Fusion of Dynamic Hypergraph and Clinical Event for Sequential Diagnosis Prediction

Xin Zhang[1,3], Xueping Peng[2], Hongjiao Guan[1,3], Long Zhao[1,3], Xinxiao Qiao[1,3*], and Wenpeng Lu[1,3*]
[1] Key Laboratory of Computing Power Network and Information Security, Ministry of Education,
Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[2]Australian Artificial Intelligence Institute, University of Technology Sydney, Sydney, Australia
[3] Shandong Provincial Key Laboratory of Computer Networks,
Shandong Fundamental Research Center for Computer Science, Jinan, China
zhangxinqlu@foxmail.com, {xueping.peng}@uts.edu.au,
{hongjiao.guan,zhaolong,qxxyn,wenpeng.lu}@qlu.edu.cn

***Abstract*—Sequential diagnosis prediction (SDP) is a challenging task, aiming to predict patients' future diagnoses based on their historical medical records. While methods based on graph neural networks (GNNs) have proven successful for this task, they typically focus on modeling pairwise diseases using a global disease combination graph. However, these approaches neglect the fine-grained higher-order relations among persistent and emerging diseases within a single visit, which may contain crucial clues to predict the next diagnosis. Additionally, they fail to fully leverage patient-related clinical information present in electronic health records (EHRs). To address these challenges, this paper proposes a novel approach called the fusion of Dynamic Hypergraph and Clinical Event (DHCE) for sequential diagnosis prediction. The proposed method aims to exploit the fine-grained higher-order relations among diagnoses within a visit and leverage clinical event information from EHRs to improve the accuracy of predicting the next diagnosis. Specifically, DHCE categorizes diagnoses within a single visit in a fine-grained granularity into persistent and emerging categories based on a patient's historical diagnoses. It then constructs dynamic hypergraphs to capture higher-order disease relations within each visit. Next, we design a transition function to extract the transitional context from previous visits in order to generate the visit representation. Furthermore, to fully leverage patient-related clinical events in a visit, we utilize Bio-Clinical BERT to encode them and generate the clinical event representation for each visit. Finally, we combine the visit representation and event representation to generate a comprehensive patient representation, which is then used to predict the patient's next diagnosis. Experimental results on two benchmark datasets consistently demonstrate that DHCE outperforms state-of-the-art methods[1].**

***Index Terms*—dynamic hypergraph, clinical event, sequential diagnosis prediction, Bio-Clinical BERT**

I. Introduction

Sequential diagnosis prediction aims to forecast future patient diagnoses utilizing their historical electronic health records (EHRs), which consist of temporally-ordered sequences of patient visits [1]–[6]. Many research efforts have focused on analyzing patient visit records as ordered sequences. Notably, recurrent neural networks (RNNs) and graph neural networks (GNNs) approaches have achieved remarkable success in this domain.

For RNN-based approaches, the key to success lies in modeling a patient's historical visits as a strictly-order sequence due to the intrinsic chronological order of visits [7], [8]. However, it should be noted that the diagnosis in neighboring visits may be independent. For example, the diagnosis in the current visit might be related to a diagnosis from a distant visit rather than the most recent one. Additionally, the diagnosis in the current visit may be an emerging disease unrelated to the patient's historical diagnoses. Moreover, the existing RNN-based SDP models often oversimplify the representations of a visit by aggregating the representation of the diagnoses, which hinders effective modeling of the relationships among diagnoses within a single visit. Consequently, RNN-based models exhibit limited performance on SDP tasks.

Graph neural networks (GNNs) have recently been reported to be effective in many areas, including sequential diagnosis prediction [9]–[13]. Unlike RNN-based methods, GNN-based methods model the diagnoses in a visit as a pairwise graph. In real scenarios, a diagnosis in a single visit may be either a persistent disease or an emerging one, which plays different impacts on the next diagnosis. It is necessary to categorize diagnoses in a visit into persistent diseases and emerging ones to investigate their influences in more fine-grained granularity. Furthermore, a disease may be caused by multiple historical diseases. There exist many-to-many/one, higher-order relations between diseases. Therefore, GNNs-based SDP methods fail to model the fine-grained, higher-order relations between diagnoses within one visit. In addition, GNN-based methods always focus on the information of diseases and neglect the patient-related clinical event information. However, it is obvious that the event information contains important clues to predict the possible diagnosis in the current visit, which is valuable and indispensable. Therefore, (Gap 2) *GNNs-based SDP methods fail to fully utilize clinical event information in EHRs*.

This paper addresses the above two gaps by proposing *sequential diagnosis prediction with Dynamic Hypergraph*

*Corresponding authors.
[1]Our source code is avaliable at https://github.com/HELJJ/DHCE.

and Clinical Event (DHCE). First, Unlike the traditional graph neural networks such as GCN that model diagnoses in a visit as a pairwise graph to capture the pairwise relations among diagnoses. To model the sophisticated information between diagnoses in a visit, according to the historical diseases of a patient, we categorize the diagnoses into two types: persistent diseases and emerging ones and then construct dynamic hypergraphs to capture fine-grained, high-order relations among them. Besides, in order to fully utilize the patient-related clinical events in a visit, we first convert the clinical features into textual descriptions, then utilize Bio-Clinical BERT to encode different types of clinical events. The main contributions of this work are summarized as follows:

- We propose a novel model for sequential diagnosis prediction with dynamic hypergraph and clinical event (DHCE). In contrast to the existing SDP models, DHCE constructs dynamic hypergraphs to capture the higher-order relations in a visit and utilizes Bio-Clinical BERT to model clinical events. Both visit representation and event representation are combined to generate a patient representation for predicting the next diagnosis. As far as we know, this is the first model for leveraging both dynamic hypergraphs and clinical events to predict sequential diagnosis.
- We propose a novel method to construct and learn dynamic hypergraphs for visit representation. We categorize diagnoses into persistent diseases and emerging ones and construct dynamic hypergraphs to learn higher-order information for two categories of diseases. We design a transition function to extract the transitional context from previous visits to generate the visit representation. We are able to perform efficient dynamic hypergraph learning to model more fine-grained relations among diseases, resulting in well-optimized visit representation.
- We propose a novel method for utilizing clinical events to enhance patient representation. We employ medical pre-trained language models to encode the textual information of clinical events in a visit and aggregate the representations of all events as the event representation of the current visit. We are able to capture more event information to represent a patient, benefiting our model to show better performance.

We conduct comprehensive experiments on two real-world EHR datasets to show the improvement of DHCE over the state-of-the-art models on prediction accuracy.

II. RELATED WORK

A. Traditional Sequential Prediction on EHRs

Most of the existing SDP models rely on RNNs and attention mechanisms, which focus on certain parts of the sequences according to different requirements. They have shown their effectiveness in learning time-series EHR data. For example, HiTANet [14] leveraged hierarchical and time-aware attention mechanisms to capture complex temporal relations in EHRs to predict the health risks of a patient. Concare [15] utilized contextual information from healthcare data to create tailored embeddings for individual patients to represent their personal health context. Dipole [16] applied various attention mechanisms on bi-directional RNN to predict diagnoses in future visits. The successes of RNN-based methods rely on the assumption that the diagnoses in neighboring visits are dependent on each other. However, this may not be true. The diagnosis in the current visit may be related to the diagnosis in a visit a long time ago, instead of the last visit. Moreover, the existing RNN-based SDP models always simply aggregate the representations of diagnoses in a visit as the representation of the visit, which is unable to capture the relations among diagnoses within a visit, resulting in unsatisfied performance.

B. Graph Learning on EHRs

In recent years, GNNs have been widely applied on SDP task, which encode EHR graphs and to learn the connections between different diagnosis nodes. The graph convolutional transformer (GCT) [17] was proposed as a general framework combining Transformer and GNNs. CGL [9] applied a collaborative graph learning model with attention regulation strategy to explore patient-disease interactions and medical domain knowledge to predict temporal event in healthcare. T-ContextGGAN [18] employed a GNN-based model with time-aware meta-paths and self-attention mechanism, to extract both temporal semantic information and inherent relations of EHR data simultaneously, performing automatic meta-path selection for clinical risk prediction. Although these GNN-based methods can achieve some improvements over the RNN-based methods, they model the diagnoses in a visit as a pairwise graph. However, in practice, diagnoses in a visit may be persistent diseases or emerging ones. Furthermore, a disease may be caused by multiple historical diseases. There exist many-to-many/one, higher-order relations between diseases. The existing GNN-based methods are unable to model the fine-grained, higher-order information between diagnoses in a visit. This hurts the performance of GNN-based methods severely.

C. Hypergraph Graph Learning on EHRs

Current popular SDP methods are hypergraph-based approaches. In hypergraphs, intricate associations are represented by hyperedges which link an indefinite number of nodes [19]. Hyperedges and hypergraphs can effectively expand the traditional binary relations in graph structures to multivariate relations, thereby allowing for the representation of intricate correlations between data [20], [21]. Compared with graph modeling, hypergraph modeling has attracted more attention from research communities in recent years due to its increased flexibility in depicting complex data associations [22]. For the SDP task, a hypergraph can be constructed from the relations between patient visit records and medical codes in EHR systems, where each patient visit is represented as a hyperedge, and each diagnosis in a visit is represented as

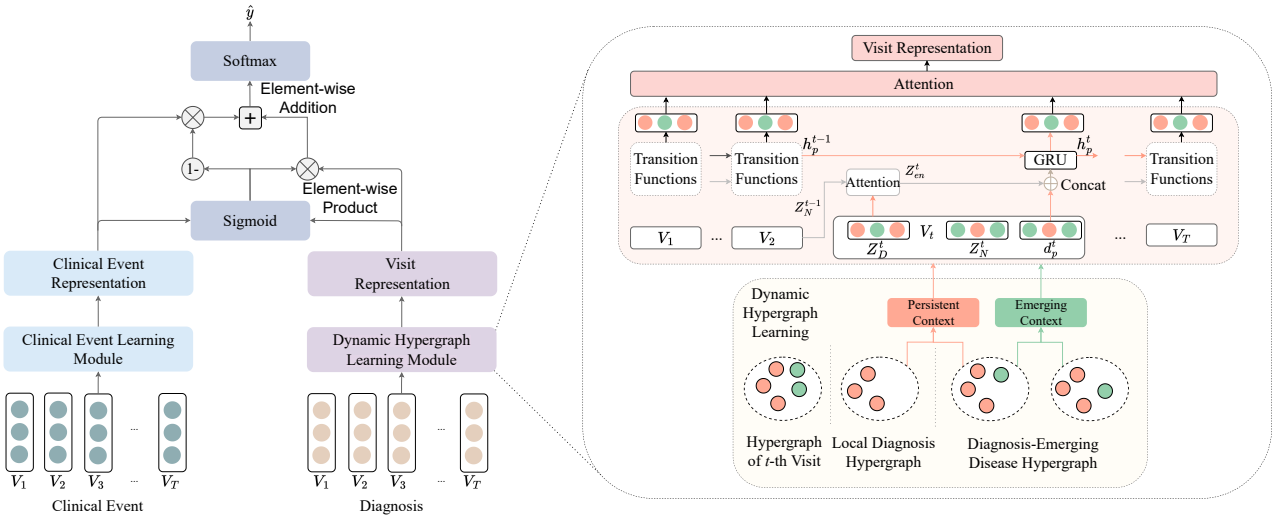

Fig. 1: The architecture of overall DHCE framework. It includes a dynamic hypergraph learning module, a clinical event learning module, and a diagnosis prediction module. The dynamic hypergraph learning module learns the visit representation using a hypergraph convolutional network. The clinical event learning module encodes the clinical event with Bio-clinical BERT to obtain the clinical event representation. The diagnosis prediction module fuses the visit representation and clinical event representation to generate the patient representation, which is utilized to predict the future diagnoses.

a node. CACHE [23] provided effective and insightful clinical predictions based on hypergraph representation learning and factual reasoning techniques. HCL [24] proposed hypergraph contrastive learning to jointly learn patient embeddings and code embeddings from EHR. Although hypergraph-based methods can model the relations among diagnoses more effectively, they still neglect to consider the fine-grained relations, such as persistent diseases and emerging ones. Besides, they also neglect to utilize the valuable information contained in clinical events. There still is great room to improve the performance of hypergraph-based methods.

III. METHODOLOGY

In this section, we first introduce notations and problem statements and then present hypergraph construction. After that, we introduce the details of the proposed DHCE model, including the dynamic hypergraph learning module, clinical event learning module and diagnosis prediction module.

A. Notations and Problem Statement

1) Notations: In each visit, the patient is diagnosed with one or more diseases, represented by medical codes, such as ICD-9-CM or ICD-10. For instance, *right heart failure* is coded as *428.2* in ICD-9-CM.

An EHR dataset is given by $\{\gamma_u | u \in \mathbb{U}\}$, where $\mathbb{U}$ is the set of patients, and $\gamma_u = (V_1^u, V_2^u, \cdots, V_T^u)$ is a visit sequence of the patient u. Each visit $V_t^u = \{C_t^u, E_t^u\}$ is recorded with a subset of medical codes $C_t^u \subset \mathbb{C}$, and clinical events $E_t^u \subset \mathbb{E}$. $\mathbb{C} = \{c_1, c_2, \cdots, c_{|\mathbb{C}|}\}$ refers to the entire set of diseases represented by medical codes in the EHR dataset, where $|\mathbb{C}|$ is the number of medical codes. Furthermore, let $\mathbb{E} = \{E_1, E_2, \cdots, E_{|\mathbb{E}|}\}$ denote the collection of clinical events from the EHR dataset, with $|\mathbb{E}|$ indicating the total number of clinical events. The i-th medical event in visit t for a patient u, denoted as E_{ti}^u, consists of a corresponding event type $q_i \in \mathbb{Q}$ (e.g., *lab test, prescription, procedure*), and a set of clinical features $\{A_i^1, \cdots, A_i^{|m_i|}\}$ associated with the event type, where $|m_i|$ denotes the number of clinical features. Each feature A_i^k can be seen as a tuple of a feature name and its value (n_i^k, v_i^k), $n_i^k \in \mathbb{N}$, $v_i^k \in \mathbb{X}$, where $\mathbb{N}$ and $\mathbb{X}$ are each a set of unique feature names (e.g., *diagnosis code*, *drug name*) and feature values (e.g., *401.9*, *vancomycin*).

2) Problem Statement: Given a patient u with T historical visit records, the goal of the sequential diagnosis prediction task is to predict the diagnoses appearing in next visit of the patient. For example, given an EHR dataset, the target is to predict the probability of the medical code appearing in the $(T+1)$-th visit, that is, $y^{T+1} \in \{0,1\}^{|\mathbb{C}|}$.

B. Hypergraph Construction

In order to capture higher-order interactions among diagnoses in a visit, we introduce hypergraphs, which can model the beyond pairwise relations. Specifically, we construct the hypergraph by defining the diagnoses in a visit as nodes, and the visit as a hyperedge. For example, as illustrated in the bottom-left panel of Fig. 1, the hypergraph for the t-th visit consists of five nodes and one hyperedge, where each node represents a diagnosis in the visit and the hyperedge connects the five nodes.

For each visit, we can construct one hypergraph. Therefore, for the sequence of visits of a patient, we can obtain the dynamic hypergraph of the patient. Let $\mathcal{G} = \{\mathcal{G}^1, \mathcal{G}^2, \cdots, \mathcal{G}^T\}$ denote the dynamic hypergraph, and $\mathcal{G}^t = (\mathbb{P}^t, \mathbb{H}^t)$ denote the hypergraph of the t-th visit. $\mathbb{P}^t \subset \mathbb{C}$ represents the set of all nodes in $\mathcal{G}^t$, that is, the set of all diagnoses of the t-th

visit. Similarly, $\mathbb{H}^t$ denotes the hyperedge (i.e. t-th visit) in $\mathcal{G}^t$. $\mathcal{G}^t$ has an incidence matrix $\mathbf{O}^t$ of size $|\mathbb{P}^t| \times |\mathbb{H}^t|$. When the patient u is diagnosed with the disease c_i in the visit t, $\mathbf{O}^t_{c_i u}$ is set to 1, otherwise, 0.

In real scenarios, a diagnosis in current visit may be a persistent disease or an emerging one. For instance, if a patient is diagnosed with *chronic heart failure* in the first visit, the diagnosis will appear in all subsequent visits. It is a persistent disease. On the other hand, a patient may be diagnosed with *appendicitis*. It may be an emerging disease, which doesn't appear in the last visit.

In order to model the fine-grained relations among persistent and emerging diseases, for the diseases in diagnosis set $\mathbb{D}^t$ of visit t ($|t| \geq 2$), we further categorize them into persistent diseases and emerging diseases:

- **Persistent diseases**: $\mathbb{D}^t_p = \mathbb{D}^t \wedge \mathbb{D}^{t-1} \in \{0,1\}^{|\mathbb{C}|}$, i.e., diagnoses in visit t that are also appear in visit t-1.
- **Emerging diseases**: $\mathbb{D}^t_{em} = \mathbb{D}^t \wedge \neg \left(\mathbb{D}^{t-1}\right) \in \{0,1\}^{|\mathbb{C}|}$, i.e., diagnoses in visit t that does not appear in visit t-1.

Once categorizing the diseases into two groups, for the visit t, we construct two kinds of dynamic hypergraphs according to the original $\mathcal{G}^t$:

- **Local diagnosis hypergraph** $\mathcal{G}^t_D$: It is a hypergraph consisting of all persistent diseases in visit t. In the hypergraph, persistent diseases are viewed as nodes and the visit is viewed as edges.
- **Diagnosis-emerging disease hypergraph** $\mathcal{G}^t_{DN}$: It is a hypergraph describing the connection of persistent and each emerging disease. In the hypergraph, diseases are used as nodes and the visit is used as edges.

In reality, a diagnosis in a visit may be triggered by the effects of the development of an existing diagnosis or the collaborative effects of multiple diagnoses. In order to model the effects, in a dynamic graph layer, we extract the local context as well as the emerging context:

- **Local Context**. For each diagnosis node, we aggregate the embeddings of connected diagnosis nodes from $\mathcal{G}^t_D$ and aggregate the embeddings of connected emerging disease nodes from $\mathcal{G}^t_{DN}$, to generate the local context $\mathbf{Z}^t_D$, described as:

$$\mathbf{Z}^t_D = HyperGCN(\mathcal{G}^t_D) + HyperGCN(\mathcal{G}^t_{DN}), \quad (1)$$

  where $HyperGCN(\cdot)$ denotes hypergraph convolutional networks.
- **Emerging Context**. For each emerging disease node, we aggregate the embeddings of connected diagnosis nodes from $\mathcal{G}^t_{DN}$, to generate the emerging context $\mathbf{Z}^t_N$, described as:

$$\mathbf{Z}^t_N = HyperGCN(\mathcal{G}^t_{DN}). \quad (2)$$

The above operations focus on the modeling of diseases in the same visit. Obviously, there exist close relations between diagnoses in different visits. In order to further capture the relations, we design transition function to extract the transition context from previous visits.

For emerging diseases $\mathbb{D}^t_{em}$, they may be triggered in a visit by the collaborative effects of existing diagnoses. We design a scaled dot-product attention as the transition function to calculate emerging transitional context, described as:

$$\mathbf{Z}^t_{en} = \mathrm{Att}(\mathbf{Z}^{t-1}_N, \mathbf{Z}^{t-1}_N, \mathbf{Z}^t_D), \quad (3)$$

where the attention function is defined as below:

$$\mathrm{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{W}_{\mathrm{q}} \left(\mathbf{K}\mathbf{W}_{\mathrm{k}}\right)^\top}{\sqrt{\mathrm{a}}}\right)\mathbf{V}\mathbf{W}_{\mathrm{v}}. \quad (4)$$

Here, $\mathbf{W}_k$, $\mathbf{W}_q$ and $\mathbf{W}_v$ are attention weights, a is the attention size.

For persistent diseases $\mathbb{D}^t_p$, they inherit information from previous diagnoses. We keep persistent diseases unchanged.

In visit t, transition output $\mathbf{h}^t_p$ is calculated with GRU according to the embeddings of persistent diseases $\mathbf{h}_{\mathbb{D}^t_p}$ in visit t and the emerging transitional context $\mathbf{Z}^t_{en}$:

$$\mathbf{h}^t_p = \mathrm{GRU}\left(\mathrm{Concat}(\mathbf{h}_{\mathbb{D}^t_p}, \mathbf{Z}^t_{en}), \mathbf{h}^{t-1}_p\right), \quad (5)$$

where $\mathbf{h}^{t-1}_p$ denotes the output from GRU in visit t-1, $\mathrm{Concat}(\cdot)$ represents concatenate operation. Specifically, when t=1, since there are no emerging diseases yet in the first visit, we let $\mathbb{D}^t_p = \mathbb{D}^1_p$ and use the GRU with an initial hidden state $\mathbf{h}^0_p = 0$ to calculate $\mathbf{h}^1_p$, describe as:

$$\mathbf{h}^1_p = \mathrm{GRU}\left(\mathbf{h}_{\mathbb{D}^t_p}, \mathbf{h}^0_p\right), \quad (6)$$

where $\mathbf{h}_{\mathbb{D}^t_p}$ denotes the embeddings of persistent diseases.

Then, we use max pooling on the transition output $\mathbf{h}^t_p$ to generate the visit embedding $\mathbf{v}^t$ of visit t, described as:

$$\mathbf{v}^t = \mathrm{maxpooling}(\mathbf{h}^t_p). \quad (7)$$

Finally, we apply an attention mechanism to calculate the visit representation $\mathbf{O}_v$ of all visits, as below:

$$\boldsymbol{\alpha} = \mathrm{softmax}\left(\left[\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^T\right]\mathbf{W}_\alpha\right), \quad (8)$$

$$\mathbf{O}_v = \boldsymbol{\alpha}\left[\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^T\right]^\top, \quad (9)$$

where $\mathbf{W}_\alpha$ is trainable parameter and $\boldsymbol{\alpha}$ is the attention score.

C. Clinical Event Learning Module

In EHR systems, each visit of a patient can be represented as a sequence of clinical events, such as *lab test*, *prescription* and *symptoms*. These events provide important clues for doctors to assess physical condition of a patient, which also benefit diagnosis prediction of the patient. For example, if a patient shows symptoms of *persistent chest pain* and *shortness of breath*, and doctors prescribes medication for the patient to prevent *blood clotting*, the combination of these clinical events mean that the patient may be diagnosed as *coronary artery disease (CAD)*.

In order to capture patient's conditional information contained in clinical events, following the work of Hur et

al. [25], we adopt text-based embeddings, where various feature values of clinical events are first converted to textual description. And, we employ Bio-Clinical BERT [26] to encode the descriptions. Then, We utilize attention mechanism to aggregate multiple events to generate the clinical event representation. The detailed operations are described as follows:

1) Preprocessing: For facilitating the modeling of clinical events with Bio-Clinical BERT, we first convert feature values of clinical events to textual descriptions. For instance, a feature value *486* is converted to *Pneumonia, organism unspecified*.

As described in Section III-A1, the i-th medical event in visit t for a patient u, denoted as E_{ti}^{u}, consists of a corresponding event type $q_i \in \mathbb{Q}$ and a set of clinical features $\{A_i^1, \cdots, A_i^{|m_i|}\}$ associated with the event type. Each feature A_i^k consists of a feature name and its value (n_i^k, v_i^k). For clinical events in each visit, it is converted as: [CLS]q_i[SEP]n_i^1[SEP]v_i^1[SEP]n_i^2[SEP]v_i^2[SEP]$\cdots$[SEP]$n_i^{|m_i|}$ [SEP]$v_i^{|m_i|}$, where [CLS] and [SEP] are special tokens.

2) Clinical Event Embedding: To capture the essential information contained in textual descriptions of clinical events, we feed them into Bio-Clinical BERT [26]. Specifically, for a sequence of clinical events of type q_i generated by a patient in visit t, the clinical event embedding of q_i can be formalized as:

$$\mathbf{m}_{q_i} = f(S(q_i), S(n_i^1), S(v_i^1), ..., S(n_i^{|m_i|}), S(v_i^{|m_i|})), \quad (10)$$

where S denotes a sub-word tokenizer, f refers to the Bio-Clinical BERT and generates the embedding representation e_{q_i} of clinical event q_i.

3) Clinical Event Embedding Aggregation: In each visit, there exists a sequence of clinical events. To obtain their overall representation, we employ attention mechanism to aggregate them, described as:

$$\mathbf{O}_e = \text{aggr}(\mathbf{m}_{q_1}, \mathbf{m}_{q_2}, ..., \mathbf{m}_{q_{|\mathrm{L}|}}), \quad (11)$$

where $\text{aggr}(\cdot)$ represents the aggregation function based on attention mechanism [27], $|L|$ represents the number of clinical event types in visit t. $\mathbf{O}_e$ is the clinical event representation.

D. Diagnosis Prediction Module

In order to fully utilize the information in dynamic hypergraph and clinical event, we utilize a gate mechanism to merge them to generate the patient representation, followed by the softmax function to predict next diagnosis of the patient, as described in Equ. (12). We employ the cross-entropy loss in Equ. (13) to optimize the model [2], [28].

$$\begin{aligned} F &= \text{sigmoid}(\mathbf{W}_e\mathbf{O}_e + \mathbf{W}_v\mathbf{O}_v + \mathbf{b}_f), \\ \mathbf{u} &= F \odot \mathbf{O}_v + (1 - F) \odot \mathbf{O}_e, \\ \hat{y} &= \text{softmax}(\mathbf{W}_y\mathbf{u} + \mathbf{b}_y), \end{aligned} \quad (12)$$

$$\mathcal{L} = \frac{1}{T+1}\sum_{t=2}^{T} -(y_{true}^T log\hat{y} + (1 - y_{true})^T log(1 - \hat{y})). \quad (13)$$

Here, $\mathcal{L}$ represents the loss function, $\mathbf{O}_e$ represents the clinical event representation, $\mathbf{O}_v$ represents the visit representation, $\mathbf{W}_e$, $\mathbf{W}_v$, $\mathbf{W}_y$ and $\mathbf{b}_y$ are learnable parameters, and $\hat{\mathbf{y}}$ is a multi-hot vector whose value is 1 if the i-th diagnosis appears in next visit, otherwise 0.

IV. EXPERIMENT

A. Datesets and Preprocessing

We employ two benchmark datasets, namely, MIMIC-III [29] and MIMIC-IV [30], which are two publicly available electronic health record datasets created by MIT Laboratory of Computational Physiology. MIMIC-III is the third version of the MIMIC dataset, which contains medical information on approximately 40,000 patients from 2001 through 2012. MIMIC-IV is the latest version of the MIMIC dataset, which contains medical information on approximately 80,000 patients from 2008 through 2019.

Following the preprocessing of Lu et al. [10], we filter out visits less than 2 times on both datasets. For MIMIC-IV, we randomly sample 10,000 patients from 2013 to 2019. After preprocessing, the statistics of datasets are summarized in Table I. For MIMIC-III, we further split it into three groups: 6,000 patients for training, 493 patients for validation, and 1000 patients for test. For MIMIC-IV, the patients for training, validation and test are 8,000, 1,000 and 1,000.

TABLE I: Statistics of datasets.

Dataset	MIMIC-III	MIMIC-IV
# patients	7,493	10,000
Max. # visit	42	97
Avg. # visit	2.64	3.72
# codes	4,753	6128
Max. # codes per visit	39	50
Avg. # codes per visit	13.41	13.38
# events	18168	92497
Max. # event per visit	3	3
Avg. # event per visit	2.21	1.88

B. Parameter Settings and Evaluation Metrics

We configure parameters in our experiments based on the validation set. Specifically, our model uses an embedding size of 48 for persistent diseases and emerging diseases, and an attention size of 32. For diagnosis prediction, the hidden units p of GRU is 256 on MIMIC-III and 512 om MIMIC-IV. The training process uses the Adam optimizer, with the learning rate set at 0.01 for 200 epochs. We implemented all of our programs on a machine with Intel(R) Xeon(R) Gold 6150 CPU, 29GB memory, and NVidia Gefoce GTX 3090, using PyTorch 1.8.1 with CUDA 11.1 and Python 3.8.6. We adopted the weighted F_1 score (w-F_1), top k recall (R@k) as evaluation metrics.

TABLE II: Performance on the MIMIC-III and the MIMIC-IV datasets in comparison with the SOTA models. The best results of each column are highlighted in boldface, the suboptimal one is underlined.

Method classification	Method	MIMIC-III			MIMIC-IV		
		w-F_1	R@10	R@20	w-F_1	R@10	R@20
CNN-based method	Deepr	18.87	24.74	33.47	24.08	26.29	33.93
RNN-based methods	RETAIN	20.69	26.13	35.08	24.71	28.02	34.46
	Dipole	19.35	24.98	34.02	23.69	27.38	35.48
	Timeline	20.46	25.75	34.83	25.26	<u>29.00</u>	37.13
	HiTANet	21.15	26.02	35.97	24.92	27.45	36.37
Graph-based methods	GRAM	21.52	26.51	35.80	23.50	27.29	36.36
	G-BERT	19.88	25.86	35.31	24.49	27.16	35.86
	CGL	21.92	26.64	36.72	25.41	28.52	37.15
	Chet	<u>23.63</u>	<u>28.15</u>	<u>37.87</u>	<u>26.93</u>	28.52	<u>38.31</u>
	DHCE (our)	**24.24**	**29.53**	**39.03**	**27.86**	**30.28**	**39.89**
	Improvement(%).	↑2.58	↑4.90	↑3.06	↑3.46	↑6.17	↑4.12

C. Baselines

We compare the performance of our proposed DHCE model against the following state-of-the-art models.

- **RETAIN** [31], which first learns the representation of medical concept and then utilizes RNN with a reverse temporal attention mechanism to predict patient visit information.
- **Dipole** [16], which employs bidirectional RNN together with three attention mechanism, including location-based, general and concatenation-based ones, to predict the diagnoses in next visit.
- **Timeline** [8], which assigns time decay factors to every medical code, and learns the impact of chronic and acute conditions. Besides, It utilizes an attention mechanisms to enhance visit representation and improve the prediction of next diagnosis.
- **HiTANet** [14], which implements a hierarchical time-aware attention network, modeling time information in local and global stages. It utilizes a time-aware Transformer and key-query attention mechanism to generate patient representation for next diagnosis prediction.
- **Deepr** [32], which learns medical concept embeddings via a convolutional neural network to predict the unplanned readmission after discharge.
- **GRAM** [33], which represents a medical concept as the combination of its ancestors on knowledge graph, and learn the representations from the graph to predict next diagnosis.
- **G-BERT** [34], which utilizes GNNs to model internal hierarchical structures of medical codes, and integrates GNN representation into transformer-based encoder to generate visit representation for sequential diagnosis prediction.
- **CGL** [9], which constructs a collaborative graph that connects patients and diseases based on their co-occurrence to model patient-disease interactions to predict clinical events.
- **Chet** [10], which implements a global disease co-occurrence graph for disease combination, and constructs dynamic subgraphs for each visit to leverage global and local contexts to predict health events.

D. Performance Evaluation

TABLE. II reports the experimental results of nine baselines and our proposed model on two real-world datasets. It can be observed that DHCE achieves the best performance across both datasets in terms of the three metrics consistently, which ascertains the effectiveness of our proposed method. According to the table, we have several observations.

First, RNN-based methods (e.g., RETAIN, Dipole) outperform CNN-based method (e.g., Deepr). The CNN-based methods can only take into account fixed-size contextual information and cannot adequately capture global, long-range dependencies, while RNN-based methods take the advantage of inherently modeling sequential data, which is able to capture temporal dependencies in sequences.

Second, RNN-based methods with attention mechanism (e.g., Timeline, HiTANet)) outperform RNN-based methods without attention mechanism. This may be caused by the fact that different visit and clinical variables usually play different roles for the final prediction. It is reasonable that selecting important features based on their importance leads to better prediction performance.

Third, graph-based (e.g., GRAM, CGL, and Chet) methods outperform RNN-based methods. This is probably because the existing RNN-based SDP models always simply aggregate the representations of diagnoses in a visit as the representation of the visit, which are unable to effectively model the relations among diagnoses within a visit. This results in the inferiority of RNN-based models compared to GNN-based models.

Fourth, our DHCE model consistently outperforms all compared baseline methods. While Chet considers the clinical relationships among diagnoses within a visit, it does not account for fine-grained higher-order information among diseases, nor does it fully utilize patient-related clinical events in EHRs. Similarly, while an RNN-based approachs with attention can select important clinical features, it fails to effectively utilize patient-related clinical events and model fine-grained higher-order information between diagnoses in a visit. Thus, our approach can beat these methods.

TABLE III: Performance of ablations on MIMIC-III dataset.

Method	w-F_1	R@10	R@20
DHCE	24.24	29.53	39.03
DHCE w/o hypergraph	22.98	27.78	37.17
DHCE w/o transition	22.71	27.52	36.80
DHCE w/o event	23.37	27.79	37.07

E. Ablation Study

We conduct an ablation study on MIMIC-III dataset to evaluate the effectivenss of each key component of our DHCE model. We design four ablations, as below:

- *DHCE w/o hypergraph*: It removes hypergraph learning from the dynamic hypergraph learning module. we replace the hypergraph with a graph of pairwise relationships among nodes.
- *DHCE w/o transition*: It removes transition function from the dynamic hypergraph learning module, but retains the local diagnosis hypergraph and diagnosis-emerging disease hypergraph. We combine the local context in Equ. (2) and the emerging context in Equ. (1) as the visit representation.
- *DHCE w/o event*: It removes clinical events representation module from the standard DHCE model.

Results of the ablations are shown in TABLE III, and we have the following observations:

First, *DHCE w/o hypergraph* leads to a huge performance degradation on both datasets. That is probably because hypergraphs can capture higher-order interaction information among diseases, which is helpful to predict next diagnosis.

Second, *DHCE w/o transition* performs poorly on both datasets. That is probably because the model fails to capture dynamic evolution features of diseases, but simply aggregates the representations of the disease in a visit. This suggests that modeling the dynamic transitional relations benefits the improvement of SDP task.

Third, *DHCE w/o event* does not perform well on both datasets. That is because clinical events that appear in a visit (i.e., *lab test*, *prescription*, and *procedure*) can indicate the physical condition of a patient, which is essential for SDP task.

F. Number of Hypergraph Layers

According to the model structure described in Section III-B, to stack more hypergraph layers is an effective way to aggregate more information from neighboring nodes. However, stacking too many hypergraph layers will result in performance degradation because of over-smoothing. In order to investigate the influence of hypergraph layers of our model, we conduct an experiment on MIMIC-III dataset, as shown in Fig. 2. According to the results in the figure, DHCE is not very sensitive to the number of layers. A 2-layer setting is the best.

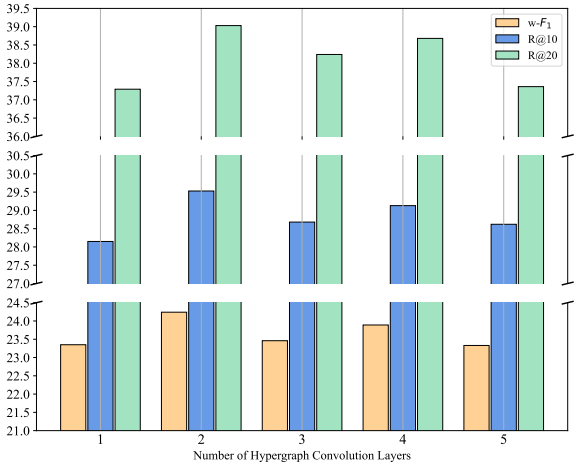

Fig. 2: Performance comparison of DHCE on MIMIC-III dataset by different numbers of hypergraph layers.

V. CONCLUSION

This paper studies the problems of the existing works on sequential diagnosis prediction (SDP), which neglect to exploit the fine-grained higher-order relations among persisitent and emerging diseases in a visit, and fail to fully utilize clinical event information in EHRs. Aiming at the problems, this paper propose a novel model for SDP with dynamic hypergraph and clinical event. Specially, it first categories diagnoses into persistent diseases and emerging ones, and constructs dynamic hypergraphs to capture higher-order disease reltions for each visit, followed by transition functions and attention mechanisms to otain the visit representation. Besides, it employs medical pre-trained language models to encode the textual information of clinical events in a visit and aggregate the representations of all events as the event representation of the current visit. Comprehensive experiments demonstrate that the proposed model show a significant superiority over the state-of-the-art baselines over two benchmark datasets consistently.

ACKNOWLEDGMENT

This work was supported by the National Key R&D Program of China (Grant No.2022YFC3302101) and the National Natural Science Foundation of China (Grant No.62376130), Nature Science Foundation of Shandong (Grant No.ZR2022MF243), Program of Science and Technology of Qilu University of Technology (Grant No.2021JC02010).

REFERENCES

[1] Z. Qiao, S. Zhao, C. Xiao, X. Li, Y. Qin, and F. Wang, "Pairwise-Ranking based Collaborative Recurrent Neural Networks for Clinical Event Prediction," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018, pp. 3520–3526.

[2] X. Peng, G. Long, T. Shen, S. Wang, J. Jiang, and M. Blumenstein, "Temporal Self-Attention Network for Medical Concept Embedding," in *Proceedings of the 19th IEEE International Conference on Data Mining (ICDM)* , 2019, pp. 498–507.

[3] Y. An, K. Tang, and J. Wang, "Time-Aware Multi-Type Data Fusion Representation Learning Framework for Risk Prediction of Cardiovascular Diseases," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, no. 6, pp. 3725–3734, 2021.

[4] H. Ren, J. Wang, and W. X. Zhao, "Generative Adversarial Networks Enhanced Pre-training for Insufficient Electronic Health Records Modeling," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2022, pp. 3810–3818.

[5] A. Caruana, M. Bandara, D. Catchpoole, and P. J. Kennedy, "Beyond Topics: Discovering Latent Healthcare Objectives from Event Sequences," in *Proceedings of the 35th Australasian Joint Conference on Artificial Intelligence (AJCAI)*, 2022, pp. 368–380.

[6] K. Niu, Z. Guo, X. Peng, and S. Pei, "P-ResUnet: Segmentation of brain tissue with Purified Residual Unet," *Computers in Biology and Medicine*, vol. 151, no. Pt B, pp. 106 294–106 305, 2022.

[7] A. N. Jagannatha and H. Yu, "Structured prediction models for RNN based sequence labeling in clinical text," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2016, pp. 856–865.

[8] T. Bai, S. Zhang, B. L. Egleston, and S. Vucetic, "Interpretable Representation Learning for Healthcare via Capturing Disease Progression through Time," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2018, pp. 43–51.

[9] C. Lu, C. K. Reddy, P. Chakraborty, S. Kleinberg, and Y. Ning, "Collaborative Graph Learning with Auxiliary Text for Temporal Event Prediction in Healthcare," *ArXiv Preprint ArXiv:2105.07542*, 2021.

[10] C. Lu, T. Han, and Y. Ning, "Context-Aware Health Event Prediction via Transition Functions on Dynamic Disease Graphs," in *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, vol. 36, no. 4, 2022, pp. 4567–4574.

[11] M. Ye, S. Cui, Y. Wang, J. Luo, C. Xiao, and F. Ma, "MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths," in *in Proceedings of th 30th World Wide Web Conference (WWW)*, 2021, pp. 1397–1409.

[12] Z. Jin, Y. Wang, Q. Wang, Y. Ming, T. Ma, and H. Qu, "GNNLens: A Visual Analytics Approach for Prediction Error Diagnosis of Graph Neural Networks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 6, pp. 3024–3038, 2023.

[13] A. Kazi, L. Cosmo, S.-A. Ahmadi, N. Navab, and M. M. Bronstein, "Differentiable Graph Module (DGM) for Graph Convolutional Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1606–1617, 2022.

[14] J. Luo, M. Ye, C. Xiao, and F. Ma, "HiTANet: Hierarchical Time-Aware Attention Networks for Risk Prediction on Electronic Health Records," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2020, pp. 647–656.

[15] L. Ma, C. Zhang, Y. Wang, W. Ruan, J. Wang, W. Tang, X. Ma, X. Gao, and J. Gao, "ConCare: Personalized Clinical Feature Embedding via Capturing the Healthcare Context," in *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, vol. 34, no. 01, 2020, pp. 833–840.

[16] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao, "Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2017, pp. 1903–1911.

[17] E. Choi, Z. Xu, Y. Li, M. Dusenberry, G. Flores, E. Xue, and A. Dai, "Learning the Graphical Structure of Electronic Health Records with Graph Convolutional Transformer," in *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, vol. 34, no. 01, 2020, pp. 606–613.

[18] Y. Xu, H. Ying, S. Qian, F. Zhuang, X. Zhang, D. Wang, J. Wu, and H. Xiong, "Time-Aware Context-Gated Graph Attention Network for Clinical Risk Prediction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 07, pp. 7557–7568, 2023.

[19] Y. Su, Y. Zhao, S. Erfani, J. Gan, and R. Zhang, "Detecting Arbitrary Order Beneficial Feature Interactions for Recommender Systems," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2022, pp. 1676–1686.

[20] Y. Ma, R. Mao, Q. Lin, P. Wu, and E. Cambria, "Multi-source aggregated classification for stock price movement prediction," *Information Fusion*, vol. 91, pp. 515–528, 2023.

[21] Q. Lin, R. Mao, J. Liu, F. Xu, and E. Cambria, "Fusing topology contexts and logical rules in language models for knowledge graph completion," *Information Fusion*, vol. 90, pp. 253–264, 2023.

[22] J. Jiang, Y. Wei, Y. Feng, J. Cao, and Y. Gao, "Dynamic HyperGraph Neural Networks," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 2019, pp. 2635–2641.

[23] R. Xu, Y. Yu, C. Zhang, M. Ali, J. Ho, and C. Yang, "Counterfactual and Factual Reasoning over Hypergraphs for Interpretable Clinical Predictions on EHR." *in Proceedings of the 2nd Machine Learning for Health (ML4H)*, vol. 193, pp. 259–278, 2022.

[24] D. Cai, C. Sun, M. Song, B. Zhang, S. Hong, and H. Li, "Hypergraph Contrastive Learning for Electronic Health Records," in *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, 2022, pp. 127–135.

[25] K. Hur, J. Oh, J. Kim, J. Kim, M. J. Lee, E. Cho, S.-E. Moon, Y.-H. Kim, and E. Choi, "UniHPF: Universal Healthcare Predictive Framework with Zero Domain Knowledge," *arXiv preprint arXiv:2211.08082*, 2022.

[26] E. Alsentzer, J. Murphy, W. Boag, W.-H. Weng, D. Jindi, T. Naumann, and M. McDermott, "Publicly Available Clinical BERT Embeddings," in *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, 2019, pp. 72–78.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," *in Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[28] Y. Tan, C. Yang, X. Wei, C. Chen, W. Liu, L. Li, J. Zhou, and X. Zheng, "MetaCare++: Meta-Learning with Hierarchical Subtyping for Cold-Start Diagnosis Prediction in Healthcare Data," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 2022, pp. 449–459.

[29] A. E. 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, pp. 1–9, 2016.

[30] A. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, B. Moody, B. Gow, L.-w. H. Lehman *et al.*, "MIMIC-IV, a freely accessible electronic health record dataset." *Scientific Data*, vol. 10, no. 1, 2023.

[31] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, "RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism," *in Proceedings of the 30th International Conference on Neural Information Processing Systems (NeurIPS)*, vol. 29, 2016.

[32] P. Nguyen, T. Tran, N. Wickramasinghe, and S. Venkatesh, "Deepr: A Convolutional Net for Medical Records," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 22–30, 2017.

[33] 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 23th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2017, pp. 787–795.

[34] J. Shang, T. Ma, C. Xiao, and J. Sun, "Pre-training of Graph Augmented Transformers for Medication Recommendation," in *Proceedings of 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 2019, pp. 5953–5959.