“© 2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Multi-Channel Hypergraph Network for Sequential Diagnosis Prediction in Healthcare

Xin Zhang[1,4], Xueping Peng[2], Weimin Chen[3], Weiyu Zhang[1,4], Xiaoqiang Ren[1,4(✉)] and Wenpeng Lu[1,4(✉)]
[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 Mental Health Center, Jinan, China
[4] Shandong Provincial Key Laboratory of Computer Networks,
Shandong Fundamental Research Center for Computer Science, Jinan, China
zhangxinqlu@foxmail.com, xueping.peng@uts.edu.au,
cwm1901@163.com, {zwy}@qlu.edu.cn
(✉) Corresponding author email: {renxq, wenpeng.lu}@qlu.edu.cn

***Abstract*—Sequential diagnosis prediction (SDP) is a complex and challenging task, aming to predict future diagnoses of patients by analyzing their historical medical records. Although graph neural networks(GNNs) has been applied to successfully address the challenge of heterogeneous data integration in electronic health records, relatively limited work has been done on GNNs for sequential diagnosis prediction. Graph neural network-based methods, aimed at capturing structural and relational patterns of EHR data for sequential diagnosis prediction, explore code-code pairwise relationships, resulting in an inability to learn fine-grained, higher-order interaction relationships among different types medical codes. As a result, they are difficult to effectively model complex, multi-dimensional interactions among different types of medical codes necessary for accurate and nuanced diagnosis predictions. To address these challenges, this paper proposes a novel approach called Multi-Channel Hypergraph Network (MCHN) predictive framework for sequential diagnosis prediction. The proposed method aims to explore the fine-grained higher-order interactions between different types of medical codes via multi-channel hypergraphs. Specifically, MCHN learns two levels of code embeddings from multi-channel hypergraph learning module and line graph learning module, respectively: (i) multi-channel hypergraph learning module, which is to learn multi-channel hypergraph level code embeddings by modeling the higher-order relationships between medical codes in different hypergraphs; and (ii) line graph learning module, which is to learn the line graph level code embeddings by modeling code-code pairwise relationships. In MCHN, we propose a novel channel-level attention mechanism to help our model attend to the informativeness of the different channel for forecasting future patient diagnoses. We also design a code-level attention mechanism, which can to pay more attention to the medical codes that are more important to the visit representation. Moreover, MCHN aggregates the learnt code embeddings in the two levels to generate the visit representation, which is used to predict the patient's next diagnosis. Experimental results on two benchmark datasets consistently demonstrate that MCHN outperforms state-of-the-art methods [1].**

***Index Terms*—sequential diagnosis prediction, graph neural network, fine-grained higher-order interactions**

I. Introduction

Electronic Health Record (EHR) data a wealth of information about patients' status, such as diagnoses and medications, and play a crucial role in identifying patient patterns and supporting clinical decision-making [1]–[8]. Most of the research efforts in this area regard the patient visit records as ordered sequences, among which recurrent neural networks (RNNs) based and graph neural networks (GNNs) based approaches have shown great performance.

In RNNs-based approaches, modeling a patient's historical visits as a strictly-ordered sequence is deemed as the key to success, since the intrinsic chronological order of the patient's historical visits [9]–[13]. However, it is worth noting that multiple diseases can appear in a single visit, with complex clinical relationships between them. For example, the patient may be simultaneously diagnosed with *diabetes*, *high blood pressure* and *cardiovascular disease* in a visit. The *diabetes* can aggravate the risk of *heart disease* and *hypertension*. Furthermore, a single visit may contain different types of medical codes, including diagnosis and medication codes. While diagnosis codes might have a direct impact on predicting next diagnosis of the patient, medications may provide less predictive information. 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 medical codes within a single visit*. Consequently, RNN-based models exhibit limited performance on SDP tasks.

Recently, the effectiveness of graph neural networks has been reported in many areas, including sequential diagnosis prediction. Unlike the RNNs-based SDP method, the GNNs-based approaches model the medical codes in a visit as a pairwise graph. The different types of medical codes may have different informativeness for representing patient, and there is also an implicit relationship within the same type of medical codes in a visit. For instance, a doctor can intuitively predict the patient's next visit based on the patient's history of diagnoses. Therefore, diagnoses in a visit might have a direct impact on predicting diagnoses appearing in next visit of the patient. Additionally, there are co-occurring relationships among medical codes. For example, in clinical practice, it is usual with one or more medications to treat a disease. It is essential the types of medical codes and the co-occurrence relationship among diagnoses and medicines in more finegrained 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 different types of medical codes within one visit.* This paper addresses these issues by proposing *Multi-Channel Hypergraph Network (MCHN) predictive framework for sequential diagnosis prediction.* First, Unlike the traditional graph neural networks such as GCN that model diagnoses in a visit as a pairwise graph to capture the code-code pairwise relations among medical codes. To model he sophisticated information between different types of medical codes in more fine-grained granularity, we constructed a multi-channel hypergraph based on the different types of medical codes and the co-occurrence relationship among medical codes in a visit. Specifically, we constructed a diagnosis hypergraph from diagnoses in a visit for the diagnosis channel and a medicine hypergraph from medicines in a visit for the medicine channel. In addition, we constructed a diagnosis-medicine co-

[1]Our source code is avaliable at https://github.com/HELJJ/MCHN.

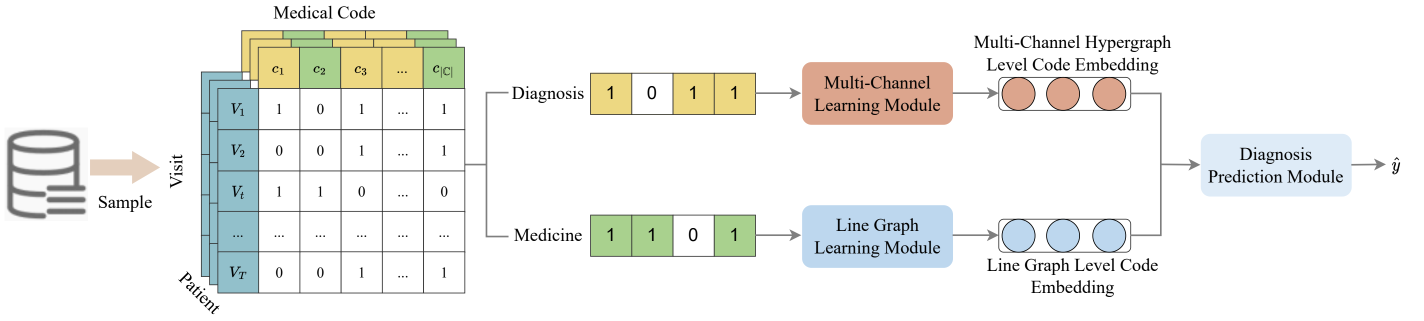

Fig. 1: The architecture of overall MCHN framework. It includes a multi-channel hypergraph learning module, a line graph learning module, and a diagnosis prediction module. The multi-channel hypergraph learning module learn multi-channel hypergraph level code embeddings by modeling the higher-order relationships between medical codes in different hypergraphs. The line graph learning module learn the line graph level code embeddings by modeling code-code pairwise relationships. The diagnosis prediction module fuses the multi-channel hypergraph level code embeddings and line graph level code embeddings to generate the visit representation, which is utilized to predict the future diagnoses.

occurrence hypergraph for the diagnosis-medicine co-occurrence channel based on the co-occurrence relationship between each diagnosis and medicines in a visit. In MCHN, we propose to learn two levels of code embeddings from multi-channel hypergraph learning module and line graph learning module, respectively: (i) multi-channel hypergraph learning module, which is to learn multi-channel hypergraph level code embeddings by modeling the higher-order relationships between medical codes in different channel hypergraphs; and (ii) line graph learning module, which is to learn the line graph level code embeddings by modeling code-code pairwise relationships. In MCHN, we propose a novel channel-level attention mechanism to help our model attend to the informativeness of the different channel for forecasting future patient diagnoses. We also design a code-level attention mechanism, which can learn to pay more attention to the medical codes that are more important to the patient representation. Moreover, MCHN aggregates the learnt code representations in the two levels. The main contributions of this work are summarized as follows:

- We propose a novel multi-channel hypergraph network for sequential diagnosis prediction. In contrast to the existing SDP models.The MCHN captures both fine-grained higher-order interactions and code-code pairwise relationships among different types of medical codes in a visit.
- We propose a unified model to improve the sequential diagnosis prediction performance by effectively leveraging two levels of code embeddings from multi-channel hypergraph learning module and line graph learning module.
- We also propose a novel channel-level attention mechanism to incorporate the code information in different channel, which shows the superiority performance for sequential diagnosis prediction.

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

II. 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 MCHN model, including the multi-channel hypergraph learning module, line graph learning module and diagnosis prediction module.

A. Notations and Problem Statement

1) Notations: The data utilized in this study from EHR (Electronic Health Records) encompasses different types of medical codes, including ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) for diagnosis and NDC (National Drug Code) for medicine.

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 $V_t^u = \{c_1, c_2, \cdots, c_{|\mathbb{C}|}\}$ is recorded with a subset of medical codes $C_t^u \subset \mathbb{C}$. C_t^u includes not only diagnosis codes $D_t^u = \{d_1, d_2, \cdots, d_{|\mathbb{D}|}\}$ but also medicine codes $M_t^u = \{m_1, m_2, \cdots, m_{|\mathbb{M}|}\}$, where $|\mathbb{D}|$, $|\mathbb{M}|$ respectively denote the number of diagnosis and medicine codes in V_t^u.

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

we construct multichannel hypergraphs to capture higher-order interactions between different types of medical codes in a visit. Specifically, we build a diagnosis hypergraph for the diagnosis channel by defining the diagnoses that appear in a visit as nodes and the visit as a hyperedge. Similarly, we construct a medicine hypergraph for the medicine channel by defining the medicines that appear in a visit as nodes and the visit as a hyperedge. Likewise, for the diagnosis-medicine co-occurrence channel, we construct a diagnosis-medicine co-occurrence hypergraph by defining the diagnoses and medicines that appear in a visit as nodes and the co-occurrence relationship between the diagnoses and medicines as the edge.

For each visit of a patient, we constructed a diagnosis hypergraph from diagnoses in a visit for the diagnosis channel and a medicine hypergraph from medicines in the visit for the medicine channel. In addition, we constructed a diagnosis-medicine co-occurrence hypergraph for the diagnosis-medicine co-occurrence channel based on the co-occurrence relationship between each diagnosis and medicines in the visit. For example, for the t-th visit of a patient, we construct a diagnosis hypergraph for diagnosis channel, a medicine hypergraph for medicine channel, and a diagnosis-medicine co-occurrence hypergraph for diagnosis-medicine co-occurrence channel, where $\mathcal{G}_d^t = (\mathbb{P}^t, \mathbb{F}^t)$, $\mathcal{G}_m^t = (\mathbb{H}^t, \mathbb{E}^t)$, and $\mathcal{G}_{dm}^t = (\mathbb{X}^t, \mathbb{K}^t)$ respectively denote the diagnosis hypergraph, the medication hypergraph, and the diagnosis-medication co-occurrence hypergraph of the visit t-th. $\mathbb{P}^t \subset \mathbb{C}$ represents the set of all nodes in $\mathcal{G}_d^t$, that is, the set of all diagnoses of the visit t. Similarly, $\mathbb{F}^t$ denotes the hyperedge (i.e. t-th visit) in $\mathcal{G}_d^t$. Likewise, for the medication channel hypergraph

$\mathcal{G}_m^t$, $\mathbb{H}^t$ represents the set of all medications of the t-th visit, and $\mathbb{E}^t$ represents the hyperedge in $\mathcal{G}_{dm}^t$. $\mathbb{X}^t$ represents the set of all medical codes of the t-th visit, and $\mathbb{K}^t$ denotes the hyperedge in $\mathcal{G}_{dm}^t$.

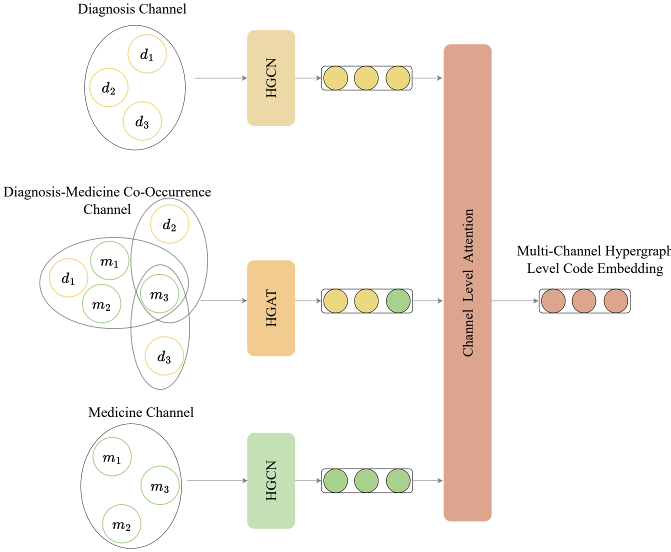

Fig. 2: The Multi-Channel Hypergraph Learning Module in MCHN framework.

C. Multi-Channel Hypergraph Learning Module

Since there are different types of medical codes in a visit, including diagnosis and medicine, we first construct a multi-channel hypergraph, then the multi-channel hypergraph learning module is used to learn multi-channel hypergraph level code embeddings from multi-channel hypergraph. As shown in Figure 2, there are two sub-modules in the multi-channel hypergraph learning module.

The first one is hypergraph neural network(HGNN) [14], HGNN is an effective neural architecture for capturing high order interaction information between medicical codes. The patient may be simultaneously diagnosed with *diabetes*, *high blood pressure* and *cardiovascular disease* in a visit. First, *diabetes* can aggravate the risk of *heart disease* and *hypertension*, and in addition, *hypertension* may lead to worsening of *heart disease* as well as is a common cause of *diabetic complications*. Therefore, we apply hypergraph neural networks to learn the interactions between multiple diseases by capturing their higher-order relationships.

For the diagnosis channel and the medicine channel, we use a hypergraph convolutional network [15] to encode the diagnosis hypergraph $\mathcal{G}_d^t$ and the medicine hypergraph $\mathcal{G}_m^t$ to obtain the diagnosis context $\mathbf{Z}_d^t$ and the medicine context $\mathbf{Z}_m^t$, respectively, described as:

$$\mathbf{Z}_d^t = HyperGCN(\mathcal{G}_d^t), \tag{1}$$

$$\mathbf{Z}_m^t = HyperGCN(\mathcal{G}_m^t), \tag{2}$$

where $HyperGCN(\cdot)$ denotes hypergraph convolutional networks.

For the diagnosis-medicine co-occurrence channel, in order to explore the interactions and combinations among diagnoses and medicines, we use the hypergraph attention network to encode the diagnosis-medicine co-occurrence graph $\mathcal{G}_{dm}^t$ to obtain the diagnosis-medicine co-occurrence context $\mathbf{Z}_{dm}^t$, as below:

$$\mathbf{Z}_{dm}^t = HyperGAT(\mathcal{G}_{dm}^t), \tag{3}$$

here, $HyperGAT(\cdot)$ denotes hypergraph attention networks.

The second one is channel-level attention network, where different channels have different information for characterizing the features of the patient. For example, among the diagnosis channel, the medicine channel, and the diagnosis-medicine channel, it is possible that the diagnosis channel makes is very informative for learning about the patient's representation, since it conveys important clues about the prediction of future visits, whereas the medicine channel may be less informative for prediction.

In this module, we propose the channel-level attention mechanism to help our model capture to capture relevant information from different channels, i.e., diagnosis channel, medicine channel and diagnosis-medicine co-occurrence channel. Taking the diagnosis channel as an example, which is formulated as:

$$\boldsymbol{\alpha}_d = \text{softmax}\left(\mathbf{Z}_d^t \mathbf{W}_d\right), \tag{4}$$

$$\mathbf{X}_d = \boldsymbol{\alpha}_d \mathbf{Z}_d^t, \tag{5}$$

According to the above equations, we first construct the attention weight of the diagnosis channel $\boldsymbol{\alpha}_d$, as shown in Equ. (4). Here, $\mathbf{W}_d$ represents trainable parameters and $\mathbf{Z}_d^t$ the diagnosis context. Subsequently, the feature representation of the diagnosis channel $\mathbf{X}_d$ is obtained with Equ. (5).

Similar with the operations on diagnosis channel, we can obtain the feature representation for medicine channel as well as diagnosis-medicien channel, denoted by $\mathbf{X}_m$ and $\mathbf{X}_{dm}$ respectively.

Given that both the diagnosis channel and the diagnosis-medicine co-occurrence channel contain the diagnoses in the t-th visit of the patient, and that both the medicine channel and the diagnosis-medicine channel contain the medicines in the same visit. Therefore, we integrate the diagnosis-related representations in the feature representation of the diagnosis channel and feature representation of the diagnosis-medicine co-occurrence channel to obtain the hidden representations of diagnosis $\mathbf{X}_d^{hy}$ in the visit t,

$$\mathbf{X}_d^{hy} = \left[\text{MLP}_L\left(\mathbf{X}_{dd}, \mathbf{X}_{ddm}\right)\right], \tag{6}$$

where MLP_L denotes Multi-Layer Perceptron. $\mathbf{X}_{dd}$, $\mathbf{X}_{ddm}$ represent the representations of diagnosis in $\mathbf{X}_d$ and $\mathbf{X}_{dm}$, respectively.

Similarly, we integrate the medicine-related representations in the feature representation of the medicine channel and diagnosis-medicine co-occurrence channel to obtain the hidden representation of medicine $\mathbf{X}_m^{hy}$,

$$\mathbf{X}_m^{hy} = \left[\text{MLP}_L\left(\mathbf{X}_{mm}, \mathbf{X}_{md}\right)\right], \tag{7}$$

where $\mathbf{X}_{mm}$, $\mathbf{X}_{md}$ represent the representations of medicine in $\mathbf{X}_m$ and $\mathbf{X}_{dm}$, respectively.

Finally, the hidden representation of diagnosis $\boldsymbol{X}_d^{hy}$ and medicine codes $\boldsymbol{X}_m^{hy}$ are aggregated to form multi-channel hypergraph level code embeddings $\mathbf{E}_{C_h}$, as below:

$$\mathbf{E}_{C_h}^t = Concat(\mathbf{X}_d^{hy}; \mathbf{X}_m^{hy}). \tag{8}$$

Here, $\text{Concat}(\cdot)$ represents concatenate operation.

D. Line Graph Learning Module

Let $\mathcal{G}_l = (C_t^u, \mathbf{A}_t^u)$ denotes the line graph. C_t^u is the medical codes in visit t. $\mathbf{A}_t^u$ is the adjacency matrix. The edge weights of the line graph $\mathcal{G}_l$ should represent the correlation among medical codes and it can be learned from data via Transformer [16].

The line graph learning module in our approach aims to learn code-code pairwise representation. In this module, a transformer module is used to build informative medical code representations. The Transformer model consists of multi-head self-attention layers, point-wise feed-forward layers, and layer normalization. Note that we do not need positional encoding here. The self-attention is obtained via

$$\text{Attn}(\mathbf{K}, \mathbf{Q}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V}, \tag{9}$$

$$\text{Self-Attention}\left(\mathbf{S}^{(l)}\right) = \text{Attn}\left(\mathbf{S}^{(l)}\mathbf{W}_K, \mathbf{S}^{(l)}\mathbf{W}_Q, \mathbf{S}^{(l)}\mathbf{W}_V\right) \tag{10}$$

where $\mathbf{S}^{(l)}$ is the input of the l-th layer, $\mathbf{W}_K$, $\mathbf{W}_Q$, $\mathbf{W}_V$ are trainable parameters and d is the column size of $\mathbf{W}_K$. Transformers can learn the underlying structure of the medical code graph since the attention weights reflect the strength of pairwise connections between codes. The output of L_t layer Transformer are aggregated into line graph level code embeddings $\mathbf{E}^t_{C_l}$ for visit t.

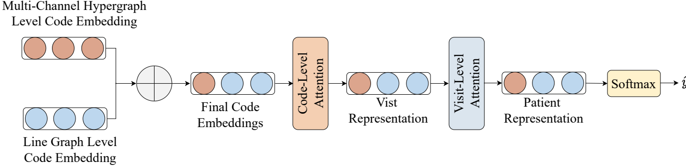

Fig. 3: The diagnosis prediction module in MCHN framework.

E. Diagnosis Prediction Module

The diagnosis prediction module is used to predict the future diagnose. As shown in Figure 3, There are two components in the diagnosis prediction module: patient representation and diagnosis prediction.

patient representation. We first obtain the final code embeddings for visit t via information integration. Subsequently, We use code-level attention to capture the importance of each medical code to visit representation of visit t. Finally, we employ visit-level attention to aggregate the representations of all visits into a comprehensive patient representation.

The information integration is used to integrate multi-channel hypergraph level code embeddings from the multi-channel hypergraph learning module and linear graph level code embeddings from the linear graph to obtain the final code embeddings, described as:

$$\mathbf{E}^t_C = \text{Concat}(\mathbf{E}^t_{C_h}, \mathbf{E}^t_{C_l}), \tag{11}$$

where $\mathbf{E}^t_C$ is the final code embeddings of visit t, $\text{Concat}(\cdot)$ represents concatenate operation.

Usually multiple medical codes are appearing in a single visit and different types of medical codes have different informativeness for learning visit representation. Therefore, we utilize code-level attention to learn the importance of each medical code in visit t for visit representation of visit t, as below:

$$\boldsymbol{\alpha}_c = \text{softmax}\left(\left[\mathbf{E}^t_{C_1}, \mathbf{E}^t_{C_2}, \ldots, \mathbf{E}^t_{C_i}\right]\mathbf{W}_c\right), \tag{12}$$

$$\mathbf{E}^t_{C_i} = \boldsymbol{\alpha}_c\left[\mathbf{E}^t_{C_1}, \mathbf{E}^t_{C_2}, \ldots, \mathbf{E}^t_{C_i}\right]^\top, \tag{13}$$

Here, $\mathbf{W}_c$ is a context vector for attention. α_c is the attention score for medical codes in visit t. $\mathbf{E}^t_{C_i}$ represents embedding of i-th medical code in visit t.

Finally, we apply a visit-level attention, to calculate the representations o of all visits, i.e., patient embedding:

$$\boldsymbol{\alpha} = \text{softmax}\left(\left[\mathbf{E}^1_v, \mathbf{E}^2_v, \ldots, \mathbf{E}^T_v\right]\mathbf{W}_v\right), \tag{14}$$

$$\mathbf{u} = \boldsymbol{\alpha}\left[\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^T\right]^\top, \tag{15}$$

where $\mathbf{W}_v$ is a context vector for attention and α is the attention score for visits. The patient embedding $\mathbf{u}$ will be used in a classifier for final prediction of future visits.

diagnosis representation. Given a patient's visit records denoted as $\gamma_u = (V^u_1, V^u_2, \cdots, V^u_T)$, we perform the sequential diagnostic prediction task to utilize the EHR's chronological visit data. The aim is to forecast the diagnoses for the next visit of the patient, as described in Equ. (16). We employ the cross-entropy loss in Equ. (17) to optimize the model [2], [17].

$$\hat{y} = \text{softmax}(\mathbf{W}_y\mathbf{u} + \mathbf{b}_y), \tag{16}$$

$$\mathcal{L} = \frac{1}{T+1}\sum_{t=2}^{T} -(y^T_{true}log\hat{y} + (1 - y_{true})^T log(1 - \hat{y})). \tag{17}$$

Here, $\mathcal{L}$ denotes the loss function, $\mathbf{u}$ is the patient representation, $\mathbf{W}_y$ 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.

III. EXPERIMENT

A. Datesets and Preprocessing

we conduct experiments on two benchmark electronic health record datasets: MIMIC-III [18] and MIMIC-IV [19], both developed by the MIT Laboratory of Computational Physiology. MIMIC-III encompasses medical data for about 40,000 patients spanning 2001 to 2012. MIMIC-IV, the more recent dataset iteration, includes data for roughly 80,000 patients from 2008 to 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 II. For MIMIC-III, we further split it into three groups: 4,500 patients for training, 399 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 II: Statistics of datasets.

Dataset	MIMIC-III	MIMIC-IV
# patients	1,899	10,000
Max. # visit	42	77
Avg. # visit	4.02	4.19
# codes	3605	5341
Max. # codes per visit	39	39
Avg. # codes per visit	14.81	9.67

B. Baselines

The performance of our proposed MCHN model is compared against different state-of-the-art models as follows.

- **RETAIN** [20], which learns the representations of medical concepts and subsequently utilizes an RNN (Recurrent Neural Network) with a reverse temporal attention mechanism for predicting patient visit information.
- **Dipole** [21], which utilizes a bidirectional RNN combined with three types of attention mechanisms: location-based, general, and concatenation-based. This approach is utilized to predict diagnoses in the subsequent visit.
- **Timeline** [22], which assigns time-decay factors to each visits and learns the impact of both chronic and acute conditions. Furthermore, it employs attention mechanisms to enhance visit representation and improve the accuracy of subsequent diagnosis predictions.
- **HiTANet** [23], which employs a hierarchical time-aware attention network, effectively models temporal information at both local and global stages. This network leverages a time-aware Transformer along with a key-query attention mechanism to construct a comprehensive patient representation for the prediction of subsequent diagnoses.
- **Deepr** [24], learns medical concept embeddings using a convolutional neural network, with the aim of predicting unplanned patient readmissions post-discharge.

TABLE I: 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	17.72	<u>24.74</u>	33.47	24.08	25.71	32.51
RNN-based methods	RETAIN	19.54	23.43	32.23	21.86	27.44	33.04
	Dipole	18.20	22.28	31.17	21.98	26.80	34.06
	Timeline	19.31	23.05	31.98	23.55	28.42	35.71
	HiTANet	20.08	23.32	33.12	23.21	26.87	34.81
Graph-based methods	GRAM	20.37	23.81	32.95	21.79	27.71	34.94
	G-BERT	18.73	23.46	32.46	22.78	26.58	34.44
	CGL	20.77	23.94	33.87	23.70	27.94	35.73
	Chet	<u>22.48</u>	24.27	<u>35.02</u>	<u>25.22</u>	<u>28.52</u>	<u>36.89</u>
	MCHN (our)	**23.92**	**25.22**	**36.94**	**26.15**	**29.03**	**37.71**
	Improvement(%).	↑6.40	↑3.92	↑5.48	↑3.68	↑1.48	↑2.24

- **GRAM** [25], which utilizes graph attention networks, learns representations of a medical code and its ancestral nodes within a knowledge graph. Moreover, it leverages these representations of medical codes to predict the diagnoses that appears in the subsequent visit predict.
- **G-BERT** [26], which utilizes Graph Neural Networks (GNNs) to model the hierarchical structures of medical codes. It integrates these representations of medical codes into a transformer-based encoder, thereby generating visit representations for sequential diagnosis prediction.
- **CGL** [9], which constructs a collaborative graph, connecting patients and diseases based on their co-occurrence. This approach models patient-disease interactions accurately predict clinical events.
- **Chet** [10], which implements a global disease co-occurrence graph to analyze disease combinations. It also constructs dynamic subgraphs for each patient visit, leveraging both global and local contexts to enhance the prediction of health events.

C. Performance Evaluation

The experimental results on sequential diagnosis prediction are summarized in TABLE. I. As shown in Table I, the MCHN model demonstrates outstanding performance across all three metrics on both datasets, which shows that our proposed method is effective. From the table we can see the following:

First, RNN-based methods (e.g., RETAIN, Dipole) demonstrate superior performance compared to CNN-based method (e.g., Deepr). The CNN-based methods are excellent for capturing fixed-size contextual information, but inadequate for modeling global, long-range dependencies. In contrast, RNN-based methods are inherently advantageous for modeling sequential data, enabling them to effectively capture temporal dependencies within sequences.

Secondly, RNN-based methods that incorporate an attention mechanism (e.g., Timeline and HiTANet) outperform those without it. The enhanced performance of RNN-based methods with an attention mechanism might be attributed to the differential impact of different visits and clinical variables on the final prediction. This suggests that it is reasonable to select important features based on their importance to improve the accuracy of the prediction.

Thirdly, graph-based (e.g., GRAM, CGL, and Chet) methods exhibit superior performance over RNN-based methods. This superiority is likely due to the limitations in existing RNN-based SDP models, which simply aggregate the representations of medical code in a visit as the representation of the visit. This approach fails to effectively model the interrelations among medical codes within a visit. Consequently, this leads to the relative inferiority of RNN-based models when compared to those based on Graph Neural Networks (GNNs).

Fourthly, the MCHN model in our study consistently excels beyond all compared baseline methods. While Chet effectively accounts for clinical relationships among diagnoses within a single visit, it falls short in capturing fine-grained higher-order relationships among different types of medical codes. In a similar vein, RNN-based approaches with attention mechanisms are adept at selecting important clinical features, yet they too fall short in capturing these fine-grained higher-order relationships among different types of medical codes.

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

Method	w-F_1	R@10	R@20
MCHN	23.92	25.22	36.94
MCHN w/o diagnosis channel	23.86	23.58	33.66
MCHN w/o medicine channel	26.10	25.69	37.01
MCHN w/o co-currence channel	26.74	27.04	36.27
MCHN w/o channel-level attention	26.51	26.62	36.83

D. Ablation Study

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

- *MCHN w/o diagnosis channel*: It removes diagnosis channel in multi-channel hypergraph learning module.
- *MCHN w/o medicine channel*: It removes medicine channel in multi-channel hypergraph learning module.
- *MCHN w/o co-currence channel*: It removes diagnosis-medicine co-ocurrence in multi-channel hypergraph learning module. We combine the diagnosis context in Eq. 1 with the medicine context in Eq. 2 as the multi-channel hypergraph level code embeddings.
- *MCHN w/o channel-level attention*: It removes channel-level attention in multi-channel hypergraph learning module， The diagnosis context in Equ. (1), the medicine context in Equ. (2),and the diagnosis-medicine co-occurrence context in Equ (3) are concatenated together according to the first dimension.

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

First, *MCHN w/o diagnosis channel* leads to a huge performance degradation on MIMIC-III datasets. This is probably because higher-order interaction information between daignoses in a visit can be captured and diagnoses have a direct impact on predicting next diagnosis of the patient.

Second, *MCHN w/o medicine channel* performs poorly for F_1 metrics, but performs slightly better for R@20 on the MIMIC-

III datasets. This may be because the medications at a visit are helpful for patient representation, but it is possible that some of the side effects between the medications may affect the effectiveness of F_1.

Third, *MCHN w/o co-currence channel* slightly worse on MIMIC-III datasets. That is because the co-occurrence of each diagnosis with the medicines is beneficial, which is essential for SDP task.

fourth, *MCHN w/o channel-level attention* does not perform well on MIMIC-III datasets. That is because the informativeness of different channels is different for patient representation, channel-level attention can adaptively incorporate information from different channels benefits the improvement of SDP task.

IV. CONCLUSION

This paper studies the problems of the existing works on sequential diagnosis prediction, which neglect to exploit the fine-grained higher-order interactions between different types of medical codes in a visit. Aiming at the problems, this paper propose a novel model for SDP with multi-channel hypergraph network. Specially, it first converts the medical record into multi-channel hypergraph and line graph. The multi-channel hypergraph level code embeddings learned from the multi-channel hypergraph and the line graph level code embeddings learned from the line graph are subsequently combined to obtain final code embeddings. Next, we use code-level attention to learn the contribution of each code to the visit representation. Finally, all the visit representations are combined to get the patient representation. Comprehensive experiments demonstrate that the proposed model show a significant superiority over the state-of-the-art baselines over two benchmark datasets consistently.

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] X. Peng, G. Long, T. Shen, S. Wang, and J. Jiang, "Sequential diagnosis prediction with transformer and ontological representation," in *Proceedings of the 21st IEEE International Conference on Data Mining (ICDM)*, 2021, pp. 489–498.

[5] 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.

[6] 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.

[7] 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.

[8] S. Wei, W. Lu, X. Peng, S. Wang, Y.-F. Wang, and W. Zhang, "Medical question summarization with entity-driven contrastive learning," *arXiv preprint arXiv:2304.07437*, 2023.

[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] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," in *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, vol. 33, no. 01, 2019, pp. 3558–3565.

[15] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui, and X. Zhang, "Self-supervised hypergraph convolutional networks for session-based recommendation," in *Proceedings of the 35th AAAI conference on artificial intelligence*, vol. 35, no. 5, 2021, pp. 4503–4511.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.