MHGRL: An Effective Representation Learning Model for Electronic Health Records

Feiyan Liu[1], **Liangzhi Li**[2]*, **Xiaoli Wang**[1]*, **Chang Liu**[3]
Feng Luo[4], **Jinsong Su**[1], **Yiming Qian**[5]
[1]School of Informatics, Xiamen University, Xiamen, China
[2]Meetyou AI Lab, Xiamen, China
[3]National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, China
[4]RMIT University, Melbourne, Australia
[5]Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore
{feiyanliu, liuchang628}@stu.xmu.edu.cn, liliangzhi@xiaoyouzi.com, {xlwang, jssu}@xmu.edu.cn
feng.luo@student.rmit.edu.au, qian_yiming@ihpc.a-star.edu.sg

Abstract

Electronic health records (EHRs) serve as a digital repository storing comprehensive medical information about patients. Representation learning for EHRs plays a crucial role in healthcare applications. In this paper, we propose a Multimodal Heterogeneous Graph-enhanced Representation Learning, denoted as MHGRL, aimed at learning effective EHR representations. To address the challenge posed by data insufficiency of EHRs, MHGRL utilizes a multimodal heterogeneous graph to model an EHR. Specifically, we construct a heterogeneous graph for each EHR and enrich it by incorporating multimodal information with medical ontology and textual notes. With the integration of pre-trained model, graph neural network, and attention mechanism, MHGRL effectively incorporates both node attributes and structural information across a multimodal heterogeneous graph. Moreover, we employ contrastive learning to ensure the consistency of representations for similar EHRs and improve the model robustness. The experimental results show that MHGRL outperforms all baselines on two real clinical datasets in downstream tasks, including EHR clustering and disease prediction. The code is available at https://github.com/emmali808/MHGRL.

Keywords: representation learning, multimodal heterogeneous graph, contrastive learning

1. Introduction

The increasing number of electronic health records (EHRs) enable deep learning methods to show impressive performance in diverse tasks, such as medical diagnosis (Wang et al., 2018; Choi et al., 2018; Zhou et al., 2021), readmission prediction (He et al., 2022; Cai et al., 2022), medication recommendation (Shang et al., 2019; Yang et al., 2021; Wu et al., 2022; Yang et al., 2023), etc.

Effective EHR representations are key to achieving high performance in these tasks. Nevertheless, gathering and organizing EHR data is complex and time-consuming due to privacy regulations, limiting the available data quantity. Data from a single healthcare system is often insufficient to train deep learning models, especially for rare diseases and uncommon conditions like intensive care units. Early works (He et al., 2016; Ni et al., 2017; Zhu et al., 2016; Suo et al., 2018) represent an EHR as unordered sets of features. Sequence models that overlook structural information need large amounts of data for training. Therefore, some studies (Choi et al., 2016b, 2018, 2020; Cai et al., 2022) incorporate structural information from EHRs and utilize graph neural networks (GNNs) to generate EHR representations. These graph-based methods are more robust in situations with limited data. However, they fail to exploit external information fully, typically focusing on a single type of data, such as ontology (Choi et al., 2016b) and clinical notes (Lee et al., 2020). So there remain challenges in effectively using diverse external information for EHR representation learning.

To solve the above challenge, we propose a multimodal heterogeneous graph (MHG) to model EHR, by considering both the internal structure and external knowledge information. Figure 1 illustrates an example of MHG that contains three types of nodes: diagnosis, procedure, and medication. We use a medical knowledge graph (MKG) to obtain the relationship between nodes. To enrich the EHR representation, each node in the MHG is supplemented with multimodal information, including medical textual notes and medical ontology. Furthermore, we propose a Multimodal Heterogeneous Graph-enhanced Representation Learning (MHGRL) to integrate structural and multimodal information from MHGs, aiming to generate effective EHR representations. MHGRL consists of four modules: multimodal encoding, neighbor aggregation, node combination, and contrastive learning. We begin by employing the multimodal encoding module to generate a multimodal embedding for each node in the MHG. Next, these node embeddings and the graph structure, are fed into the neighbor aggregation module to get high-order neighborhood informa-

* Corresponding authors

LREC-COLING 2024, pages 11272–11282
20-25 May, 2024.

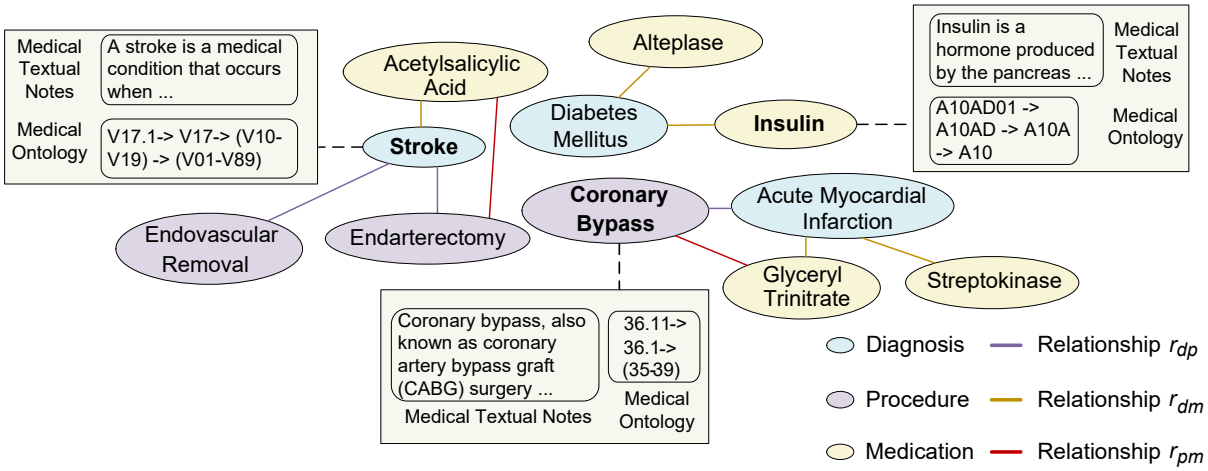

Figure 1: An example of MHG. The blue, purple, and yellow nodes represent the medical codes of diagnosis, procedure, and medication. Different colored edges represent various types of relationships between medical codes. Each node is attached with extra multimodal information, including medical textual notes and medical ontology.

tion for each node. Utilizing an attention mechanism, the node combination module combines all node embeddings to construct a comprehensive graph representation. Lastly, we implement the contrastive learning module to ensure consistency in representations among similar EHRs obtained from the same cohorts. In summary, our main contributions are as follows:

- We construct a novel MHG to accurately model EHR for representing real clinical conditions.
- We propose an EHR representation learning model, denoted by MHGRL, to incorporate both the internal structure and external knowledge information.
- The experimental results show that MHGRL outperforms all baselines on two real-world EHR datasets in two downstream tasks: EHR clustering and disease prediction, demonstrating that MHGRL can learn effective EHR representations.

2. Related Work

2.1. EHR Representation Learning

Deep learning methods are widely used in the field of EHR representation. Choi et al. (2016c) utilize the sequential order of an EHR and the co-occurrence of medical codes to simultaneously learn the representation. Hong et al. (2017) expand upon the approach of (Choi et al., 2016c) by considering the temporal relationships within clinical events. Choi et al. (2016a) propose Med2Vec, a scalable two-layer neural network for learning lower dimensional representations for medical concepts. However, they ignore the valuable structural information that can enhance medical practice.

Later approaches (Choi et al., 2016b, 2018; Wu et al., 2019; Choi et al., 2020; Liu et al., 2020; Lee et al., 2020; Cai et al., 2022) learn the representations by incorporating the hierarchical or graphical information. GRAM (Choi et al., 2016b) employs a static hierarchical ontology and MiME (Choi et al., 2018) encodes multilevel relationships between medical codes, both of which can be viewed as references for the graph structure. Recently, GNNs have been widely used for graph representation learning. And GNNs with different neighbor aggregation schemes are proposed (Kipf and Welling, 2016; Veličković et al., 2018; Schlichtkrull et al., 2018; Busbridge et al., 2019; Rampášek et al., 2022; Gravina et al., 2023; Hu et al., 2020), aiming to map the intricate information in graphs into a vector by capturing both the feature and topological information. Some approaches (Liu et al., 2020; Cai et al., 2022) utilize GNN to learn the relationships between medical codes. Nevertheless, most of them include only limited information to represent the clinical conditions. In this work, we construct each EHR into a MHG that contains structural information and external knowledge to solve the issue of data insufficiency.

2.2. Contrastive Learning

Contrastive learning is an effective framework to capture the consistency of feature representations under different views (Qiu et al., 2020; Zhu et al., 2021; Wei et al., 2022; Xia et al., 2022). Cai et al. (2022) integrates hypergraph learning and contrastive learning in EHR representation learning.

They treat different views of patient embeddings as positive samples to maximize the mutual information. In this work, the contrastive learning method is applied to ensure the consistency of representations for similar EHRs. Positive samples are constructed by utilizing EHRs from the same cohort, while negative samples are derived from distinct cohorts. We optimize representations by measuring the similarities between each EHR pair. Consequently, the representations of similar EHRs will demonstrate consistency, whereas those of dissimilar pairs will manifest divergence.

3. Preliminaries

Electronic Health Records. Each EHR consists of medical codes, including diagnosis, procedure, and medication codes. The D, P, M are the overall diagnosis, procedure, and medication sets, while $|*|$ is the cardinality of the set. So an EHR is defined as $X = \{\mathcal{V}^d, \mathcal{V}^p, \mathcal{V}^m\}$, where $\mathcal{V}^d \in D$, $\mathcal{V}^p \in P$ and $\mathcal{V}^m \in M$. The medical codes of diagnosis and procedure are mapped to the International Classification of Diseases (ICD-9) (Slee, 1978), while the medical codes of medication are obtained from the National Drug Code [1] (NDC).

We propose a graph model to represent EHRs. Specifically, we build EHRs into heterogeneous graphs based on the MKG constructed from EHRs. To extract relationships from EHRs, we utilize co-occurrence information to establish connections between diverse medical entities.

Medical Knowledge Graph. The MKG $K = (N, R)$ is a set of triples in the form (h, r, t), where $N = \{D \bigcup P \bigcup M\}$ is a set of entities, R is a set of relationships, $h, t \in N$ and $r \in R$. MKG contains three entity types $\mathcal{T}_{\mathcal{N}} = \{diagnosis, procedure, medication\}$ and three relationship types $\mathcal{T}_{\mathcal{R}} = \{r_{dp}, r_{dm}, r_{pm}\}$. r_{dp} represents the diagnosis-procedure relationship, r_{dm} represents the diagnosis-medication relationship, and r_{pm} represents the procedure-medication relationship. MKG can reveal the clinical relationships between medical entities. Thus, we use it to construct a heterogeneous graph for an EHR as below.

Heterogeneous Graph. Given an EHR, we employ the MKG to construct its heterogeneous graph, represented by $G = \{\mathcal{V}, \mathcal{E}, \mathcal{T}_{\mathcal{N}}, \mathcal{T}_{\mathcal{R}}\}$, with a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$. We have $\mathcal{V} = \{\mathcal{V}^d \bigcup \mathcal{V}^p \bigcup \mathcal{V}^m\}$. Each node in $\mathcal{V}$ is mapped to one entity in the MKG. Edge set $\mathcal{E}$ can be obtained by Algorithm 1.

To capture more useful information from EHRs, we enhance the heterogeneous graph by incorporating multimodal data, such as medical textual notes and medical ontology, into its nodes. A formal definition is as follows.

Algorithm 1: Edge Set Construction

Input: The medical knowledge graph K, the node sets of $\mathcal{V}^d = \{v_1^d, \ldots, v_{|\mathcal{V}^d|}^d\}$, $\mathcal{V}^p = \{v_1^p, \ldots, v_{|\mathcal{V}^p|}^p\}$, and $\mathcal{V}^m = \{v_1^m, \ldots, v_{|\mathcal{V}^m|}^m\}$
Output: The edge set $\mathcal{E}$

```
E ← ∅
for each v^d ∈ V^d do
    for each v^p ∈ V^p do
        if (v^d, r_dp, v^p) ∈ K then
            E ← E ∪ {(v^d, v^p)}
        end
    for each v^m ∈ V^m do
        if (v^d, r_dm, v^m) ∈ K then
            E ← E ∪ {(v^d, v^m)}
        end
for each v^p ∈ V^p do
    for each v^m ∈ V^m do
        if (v^p, r_pm, v^m) ∈ K then
            E ← E ∪ {(v^p, v^m)}
        end
return E
```

Multimodal Heterogeneous Graph. Given a heterogeneous graph G, we enrich it with multimodal data to construct a MHG, denoted by $\mathcal{G} = \{\hat{\mathcal{V}}, \mathcal{E}, \mathcal{T}_{\mathcal{N}}, \mathcal{T}_{\mathcal{R}}\}$. Each node in $\hat{\mathcal{V}}$ is associated with multimodal contents of medical textual notes and medical ontology.

To achieve this enrichment, we employ large language models (LLMs) like ChatGPT (Ouyang et al., 2022) to generate high-quality medical textual notes for each node, by designing appropriate prompt templates. We utilize the prompt: "Please give a brief definition of @NAME.", where @NAME represents medical terminology. To construct medical ontology for each node, we leverage the ICD-9 (Slee, 1978) and ATC ontology (Cheng et al., 2017). Figure 1 illustrates an example of MHG. The blue, purple, and yellow nodes represent the diagnosis, procedure, and medication codes, respectively. The purple, yellow, and red lines represent the relationship r_{dp}, r_{dm}, and r_{pm}, respectively. We use the medical terminology of node instead of its corresponding medical code for illustration. Figure 1 shows the relevant multimodal information of "Stroke", "Insulin", and "Coronary Bypass".

4. Our Model

We propose a novel model, denoted by MHGRL, that learns representations of EHRs based on the

[1] https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory

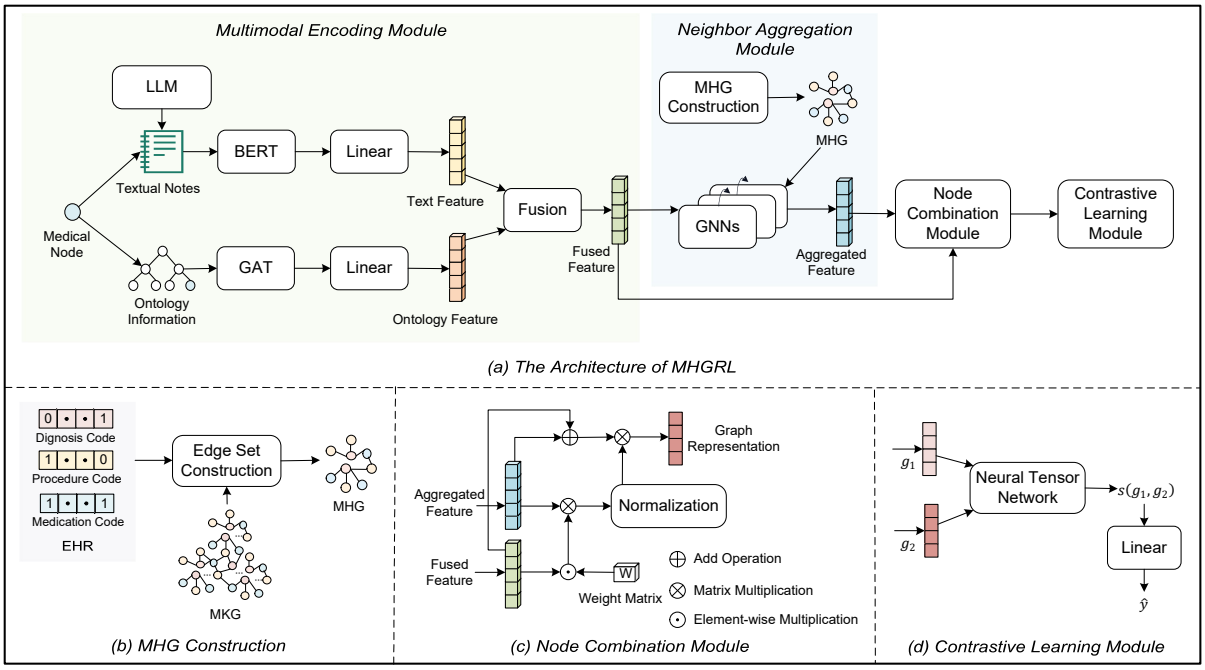

Figure 2: The architecture of our MHGRL model. There are four modules: multimodal encoding, neighbor aggregation, node combination, and contrastive learning.

proposed MHG. Figure 2 shows an overview of MHGRL. There are four modules: multimodal encoding, neighbor aggregation, node combination, and contrastive learning.

Problem Statement. Given two MHGs, represented by $\mathcal{G}_1$ and $\mathcal{G}_2$, the objective is to learn effective representations for evaluating the similarity between $\mathcal{G}_1$ and $\mathcal{G}_2$.

4.1. Multimodal Encoding Module

Given an MHG, denoted as $\mathcal{G} = \{\hat{\mathcal{V}}, \mathcal{E}, \mathcal{O}_{\mathcal{V}}, \mathcal{R}_{\mathcal{E}}\}$, this module encodes medical information of each node $v_n \in \hat{\mathcal{V}}$ ($n \in [1, |\hat{\mathcal{V}}|]$) into a fixed-size embedding $m_n \in \mathbb{R}^d$, where d is the embedding dimension. Given a node v_n, we start by deriving its textual notes embedding $t_n \in \mathbb{R}^{d_t}$ and ontology embedding $o_n \in \mathbb{R}^{d_o}$, separately. We adopt BERT (Devlin et al., 2019) as the language representation model to generate t_n with 768 dimensions. We model the medical ontology as a directed acyclic graph (DAG) (Christofides, 1975). The DAG is constructed based on the relationships between medical codes, and the ontology embedding of node v_n can be generated from its corresponding nodes in the DAG. To incorporate medical ontology knowledge, we utilize the graph attention network (GAT) (Veličković et al., 2018) to capture the structural information from DAG. We initialize each node c_i in the DAG with a vector $e_i \in \mathbb{R}^{d_x}$. Then, we employ the bottom-to-top and top-to-bottom strategies to update the node embeddings. First, the bottom-to-top strategy enhances each node by embedding the information from its children. Second, the top-to-bottom strategy updates each leaf node embedding by incorporating the information learned from its ancestor in the previous strategy. Through these steps, we derive the ontology embedding o_n of node v_n^*.

Given a DAG, the bottom-to-top strategy obtains the embedding $e'_i \in \mathbb{R}^{d_x}$ of c_i in the DAG by integrating the representations of its children.

$$
\begin{aligned}
e'_i &= \sum_{j \in ch(i)} \alpha_{ij} W e_j \\
\alpha_{ij} &= \frac{\exp(LReL(a^T[We_i || We_j]))}{\sum_{k \in ch(i)} \exp(LReL(a^T[We_i || We_k]))}
\end{aligned}
\tag{1}
$$

Here, $ch(i)$ is the indices of c_i's children (including i). $W \in \mathbb{R}^{d_x \times d_x}$ and $a \in \mathbb{R}^{2d_x}$ are weight matrix and weight vector, respectively. $LReL(\cdot)$ denotes the LeakyReLU nonlinearity and $||$ represents the concatenation operation.

Then, the top-to-bottom strategy generates the embedding $o_i \in \mathbb{R}^{d_o}$ of c_i by combining the embeddings from its ancestors.

$$
\begin{aligned}
o_i &= \sum_{j \in anc(i)} \alpha'_{ij} W' e'_j \\
\alpha'_{ij} &= \frac{\exp(LReL(a'^T[W' e'_i || W' e'_j]))}{\sum_{k \in anc(i)} \exp(LReL(a'^T[W' e'_i || W' e'_k]))}
\end{aligned}
\tag{2}
$$

Here, $anc(i)$ is the indices of c_i's ancestors (including i). $W' \in \mathbb{R}^{d_o \times d_x}$ and $a' \in \mathbb{R}^{2d_o}$ are the

learnable parameters as W and a in Equation 1.

To combine the text embedding t_n and the ontology embedding o_n into a joint embedding $m_n \in \mathbb{R}^d$, we normalize them into a shared space using the following procedure.

$$t'_n = f(t_n, W_1, b_1), \quad o'_n = f(o_n, W_2, b_2) \quad (3)$$

Here, $f(\cdot)$ is a linear project function, $W_1 \in \mathbb{R}^{\frac{d}{2} \times d_t}$, $W_2 \in \mathbb{R}^{\frac{d}{2} \times d_o}$ and $b_1, b_2 \in \mathbb{R}^{\frac{d}{2}}$ are the projection parameters.

Finally, we generate the embedding of node v_n by concatenating the projected embeddings as: $m_n = [t'_n || o'_n]$, $m_n \in \mathbb{R}^d$, where $||$ represents the concatenation operation.

4.2. Neighbor Aggregation Module

With the multimodal encoding module, the node embeddings of $\mathcal{G}$ can be denoted as $M \in \mathbb{R}^{|\hat{\mathcal{V}}| \times d}$. We then develop a neighbor aggregation module with L layers of GNNs to consolidate high-order neighborhood information for each node.

Given the adjacency matrix A of $\mathcal{G}$ as the graph structure and the initial node embeddings of $H^{(0)} = M$, the aggregated embeddings at the l^{th} layer of $H^{(l)} \in \mathbb{R}^{|\hat{\mathcal{V}}| \times d_l}$ can be formulated as follows.

$$H^{(l)} = \sigma(UpdateFunction(H^{(l-1)}, A)) \quad (4)$$

where $UpdateFunction(\cdot)$ represents the function used to update node embeddings and varies depending on the chosen GNN, and $\sigma(\cdot)$ is an activation function. Finally, we extract the aggregated embedding of $h_n \in \mathbb{R}^d$ for the node v_n from the node embeddings $H^{(L)} \in \mathbb{R}^{|\hat{\mathcal{V}}| \times d}$ in the last layer of GNNs. The aggregated node embeddings can exploit higher-order heterogeneous graph structures. Thus, h_n incorporates both multimodal and structural information for v_n.

4.3. Node Combination Module

To aggregate node embeddings into a graph representation of $g \in \mathbb{R}^d$, we adopt an attention mechanism in this module. Different from the multimodal embedding m_n, the aggregated embedding h_n incorporates the heterogeneous neighborhood information of v_n. That is, if the difference is bigger, v_n may attain more complex relationships with other nodes, and should receive a higher attention weight. By summing the initial multimodal embedding m_n and the aggregated embedding h_n, both the original and structural information are preserved in the graph representation. This helps to capture more important information and make our model more robust to the noisy neighborhood data in structural information. Thus, the graph representation $g \in \mathbb{R}^d$ can be formulated as follows.

$$g = \sum_{n=1}^{|\hat{\mathcal{V}}|} sigmoid((m_n \odot w)^T h_n)(m_n + h_n) \quad (5)$$

Here, $sigmoid((m_n \odot w)^T h_n)$ calculates the attention weight for node v_n. $w \in \mathbb{R}^d$ is the learnable parameters and $\odot$ is element-wise multiplication.

4.4. Contrastive Learning Module

This module uses contrastive learning to guide the model in learning effective graph representations. The representations generated by the node combination module reflect the node and structure information in EHRs. To ensure the consistency of representations for similar EHRs, we employ contrastive learning to maximize mutual information. Given two MHGs of $\mathcal{G}_1$ and $\mathcal{G}_2$, we first obtain two graph representations of g_1 and g_2 from the node combination module. Then this module utilizes a neural tensor network (Socher et al., 2013) to capture the interaction signals, denoted by $s(\mathcal{G}_1, \mathcal{G}_2)$. The interaction signals are computed below.

$$s(\mathcal{G}_1, \mathcal{G}_2) = \sigma({g_1}^T W_3 g_2 + W_4 \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + b_3) \quad (6)$$

Here, $W_3 \in \mathbb{R}^{d \times N \times d}$ and $W_4 \in \mathbb{R}^{N \times 2d}$ are the weight matrices, and $b_3 \in \mathbb{R}^N$ is a bias vector. N is a hyperparameter controlling the grain size of interaction signals for graph representation pairs.

After that, we feed the interaction signals of $s(\mathcal{G}_1, \mathcal{G}_2)$ into a fully connected softmax layer to output similarity score $\hat{y}$. The end-to-end model is trained to minimize the cross-entropy loss function:

$$\mathcal{L}(y, \hat{y}) = -[ylog(\hat{y}) + (1-y)log(1-\hat{y})] \quad (7)$$

where $y \in \{0, 1\}$ indicates whether two EHRs are similar or not.

5. Experimental Results

5.1. Setup

5.1.1. Datasets

MIMIC-III (Johnson et al., 2016) and MIMIC-IV (Johnson et al., 2021) are two real-world EHR datasets, containing demographics, procedures, medications, laboratory results, etc. MIMIC-III is obtained from over 40,000 patients admitted to hospitals from 2001 to 2012. MIMIC-IV is an update for MIMIC-III, incorporating data on admissions to the intensive care unit from 2008 to 2019.

MIMIC-III		MIMIC-IV	
Cohort	**# of EHRs**	**Cohort**	**# of EHRs**
CA	2,532	CA	806
CS	1,290	CS	1,627
ND	1,131	ND	3,445
SEI	1,121	EH	2,465
AVD	857	ER	1089
SEP	1021	SID	855

Table 1: The statistics of EHR cohorts.

We construct two MKGs respectively for two datasets. Then, we conduct cohort studies on diseases (Zhu et al., 2016). We select six diseases that appear frequently in two datasets respectively. These diseases are widely studied and have significant implications in healthcare applications. Moreover, they exhibit correlations that present intricate challenges in the context of joint diagnosis. The six cohorts of MIMIC-III include Coronary Atherosclerosis (CA), Cesarean Section (CS), Normal Delivery (ND), Subendocardial Infarction (SEI), Aortic Valve Disorders (AVD), and Septicaemia (SEP), and six cohorts of MIMIC-IV consist of CA, CS, ND, Essential Hypertension (EH), Suspected Infectious Disease (SID) and Esophageal Reflux (ER). The statistics of EHR cohorts are in Table 1.

We conduct extensive experiments on two tasks: EHR clustering and disease prediction. For EHR clustering, we generate MHGs using diagnosis, procedure, and medication nodes. Then, we perform the k-means clustering based on EHR representations. We use the cohorts as ground truth for evaluating the clustering results. For disease prediction, we aim to predict the cohort to which each EHR belongs accurately. Notably, for this task, we only rely on procedure and medication nodes to construct MHGs. Our approach is grounded in the assumption that procedure and medication information adequately reflects the patient's disease condition. During testing, for each EHR in the test set, we retrieve the K most similar EHRs from the training set based on their representations. The final predicted disease label is determined through a voting mechanism, where the label that appears most frequently among the K retrieved records is selected.

We generate our training, validation, and test set with a ratio of 6:2:2. For each EHR, we construct five similar pairs as positive samples by randomly selecting five other EHRs from the same cohort. Simultaneously, we build five dissimilar pairs as negative samples by randomly selecting EHRs from distinct cohorts.

5.1.2. Competitors

Given our emphasis on EHR representation learning, we select some EHR representation learning models as competitors. Besides, as our method performs EHR representation learning based on GNNs, we evaluate the performance of various GNNs. Moreover, the current LLMs are tuned to follow instructions and trained on extensive datasets to obtain zero-shot capabilities. So we design prompts to call the interface of ChatGPT in the disease prediction task, as shown in Table 4. So we select the following competitors.

- Static Representation Model: **One-hot**.
- EHR Representation Learning (RL) Model: **GRAM** (Choi et al., 2016b), **MiME** (Choi et al., 2018), and **GCT** (Choi et al., 2020).
- GNN-based Model: **GCN** (Kipf and Welling, 2016), **GAT** (Veličković et al., 2018), **RGCN** (Schlichtkrull et al., 2018), **RGAT** (Busbridge et al., 2019), **GPS** (Rampášek et al., 2022), and **A-DGN** (Gravina et al., 2023).
- LLM: **ChatGPT** (Ouyang et al., 2022).

5.1.3. Parameter Settings

We implement the model using Pytorch 2.0.1 and utilize the Adam optimizer with 30 epochs. We set the default batch size to 256, the learning rate to 0.0001, and the dropout rate to 0.4. The GNN we used is A-DGN and we set the number of layers L to 2. The dimension of the projected embeddings is 250 on MIMIC-III and 100 on MIMIC-IV, and N in our neural tensor network is 25 on MIMIC-III and 30 on MIMIC-IV. For baselines, the feature dimension is 250. We carry the experiments on a Dell server with 4 NVIDIA GeForce RTX 3090.

5.2. Performance

For EHR clustering, we choose three widely used evaluation metrics: $Purity$, normalized mutual information (NMI), and rand index (RI). We set k=6 for k-means algorithm. For disease prediction, we adopt the evaluation metric $Accuracy$ and set K=1, 3 and 5. As the experimental results in Table 2 and Table 3, our MHGRL significantly outperforms all baselines on two tasks.

Table 2 shows the results of EHR clustering. On MIMIC-III, our MHGRL achieves a $Purity$ of 0.9761, a NMI of 0.9390, and a RI of 0.9811, outperforming the strongest baseline of RGCN by 2.89%, 6.96%, and 2.21%, respectively. On MIMIC-IV, our model achieves a $Purity$ of 0.8455 and a NMI of 0.7515, outperforming the strongest baseline of GPS by 0.39% and 2.14%, respectively.

Category	Model	MIMIC-III			MIMIC-IV		
		Purity	NMI	RI	Purity	NMI	RI
Static	One-hot	0.6052	0.5023	0.7895	0.6093	0.4840	0.7587
RL	MiME	0.4368	0.2219	0.7041	0.5666	0.3773	0.7346
	GRAM	0.4840	0.3315	0.7589	0.6050	0.4517	0.7615
	GCT	0.6224	0.5137	0.8102	0.7473	0.5789	0.8142
GNN-based	GCN	0.9265	0.8243	0.9489	0.7464	0.6208	0.8738
	GAT	0.7957	0.7345	0.8753	0.7935	0.6905	0.8901
	RGCN	0.9472	0.8694	0.9590	0.7799	0.6744	0.8866
	RGAT	0.9170	0.8373	0.9331	0.7823	0.6711	0.8817
	A-DGN	0.6285	0.5101	0.8015	0.7410	0.6079	0.8681
	GPS	0.9334	0.8455	0.9530	0.8416	0.7301	**0.9146**
	MHGRL	**0.9761**	**0.9390**	**0.9811**	**0.8455**	**0.7515**	0.9141

Table 2: The EHR clustering results on two datasets. The best performance is highlighted in bold while the second best is marked with an underline.

Category	Model	MIMIC-III			MIMIC-IV		
		K=1	K=3	K=5	K=1	K=3	K=5
Static	One-hot	0.6524	0.6732	0.6795	0.4315	0.5010	0.5272
RL	GRAM	0.5368	0.5493	0.5644	0.4898	0.5078	0.5292
GNN-based	GCN	0.6952	0.7165	0.7197	0.4820	0.5194	0.5408
	GAT	0.7033	0.7285	0.7234	0.4913	0.5287	0.5603
	RGCN	0.6920	0.7291	0.7454	0.4810	0.5078	0.5384
	RGAT	0.7008	0.7285	0.7341	0.5005	0.5282	0.5646
	A-DGN	0.5632	0.5820	0.5833	0.4456	0.4830	0.5131
	GPS	0.6983	0.7040	0.7209	0.5102	0.5292	0.5539
LLM	ChatGPT	0.3206	0.3226	0.3116	0.0782	0.0779	0.0771
	MHGRL	**0.7376**	**0.7514**	**0.7657**	**0.5398**	**0.5661**	**0.5855**

Table 3: The disease prediction results on two datasets. The best performance is highlighted in bold while the second best is marked with an underline.

Table 3 shows the results of disease prediction. On MIMIC-III, MHGRL achieves the $Accuracy$ of 0.7376, 0.7514 and 0.7657, when K=1, 3 and 5, respectively. The improvement is 3.43% (vs. GAT), 2.23% (vs. RGCN), and 2.03% (vs. RGCN), respectively. On MIMIC-IV, MHGRL achieves the $Accuracy$ of 0.5398, 0.5661 and 0.5855, when K=1, 3 and 5, respectively. The improvement is 2.96% (vs. GPS), 3.69% (vs. GPS), and 2.09% (vs. RGAT), respectively. Moreover, we observe that:

(1) Both the static and early EHR representation learning models show worse performance compared to the GNN-based models. Among them, One-hot and GRAM simply treat an EHR as a medical code sequence, ignoring relationships between medical codes. MiME and GCT only capture limited relationships between medical codes. Compared to baselines, our model incorporates more information and captures important medical information using the attention mechanism.

(2) We have designed prompts and utilized the ChatGPT for disease prediction, by obtaining its top-K responses. The performance of ChatGPT on MIMIC-III, with $K = 1$, yields an $Accuracy$ of 0.3206. However, on MIMIC-IV, the $Accuracy$ drops seriously to 0.0782. Due to the limitation of tokens, it is difficult to provide additional information to ChatGPT. Predicting diseases only based on limited information might be challenging.

(3) The results on MIMIC-IV are mostly worse than MIMIC-III. The reason might be that MIMIC-IV contains more overlapping diseases compared to MIMIC-III, as shown in Figure 3. So it is difficult to distinct dissimilar EHRs of MIMIC-IV.

5.3. Ablation Study

We conduct an ablation study to analyze the effects of different modalities of data, attention mechanism, and contrastive learning in our model. Five variants are implemented:

- without (w/o) multimodal information: remove the medical ontology and textual notes informa-

Prompt	Response			Label
	K=1	K=3	K=5	
Requirement: please predict the disease based on the procedure and medication information. **Procedures**: ['Other incision with drainage of skin and subcutaneous tissue', 'Nonexcisional debridement of wound, infection or burn'] **Medications**: ['Anilides', 'Other irrigating solutions', 'Softeners, emollients', 'Contact laxatives', 'Antibiotics', 'Heparin group'] Please predict the disease name from the following six types: ['Normal Delivery', 'Essential Hypertension', 'Cesarean Section', 'Esophageal Reflux', 'Suspected Infectious Disease', 'Coronary Atherosclerosis'] NO NEED FOR ANY EXPLANATION!	ER	SID	SID	ER

Table 4: An example of prompt and ChatGPT responses for the disease prediction task. The error results are marked in red. The prompt generation process involves the following steps: (i) converting each medical code in an EHR into medical terminology; (ii) filling in the underlined positions with actual procedure and medication information extracted from the EHR.

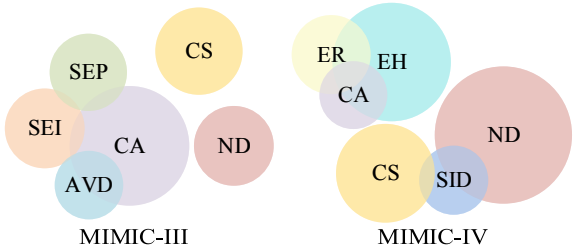

Figure 3: Disease overlapping on two datasets.

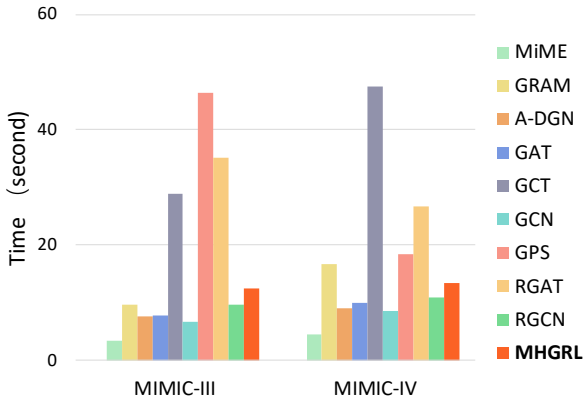

Figure 4: The training time of one epoch.

tion, and randomly initialize the representation of nodes.

- without (w/o) medical textual notes: remove textual notes information.
- without (w/o) medical ontology: remove medical ontology information.
- without (w/o) attention mechanism: not use attention mechanism in node combination.
- without (w/o) contrastive learning: without constructing similar and dissimilar EHR pairs, directly predict the cohorts.

Since the results of the two datasets are consistent, only the results of MIMIC-III are shown in Table 5. All modalities of data contribute to EHR representation learning. Using the attention mechanism can further boost the performance of representation learning. As shown in Table 5, contrastive learning also contributes to performance.

5.4. Efficiency

We further evaluate the efficiency of our model. We record the average training time of one epoch. Figure 4 shows that MHGRL takes comparable training time to strong baselines for achieving the best performance on downstream tasks.

5.5. Visualization

We use t-SNE[2] to visualize the high-dimensional EHR representations in the test set on MIMIC-III. Each EHR is plotted as a point in a two-dimensional space. In Figure 5, the points with different colors represent the EHRs in different cohorts. Our MHGRL shows the best results. The EHRs from the same cohort are plotted closely together, while those from different cohorts are separated. MiME and GRAM show the worst results. The results of

[2]https://lvdmaaten.github.io/tsne/

Model	EHR clustering			Disease prediction		
	Purity	NMI	RI	K=1	K=3	K=5
MHGRL	**0.9761**	**0.9390**	**0.9811**	**0.7376**	**0.7514**	**0.7657**
w/o multimodal information	0.6562	0.5590	0.8135	0.5931	0.6126	0.6010
w/o medical textual notes	0.9755	0.9343	0.9805	0.7319	0.7469	0.7531
w/o medical ontology	0.9617	0.9089	0.9719	0.7131	0.7236	0.7382
w/o attention mechanism	0.9679	0.9177	0.9753	0.6714	0.6942	0.7003
w/o contrastive learning	0.7142	0.6670	0.8564	0.7344	0.7502	0.7542

Table 5: Ablation study results on MIMIC-III.

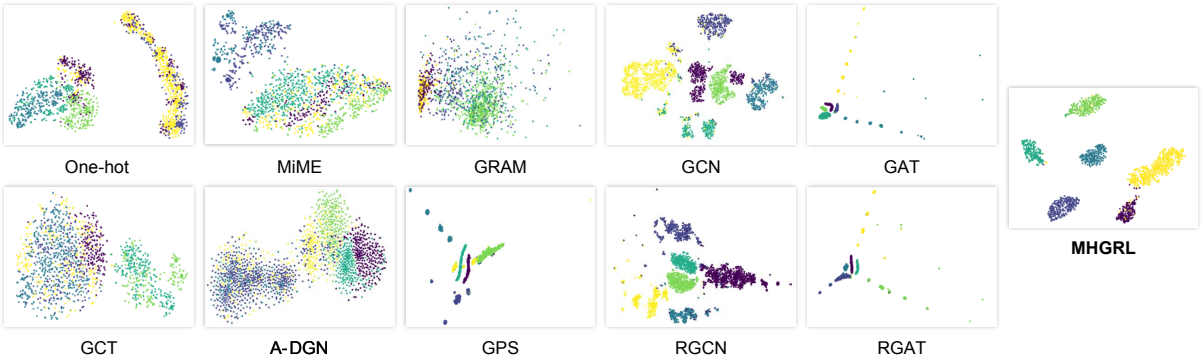

Figure 5: The visualization results of EHR clustering on MIMIC-III.

One-hot, GCT, and A-DGN look a little better because the EHRs can be separated into two parts. GCN, GAT, RGCN, RGAT, and GPS perform better by dividing the EHRs into several groups.

6. Conclusion

This paper proposes a novel model, denoted by MHGRL, to learn effective EHR representations. We first convert each EHR into a heterogeneous graph using the multimodal knowledge graph. To solve the problem of data insufficiency, every node in the graph is attached with multimodal information, including medical textual notes and medical ontology. In addition, we design an attention mechanism to aggregate node information and apply contrastive learning to ensure consistency among representations of similar EHRs and improve model robustness. The experimental results show that our MHGRL outperforms all baselines on two real datasets by learning effective EHR representations. In the future, we will incorporate additional information into our model, such as medical images, demographic information and temporal clinical data, which would help us capture the evolving information and more important clinical conditions.

7. Ethics Statement

Our work focuses on the EHR representation learning and conducts experiments on two public datasets, which poses no additional ethical issues. We ensure that our work is ethical.

8. Acknowledgments

This work was done when Feiyan Liu worked on the project in the Meetyou AI Lab. Xiaoli Wang was supported by the Natural Science Foundation of Fujian Province of China (No. 2021J01003) and the Central Guidance on Local Science and Technology Development Funds (No. 2022ZYDF082).

9. References

Dan Busbridge, Dane Sherburn, Pietro Cavallo, and Nils Y Hammerla. 2019. Relational graph attention networks. *arXiv preprint arXiv:1904.05811*.

Derun Cai, Chenxi Sun, Moxian Song, Baofeng Zhang, Shenda Hong, and Hongyan Li. 2022. Hypergraph contrastive learning for electronic health records. In *SDM*, pages 127–135, Virginia, USA.

Xiang Cheng, Shu-Guang Zhao, Xuan Xiao, and Kuo-Chen Chou. 2017. iATC-mHyb: a hybrid multi-label classifier for predicting the classification of anatomical therapeutic chemicals. *Oncotarget*, 8(35):58494.

Edward Choi, Mohammad Taha Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, Javier Tejedor-Sojo, and Jimeng Sun. 2016a. Multi-layer representation learning for medical concepts. In *KDD*, pages 1495–1504, California, USA. Association for Computing Machinery.

Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun. 2016b. GRAM: Graph-Based Attention Model for Healthcare Representation Learning. In *KDD*, pages 787–795, California, USA. Association for Computing Machinery.

Edward Choi, Cao Xiao, Jimeng Sun, and Walter F Stewart. 2018. Mime: Multilevel medical embedding of electronic health records for predictive healthcare. In *NeurIPS*, volume 2018, pages 4547–4557.

Edward Choi, Zhen Xu, Yujia Li, Michael Dusenberry, Gerardo Flores, Emily Xue, and Andrew Dai. 2020. Learning the Graphical Structure of Electronic Health Records with Graph Convolutional Transformer. In *AAAI*, volume 34, pages 606–613, New York, USA.

Youngduck Choi, Chill Yi-I Chiu, and David Sontag. 2016c. Learning low-dimensional representations of medical concepts. *AMIA Summits on Translational Science Proceedings*, 2016:41.

Nicos Christofides. 1975. *Graph theory: An algorithmic approach (Computer science and applied mathematics)*. Academic Press, Inc.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL*, Minneapolis, USA.

Alessio Gravina, Davide Bacciu, and Claudio Gallicchio. 2023. Anti-Symmetric DGN: a stable architecture for Deep Graph Networks. In *ICLR*, Kigali, Rwanda.

Lu He, Haifeng Wang, Mandana Rezaeiahari, and Chun-An Chou. 2022. An embedded machine learning model for early detection and intervention of high-risk intensive care unit readmission patients. In *IEEE BIBM*, pages 1544–1549.

Ziping He, Jijiang Yang, Qing Wang, and Jianqiang Li. 2016. A method of electronic medical record similarity computation. In *ICSH*, pages 182–191, Haikou, China. Springer Cham.

Shenda Hong, Meng Wu, Hongyan Li, and Zhengwu Wu. 2017. Event2vec: Learning representations of events on temporal sequences. In *APWeb-WAIM*, pages 33–47, Cham. Springer International Publishing.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. 2020. Strategies for pre-training graph neural networks. In *ICLR*.

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Mark Roger. 2021. MIMIC-IV (version 1.0).

Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1):1–9.

Thomas N. Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. In *ICLR*.

Dongha Lee, Xiaoqian Jiang, and Hwanjo Yu. 2020. Harmonized representation learning on dynamic ehr graphs. *Journal of Biomedical Informatics*, 106:103426.

Zheng Liu, Xiaohan Li, Hao Peng, Lifang He, and S Yu Philip. 2020. Heterogeneous similarity graph neural network on electronic health records. In *IEEE BigData*, pages 1196–1205. IEEE.

Jiazhi Ni, Jie Liu, Chenxin Zhang, Dan Ye, and Zhirou Ma. 2017. Fine-grained patient similarity measuring using deep metric learning. In *CIKM*, pages 1189–1198, Singapore. Association for Computing Machinery.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. In *NeurIPS*, volume 35, pages 27730–27744.

Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. 2020. Gcc: Graph contrastive coding for graph neural network pre-training. In *KDD*, pages 1150–1160.

Ladislav Rampášek, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. 2022. Recipe for a general, powerful, scalable graph transformer. In *NeurIPS*, volume 35, pages 14501–14515.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *ESWC*, pages 593–607, Heraklion, Crete, Greece,. Springer.

Junyuan Shang, Cao Xiao, Tengfei Ma, Hongyan Li, and Jimeng Sun. 2019. Gamenet: Graph augmented memory networks for recommending medication combination. In *AAAI*, volume 33, pages 1126–1133.

Vergil N Slee. 1978. The international classification of diseases: ninth revision (icd-9).

Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *NeurIPS*, pages 926–934, Carlifornia, USA.

Qiuling Suo, Fenglong Ma, Ye Yuan, Mengdi Huai, Weida Zhong, Jing Gao, and Aidong Zhang. 2018. Deep patient similarity learning for personalized healthcare. *IEEE transactions on nanobioscience*, 17(3):219–227.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *ICLR*.

Xiaoli Wang, Yuan Wang, Chuchu Gao, Kunhui Lin, and Yadi Li. 2018. Automatic diagnosis with efficient medical case searching based on evolving graphs. *IEEE Access*, 6:53307–53318.

Wei Wei, Chao Huang, Lianghao Xia, Yong Xu, Jiashu Zhao, and Dawei Yin. 2022. Contrastive meta learning with behavior multiplicity for recommendation. In *WSDM*, pages 1120–1128. Association for Computing Machinery.

Rui Wu, Zhaopeng Qiu, Jiacheng Jiang, Guilin Qi, and Xian Wu. 2022. Conditional generation net for medication recommendation. In *WWW*, pages 935–945.

Tong Wu, Yunlong Wang, Yue Wang, Emily Zhao, Yilian Yuan, and Zhi Yang. 2019. Representation learning of ehr data via graph-based medical entity embedding. *arXiv preprint arXiv:1910.02574*.

L Xia, C Huang, Y Xu, J Zhao, D Yin, and J Huang. 2022. Hypergraph contrastive collaborative filtering. In *SIGIR*. ACM.

Chaoqi Yang, Cao Xiao, Fenglong Ma, Lucas Glass, and Jimeng Sun. 2021. Safedrug: Dual molecular graph encoders for safe drug recommendations. In *IJCAI*.

Nianzu Yang, Kaipeng Zeng, Qitian Wu, and Junchi Yan. 2023. Molerec: Combinatorial drug recommendation with substructure-aware molecular representation learning. In *WWW*, pages 4075–4085.

Xiaokang Zhou, Yue Li, and Wei Liang. 2021. Cnn-rnn based intelligent recommendation for online medical pre-diagnosis support. *TCBB*, 18(3):912–921.

Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. 2021. Graph contrastive learning with adaptive augmentation. In *WWW*, pages 2069–2080, Ljubljana, Slovenia. Association for Computing Machinery.

Zihao Zhu, Changchang Yin, Buyue Qian, Yu Cheng, Jishang Wei, and Fei Wang. 2016. Measuring patient similarities via a deep architecture with medical concept embedding. In *ICDM*, pages 749–758, Barcelona, Spain.