HHGCN-DrugRec: Hierarchical HyperGraph Convolution Network for Drug Combination Recommendation

Zihan Zhang[1], Hongzhi Liu[1(✉)], Tianqi Sun[1], and Xiaoshuang Guo[1], Zhonghai Wu[2(✉)]

[1] School of Software and Microelectronics, Peking University, China
lusia@stu.pku.edu.cn,liuhz@pku.edu.cn,{suntq,gxs}@stu.pku.edu.cn
[2] National Engineering Center of Software Engineering, Peking University, China
wuzh@pku.edu.cn

Abstract. Drug combination recommendation aims to provide personalized medication prescriptions for patients according to their historical visit records and current condition. Several recent studies try to utilize graph structures to better model the relationships among diagnoses, procedures, and drugs for drug combination recommendation. However, they primarily focus on pairwise relationships and fail to capture the higher-order connections and combination effects among different medical entities. Additionally, existing approaches often overlook the multifaceted nature of diagnoses and procedures within patients' medical records. To address these issues, we propose a hierarchical hypergraph convolution network for drug combination recommendation. Specifically, we design an innovative hierarchical hypergraph structure and corresponding convolution network to capture the complex, higher-order relationships among medical entities. Furthermore, we design a capsule-based representation learning method to model the multifaceted nature of patients' medical records for drug combination recommendation. Extensive experiments on several real-world datasets of different hospital departments demonstrate the effectiveness of the proposed approach compared with state-of-the-art methods.

Keywords: Drug combination recommendation · Recommender system · Hierarchical hypergraph · Capsule network

1 Introduction

Recent advances in personalized medicine make it possible to prescribe drug combinations tailored to individual patients. Unlike traditional standardized treatments, this kind of approach considers each patient's unique medical history to provide more effective therapeutic options [1, 2], thereby enhancing treatment outcomes. A huge amount of accumulated electronic health records (EHR) [3] enables this kind of approach by providing comprehensive and accessible health record data, which typically include patient demographic data, clinical information, etc.

The records of a patient's visit often include multiple types of diagnoses or procedures, or even various complications. For example, diabetic patients often develop comorbid conditions, such as cardiovascular diseases (e.g., coronary artery disease or hypertension) and diabetic nephropathy. These interconnected conditions necessitate careful consideration of potential drug interactions, such as those between cardiovascular drugs and antidiabetic agents, or the impact of antidiabetic medications on renal function. Nevertheless, existing studies often fail to adequately address these complexities and overlook the multifaceted nature of these medical events.

To better model the relationships among different medical entities, a few recent studies try to utilize graph structures to represent EHR data and use GNN-based models to recommend drugs for patients [4–7]. However, these models often fail to capture the complex interactions among diagnoses, procedures, and drugs. They primarily focus on pairwise relationships and cannot effectively model the higher-order connections essential for understanding the intricate nature of medical data. This limitation results in a loss of crucial information and reduces the accuracy of drug combination recommendations.

To address the above problems, we propose a **H**ierarchical **H**yper**G**raph **C**onvolution **N**etwork for **Drug** Combination **Rec**ommendation method named **HHGCN-DrugRec**. Specifically, we design a hierarchical hypergraph structure to better capture the intricate connections among medical entities, i.e. capturing group relationships rather than just pairwise connections to more effectively model the intricate and interdependent relationships among medical entities. So that, it enables more accurate modeling of each patient visit as well as drug combination effects for drug combination recommendation. Furthermore, given that a patient visit may involve multiple types of diagnoses and procedures, we design a capsule-based patient representation learning module to capture the multifaceted nature of these medical events, thereby improving the modeling of patient query.

The main contributions of this paper can be summarized as follows:

- We propose a hierarchical hypergraph convolution network for drug combination recommendation. By designing an innovative hierarchical hypergraph structure, our method can better capture the complex, higher-order relationships among diagnoses, procedures, and drugs in EHR data, which are crucial for accurate drug combination recommendation.
- We design a capsule-based patient representation learning method. It allows our model to effectively capture the multifaceted nature of patient diagnoses and procedures, ensuring a more comprehensive understanding of the patient's query.
- Experimental results on the departmental partitioned MIMIC-III and MIMIC-IV datasets demonstrate the effectiveness and feasibility of the proposed HHGCN-DrugRec model compared with state-of-the-art methods.

2 Related Work

Early research in drug combination recommendation, such as LEAP [8], predominantly adopts an instance-based approach [9]. These methods primarily focus on single-visit data and treat drug recommendation as a multi-instance and multi-label classification problem. Recent research has shifted towards modeling the sequential relationships in patients' medical records to enhance personalization. By considering the entire history of patient visits, these approaches aim to capture temporal patterns and evolving medical conditions, thereby providing more tailored drug combination recommendations. Among them, Carmen [10] integrates patient history with molecular learning. AMANet [11] integrates both self-attention and inter-attention to explore interactions in heterogeneous medical sequences. Moreover, in order to solve the drug safety issues, many models leverage external clinical knowledge to avoid unfavorable drug combinations. For example, SafeDrug [12] enhances safety and accuracy by encoding drug molecule structure information. Additionally, 4SDrug [13] recommends fewer drugs to ensure safety. StratMed [14] utilizes a dual-property network to balance safety and accuracy in medication recommendation.

Graph Neural Networks (GNNs) have garnered significant attention in the domain of recommendation systems due to their proficiency in modeling structural data and capturing high-order connectivity [15, 16]. In the field of drug combination recommendation, GNNs have also demonstrated remarkable effectiveness in modeling complex relationships within medical data. For example, GAMENet [17] integrates the drug-drug interactions (DDI) knowledge graph with a memory module implemented as graph convolutional networks. COGNet [6] employs both EHR and DDI graphs to extract detailed drug information. SMGCN [18] leverages the bipartite symptom-herb graph to learn node embeddings for target herbs.

These GNN-based models illustrate the significant progress made in drug combination recommendation systems by leveraging graph structures to capture complex patterns and dependencies inherent in medical data. However, they are primarily designed for traditional graphs, limiting their ability to capture higher-order dependencies between medical entities.

3 Problem Definition

To make safe and accurate drug combination recommendation, we need to take consideration of both the EHR data and the domain knowledge of DDI (drug-drug interactions) relationships.

Electronic Health Records (EHR). EHR data contains detailed medical histories of patients, represented as longitudinal sequences of medical codes, including diagnoses, procedures, and drugs. Formally, the EHR of a patient u_i can be denoted as a sequence $\boldsymbol{V}_{u_i} = [\boldsymbol{v}_{u_i}^1, \boldsymbol{v}_{u_i}^2, \ldots, \boldsymbol{v}_{u_i}^N]$ where $\boldsymbol{v}_{u_i}^t$ represents the t-th visit and N is the total number of visits. In particular, $\boldsymbol{v}_{u_i}^t$ can be described by a concatenation of multi-hot encoded vectors: $\boldsymbol{v}_{u_i}^t = [\boldsymbol{d}_{u_i}^t, \boldsymbol{p}_{u_i}^t, \boldsymbol{m}_{u_i}^t]$,

$\boldsymbol{d}_{u_i}^t \in \{0,1\}^{|\boldsymbol{\mathcal{D}}|}$, $\boldsymbol{p}_{u_i}^t \in \{0,1\}^{|\boldsymbol{\mathcal{P}}|}$, and $\boldsymbol{m}_{u_i}^t \in \{0,1\}^{|\boldsymbol{\mathcal{M}}|}$, where $\boldsymbol{\mathcal{D}}$, $\boldsymbol{\mathcal{P}}$, and $\boldsymbol{\mathcal{M}}$ denote the sets of diagnosis, procedure, and medicine (drug) codes, respectively.

Domain Knowledge of DDI Relationships. The DDI relationships can be represented as a graph $\mathcal{G}_{\text{ddi}} = \{\boldsymbol{\mathcal{M}}, \mathcal{E}_{\text{ddi}}\}$, where $\mathcal{E}_{\text{ddi}}$ represents the edge set of known DDIs between pairs of drugs, serving as prior knowledge. Formally, we can use the adjacency matrix $\boldsymbol{A}_{\text{ddi}} \in \mathbb{R}^{|\boldsymbol{\mathcal{M}}| \times |\boldsymbol{\mathcal{M}}|}$ to illustrate the construction of the edge set $\mathcal{E}_{\text{ddi}}$. Specifically, $\boldsymbol{A}_{\text{ddi}}[i,j] = 1$ means that the i-th drug and the j-th drug interact with each other.

Drug Combination Recommendation. Given the historical EHR set and the domain knowledge of DDI relationships, our goal is to learn a drug combination recommendation function $f(\cdot)$. It can predict the drug set $\hat{\boldsymbol{y}}^t = f(\boldsymbol{d}_{u_i}^t, \boldsymbol{p}_{u_i}^t, \ldots, \boldsymbol{V}_{u_i}^{t-1})$ for each patient u_i at his/her t-th visit, where $\boldsymbol{d}_{u_i}^t$ is patient u_i's diagnosis set at time t, $\boldsymbol{p}_{u_i}^t$ is the procedure set at time t, $\boldsymbol{V}_{u_i}^{t-1}$ denotes historical visit sequences of patient u_i, and $\hat{\boldsymbol{y}}^t$ is a multi-label output where each drug takes a value of 0 or 1.

4 Method

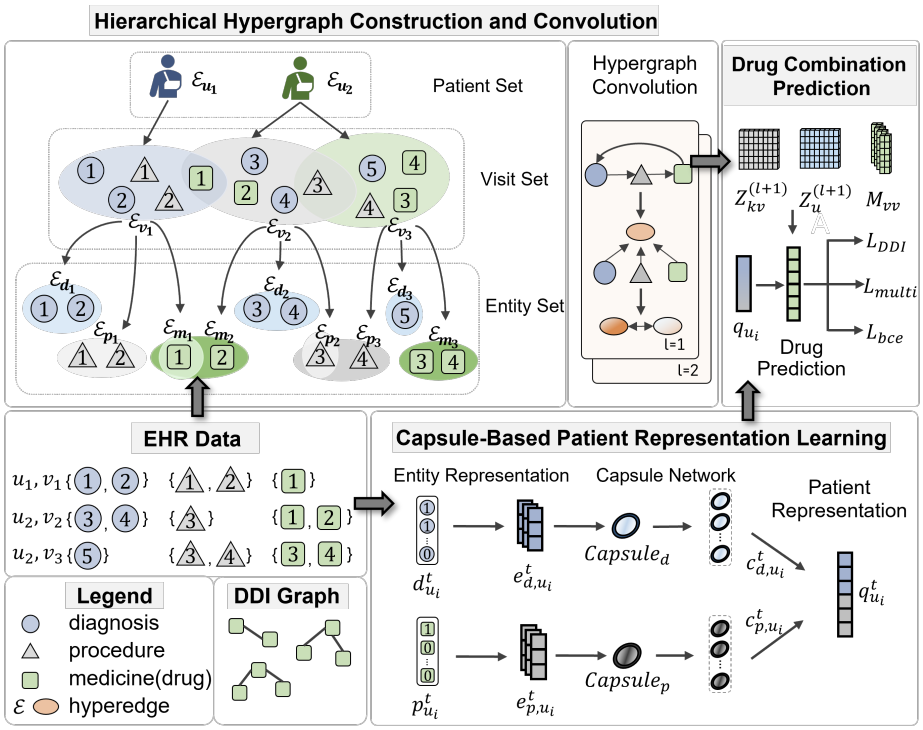


Fig. 1: Framework of the proposed HHGCN-DrugRec model.

As shown in Figure 1, the proposed HHGCN-DrugRec model is composed of three main modules, including capsule-based patient representation learning, hierarchical hypergraph construction and convolution, and drug combination prediction. The capsule-based patient representation learning module aims to learn the multifaceted representation of patients. The hierarchical hypergraph construction and convolution module aims to apply the designed hierarchical hypergraph structure and hypergraph convolution to model EHR graph and patient's historical visit sequences. The drug combination prediction module aims to generate the final recommended drug combination based on the output of the previous two modules.

4.1 Capsule-Based Patient Representation Learning

To capture the multifaceted nature of patients' health status, we construct a capsule network to learn patient's representation. It consists of multiple capsule units with each capsule unit encoding a specific type of medical record entities. In addition, the dynamic routing mechanism of capsule network is used to adaptively allocate weights among different capsules to capture complex relationships among medical entities.

Given patient u_i's diagnosis set $\boldsymbol{d}_{u_i}^t$ and procedure set $\boldsymbol{p}_{u_i}^t$ at time t, we first embed them into a latent vector space to obtain diagnosis and procedure embedding representations $\boldsymbol{e}_{d,u_i}^t \in \mathbb{R}^{|\boldsymbol{\mathcal{D}}| \times d}$ and $\boldsymbol{e}_{p,u_i}^t \in \mathbb{R}^{|\boldsymbol{\mathcal{P}}| \times d}$, where d denotes the embedding dimension. Then we transform each of them into a capsule-based feature representation. Specifically, the representation for each capsule is computed as follows:

$$\boldsymbol{v}_{d,u_i}^{j,t} = \text{squash}\left(\sum_k b_{j_1 k} \mathbf{W}_{j_1} \boldsymbol{e}_{d,u_i}^t\right), \quad \boldsymbol{v}_{p,u_i}^{j,t} = \text{squash}\left(\sum_k b_{j_2 k} \mathbf{W}_{j_2} \boldsymbol{e}_{p,u_i}^t\right) \quad (1)$$

where b_{jk} is the coupling coefficient calculated via the dynamic routing mechanism, the learnable parameter $\mathbf{W}_j$ is used for linear projection, and squash$(\cdot)$ is a non-linear activation function that restricts the magnitude of the output vector to the range $(0, 1)$. The vector $\boldsymbol{v}_{d,u_i}^{j,t}$ and $\boldsymbol{v}_{p,u_i}^{j,t}$represent the output of the j-th capsule for the diagnoses and procedures of patient u_i.

Each capsule reflects the hidden attributes inherent in diagnoses and procedures. To aggregate the characteristics of all capsules and better capture the multifaceted nature of patient u_i, we concatenate the outputs of all capsules. By doing so, the capsule networks provide a richer representation that goes beyond simple feature extraction, effectively encapsulating the complexity and multidimensionality of medical decision-making processes.

$$\boldsymbol{c}_{d,u_i}^t = [\boldsymbol{v}_{d,u_i}^{1,t}, \boldsymbol{v}_{d,u_i}^{2,t}, \cdots, \boldsymbol{v}_{d,u_i}^{J,t}], \quad \boldsymbol{c}_{p,u_i}^t = [\boldsymbol{v}_{p,u_i}^{1,t}, \boldsymbol{v}_{p,u_i}^{2,t}, \cdots, \boldsymbol{v}_{p,u_i}^{J,t}] \quad (2)$$

where J is the total number of capsules that should be tuned according to the dataset. Since different datasets contain varying numbers and types of diseases

and procedures, the required number of capsules will vary. Datasets with more complex or diverse medical conditions may need more capsules, while simpler datasets may need fewer.

Finally, we concatenate the hidden diagnosis state $\boldsymbol{c}_{d,u_i}^t$ and hidden procedure state $\boldsymbol{c}_{p,u_i}^t$ to form the representation of patient u_i at time t.

$$\boldsymbol{q}_{u_i}^t = \text{CONCAT}[\boldsymbol{c}_{d,u_i}^t, \boldsymbol{c}_{p,u_i}^t] \tag{3}$$

4.2 Hierarchical Hypergraph Construction

To capture both the collective patterns within EHR data across all visits and the unique medical feature of each patient, we design a hierarchical hypergraph to represent the EHR data. It consists of multiple types of hyperedges designed for different purposes, including entity hyperedge, visit hyperedge, and patient hyperedge as shown in the top left corner of Fig.1. The entity hyperedge is used to represent and update the same types of nodes in the specific visit. The visit hyperedge is designed to learn the prescribing logic for each visit record, and the patient hyperedge is designed to learn the characteristics of individual patients.

In order to learn the complex relationships between diagnosis, procedure, and drug nodes within each visit record in the EHR data and ultimately capture the logic behind medical prescriptions, we construct visit hyperedges with consideration of the specific combination effects of each entity type. Specifically, each visit hyperedge is consisted of three types of entity hyperedges: diagnosis hyperedge, procedure hyperedge, and drug hyperedge. Formally, the visit hyperedge $\mathcal{E}_{v_i}$ for visit v_i connects the diagnosis hyperedge $\mathcal{E}_{d_i}$, procedure hyperedge $\mathcal{E}_{p_i}$, and medicine (drug) hyperedge $\mathcal{E}_{m_i}$ in the same visit, i.e.

$$\mathcal{E}_{v_i} = \{\mathcal{E}_{d_i}, \mathcal{E}_{p_i}, \mathcal{E}_{m_i}\}, \quad \mathcal{E}_{d_i} = \{v \in \mathcal{V}_d^{v_i}\}, \quad \mathcal{E}_{p_i} = \{v \in \mathcal{V}_p^{v_i}\}, \quad \mathcal{E}_{m_i} = \{v \in \mathcal{V}_m^{v_i}\} \tag{4}$$

where $\mathcal{V}_d^{v_i}$, $\mathcal{V}_p^{v_i}$, and $\mathcal{V}_m^{v_i}$ represent the set of diagnoses, procedures and drugs within visit v_i respectively.

To capture the characteristics of each individual patient's medical feature, we construct a specific type of hyperedge, i.e. patient hyperedge, which connects all visit hyperedges from the patient's historical visit records into a large hyperedge. The patient hyperedge $\mathcal{E}_{u_j}$ represents all medical information involved in different visit records for the same patient u_j, i.e.

$$\mathcal{E}_{u_j} = \{\mathcal{E}_{v_m}, \mathcal{E}_{v_{m+1}}, \ldots, \mathcal{E}_{v_n}\} \tag{5}$$

where m and n represent the start and end visits associated with patient u_j, respectively.

4.3 Hierarchical Hypergraph Convolution

To capture complex connections among different types of vertex and hyperedge, we design a hierarchical hypergraph convolution strategy as shown in Figure 2.

It can aggregate features from different nodes and different types of hyperedges, providing a richer representation that captures both local interactions (within a single visit) and global interactions (across multiple visits). This allows us to model the progression of a patient's medical condition in a more comprehensive manner, improving the effectiveness of the medication recommendation process.

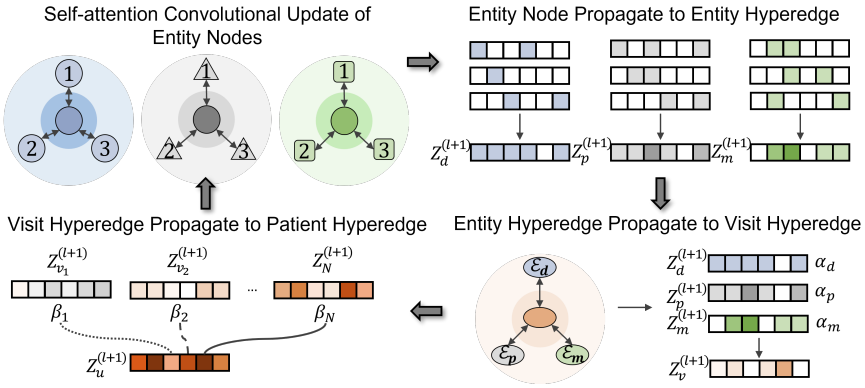


Fig. 2: Illustration of the proposed hierarchical hypergraph convolution method.

Let $\boldsymbol{X}_d \in \mathbb{R}^{|\boldsymbol{\mathcal{D}}|\times d}$, $\boldsymbol{X}_p \in \mathbb{R}^{|\boldsymbol{\mathcal{P}}|\times d}$, and $\boldsymbol{X}_m \in \mathbb{R}^{|\boldsymbol{\mathcal{M}}|\times d}$ be the feature matrices of diagnosis, procedure, and drug respectively. The corresponding incidence matrices are $\boldsymbol{H}_d \in \mathbb{R}^{|\mathcal{E}_d|\times|\boldsymbol{\mathcal{D}}|}$, $\boldsymbol{H}_p \in \mathbb{R}^{|\mathcal{E}_p|\times|\boldsymbol{\mathcal{P}}|}$, and $\boldsymbol{H}_m \in \mathbb{R}^{|\mathcal{E}_m|\times|\boldsymbol{\mathcal{M}}|}$, where $|\mathcal{E}_d|$, $|\mathcal{E}_p|$, $|\mathcal{E}_m|$ denote the number of diagnosis hyperedge, procedure hyperedge and medicine (drug) hyperedge, respectively. The convolutional update for each entity node uses self-attention within the corresponding entity hyperedges.

$$\boldsymbol{X}_i^{(l+1)} = \sigma(\boldsymbol{H}_i \,\mathrm{Att}(\boldsymbol{X}_i)\, \boldsymbol{W}_{ii}^{(l)}), \quad i \in \{d, p, m\} \tag{6}$$

where $\mathrm{Att}(\boldsymbol{X})$ represents the self-attention mechanism applied to the feature matrix $\boldsymbol{X}$, $\boldsymbol{W}_{dd}^{(l)}, \boldsymbol{W}_{pp}^{(l)}, \boldsymbol{W}_{mm}^{(l)}$ are the learnable weight matrices at the l-th layer, and σ denotes the activation function sigmoid.

We subsequently perform a mean pooling operation on the updated node features to obtain the representations for each entity hyperedge.

$$\boldsymbol{Z}_i^{(l+1)} = \mathrm{Agg}(\boldsymbol{X}_i^{(l+1)}), \quad i \in \{d, p, m\} \tag{7}$$

where $Z_d \in \mathbb{R}^{|\mathcal{E}_d|\times d}$, $Z_p \in \mathbb{R}^{|\mathcal{E}_p|\times d}$, and $Z_m \in \mathbb{R}^{|\mathcal{E}_m|\times d}$ denote the entity hyperedge embeddings for diagnosis, procedure, and drug, respectively.

To aggregate the diagnosis, procedure, and drug entity hyperedges within each visit into a visit hyperedge, we introduce learnable weights $\alpha_d, \alpha_p, \alpha_m$, which are normalized using the softmax function to determine the relative importances

of the three kinds of entity hyperedges.

$$\boldsymbol{Z}_v^{(l+1)} = \alpha_d \boldsymbol{Z}_d^{(l+1)} + \alpha_p \boldsymbol{Z}_p^{(l+1)} + \alpha_m \boldsymbol{Z}_m^{(l+1)} \tag{8}$$

where w_i are learnable scalar parameters.

To propagate information to the patient level, we aggregate the visit hyperedges to form the patient hyperedge. Each patient hyperedge is composed of multiple visit hyperedges. The incidence matrix $\boldsymbol{R} \in \mathbb{R}^{|\mathcal{E}_u| \times |\mathcal{E}_v|}$ represents the relationship between patients and their associated visits, where $|\mathcal{E}_u|$ is the number of patient hyperedges and $|\mathcal{E}_v|$ is the number of visit hyperedges. Given that different visits may contribute differently to the patient-level representation, we introduce learnable weight parameters β_j for each visit hyperedge. These weights are normalized using the softmax function and further modulated by a temporal decay factor, reflecting the relative importance of each visit based on its recency:

$$\beta_j = \frac{\exp(b_j)}{\sum_{k=1}^{N} \exp(b_k)} \cdot \gamma_j, \quad j = 1, 2, \ldots, N \tag{9}$$

where b_j are learnable scalars, and N represents the number of visits associated with patient u. The temporal decay factor γ_j is defined as $\gamma_j = e^{-\lambda t_j}$, where t_j is the time interval of the j-th visit relative to the current time, and λ is a hyperparameter that controls the rate of decay. γ_j ensures that earlier visits contribute less to the final representation compared to more recent visits.

The final patient hyperedge representation $\boldsymbol{Z}_u^{(l+1)}$ is obtained by performing a weighted sum over all associated visit hyperedges, considering both the incidence matrix $\boldsymbol{R}$ and the learned importance of each visit.

$$\boldsymbol{Z}_u^{(l+1)} = \sigma\left(\boldsymbol{R}\left(\sum_{j=1}^{N} \beta_j \boldsymbol{Z}_{v_j}^{(l+1)}\right)\boldsymbol{W}_{pv}^{(l)}\right) \tag{10}$$

where $\boldsymbol{Z}_{v_j}^{(l+1)}$ denotes the representation of the j-th visit hyperedge, and β_j is the temporal decay weight for visit v_j. $\boldsymbol{W}_{pv}^{(l)}$ is the learnable weight matrix used for feature transformation at the patient level, and σ is the activation function.

This representation can capture a holistic view of the patient's medical history, effectively integrating the diverse information from all visits while allowing the model to learn the relative importance of each visit. In addition, it takes consideration of both the visits' structural relationships and chronological sequence with temporal decay.

4.4 Drug Combination Prediction

The drug combination prediction module is responsible for generating the final drug recommendation based on the patient query $\boldsymbol{q}_{u_i}^t$, the patient hyperedge representation $\boldsymbol{Z}_u^{(l+1)}$, the visit hyperedge representation $\boldsymbol{Z}_v^{(l+1)}$, and the corresponding drug set $\boldsymbol{\mathcal{M}}_{vv}$.

To identify similar visits for the current visit, we compute the similarity between the query representation and the visit hyperedge representation matrix $\boldsymbol{Z}_v^{(l+1)}$. The objective is to assess the relevance of each historical visit to the current query, thereby leveraging the idea of collaborative filtering to support subsequent drug recommendation. The results are stored in a matrix $Q_v \in \mathbb{R}^{|\mathcal{E}_v| \times 2}$. We then adopt a similar approach to compute the similarity between patients. Notably, a learnable weight is introduced when calculating the similarity between the patients and their own embedding. This ensures that the self-similarity does not disproportionately dominate the results, thereby preserving the influence of visit-level similarities in downstream tasks. The resulting similarity matrix, $Q_u \in \mathbb{R}^{|\mathcal{E}_u| \times 2}$, consists of two columns: the first column records the patient IDs, and the second column contains the corresponding similarity scores.

$$Q_v[i,2] = \frac{\boldsymbol{q}_{u_i}^t \cdot \boldsymbol{Z}_v^{(l+1)}[i]}{\|\boldsymbol{q}_{u_i}^t\| \cdot \|\boldsymbol{Z}_v^{(l+1)}[i]\|}, \quad Q_u[i,2] = \mathrm{sim}(\boldsymbol{Z}_u^{(l+1)}[i], \boldsymbol{Z}_u^{(l+1)}[j]) \cdot w \tag{11}$$

where w is a learnable weight applied when $i = j$.

The final similarity score S is computed as a weighted sum of the patient-level similarity $Q_u[i,2]$ and the visit-level similarity $Q_v[j,2]$, using learnable weights α_u and β_v that satisfy $\alpha_u + \beta_v = 1$. This approach allows the model to dynamically balance the contributions of patient- and visit-level similarities during training.

$$\boldsymbol{S} = \alpha_u \cdot Q_u + \beta_v \cdot Q_v, \quad \forall \boldsymbol{R}[u_i, v_j] = 1 \tag{12}$$

Finally, using the similarity matrix $\boldsymbol{S}$, we identify the top-k most similar historical visits for the current visit. For each selected visit v_i, we consider its corresponding drug set $\boldsymbol{\mathcal{M}}_{vv}[v_i]$. The final drug recommendation set is generated by aggregating the drugs from the selected visits, where each drug is weighted by the similarity score in $\boldsymbol{S}$. The weight of a drug m_k in the final recommendation is computed as:

$$w_{m_k} = \sum_{v_i \in \text{top-}k} \boldsymbol{S}[v_i] \cdot \delta(m_k \in \boldsymbol{\mathcal{M}}_{vv}[v_i]) \tag{13}$$

where $\boldsymbol{S}[v_i]$ is the similarity score of visit v_i, and $\delta(m_k \in \boldsymbol{\mathcal{M}}_{vv}[v_i])$ is an indicator function that equals 1 if the drug m_k is present in the drug set $\boldsymbol{\mathcal{M}}_{vv}[v_i]$, and 0 otherwise.

4.5 Combined Loss for Parameter Learning

To optimize all learnable parameters, we introduce a combined loss to find a better balance between accuracy and safety. For accuracy, we employ binary cross-entropy loss L_{bce} and multi-label margin loss L_{multi}. The binary cross-entropy loss aims to measure the discrepancy between predicted probabilities and actual binary labels for each drug, and the multi-label margin loss measures

the margin between scores of relevant and irrelevant drugs to ensure proper ranking. For safety, we design an adverse DDI loss based on the DDI adjacency matrix $\boldsymbol{A}_{\text{ddi}}$.

$$L_{\text{bce}} = -\sum_{i=1}^{|\boldsymbol{\mathcal{M}}|} \left[y_i^t \log(\hat{y}_i^t) + (1 - y_i^t) \log(1 - \hat{y}_i^t) \right]$$

$$L_{\text{multi}} = \sum_{i,j:y_i^t=1,y_j^t=0} \frac{1}{|\boldsymbol{\mathcal{M}}|} \max\left(0, 1 - (\hat{y}_i^t - \hat{y}_j^t)\right), \quad L_{\text{DDI}} = \sum_{i=1}^{|\boldsymbol{\mathcal{M}}|} \sum_{j=1}^{|\boldsymbol{\mathcal{M}}|} \boldsymbol{A}_{ij} \cdot \hat{y}_i^t \cdot \hat{y}_j^t \tag{14}$$

where y_i^t represents the ground truth label for the i-th drug, $\hat{y}_i^t$ denotes the predicted probability of the i-th drug at time t, and $\cdot$ is the product between scalars. Note that the above loss functions are defined for a single visit. During model training, loss back propagation will be conducted at patient level by the averaged losses across all visits.

To balance accuracy and safety during model training, we alternate between the multi-label prediction loss and DDI loss consistent with prior research in [26]:

$$L = \alpha(\beta L_{\text{bce}} + (1 - \beta) L_{\text{multi}}) + (1 - \alpha) L_{\text{DDI}}$$

$$\alpha = \begin{cases} 1, & \text{rate}_{\text{ddi}} \leq \gamma \\ \max\left\{0, 1 - \frac{\text{rate}_{\text{ddi}} - \gamma}{kp}\right\}, & \text{rate}_{\text{ddi}} > \gamma \end{cases} \tag{15}$$

Here, β controls model complexity or regularization strength. α, related to rate_{ddi}, adjusts how the DDI rate affects the model's decisions. γ, ranging from 0 to 1, sets the DDI acceptance threshold for clinical use. And kp is a correction factor that scales the information based on proportionality.

5 Experiments

5.1 Experimental Settings

Datasets. We evaluate the model performance on both MIMIC-III [19] and MIMIC-IV [20], which are two real-world EHR datasets that are publicly available. The two datasets differ in terms of data sources, time span of data collection, data organization, the version of ICD codes, etc. To better align with actual hospital workflows, we do not apply the entire MIMIC dataset directly. Instead, we select specific departments for training and testing, using the services table from hosp module for department filtering. Additionally, we retain all visit records from the selected departments, including patients with only a single visit.

- MED: Medical - general service for internal medicine.
- SURG: Surgical - general surgical service not classified elsewhere.

– GU: Genitourinary - reproductive organs/urinary system.
– OMED: Oncologic Medical - non-surgical, relating to cancer.

Evaluation Metrics. To evaluate the prediction accuracy and drug safety, we utilize four commonly used metrics as in [17,26]. They are Jaccard Similarity Score (Jaccard), Average F1 Score (F1), Precision-Recall AUC (PRAUC), and DDI Rate .

Baseline Methods. To construct a comprehensive comparison, we adopt several different types of baseline methods, including LR, ECC [21], RETAIN [22], DMNC [23], GAMENet [17], COGNet [6], MICRON [24], MoleRec [26] and RASNet [25].

Logistic Regression (LR) and Ensemble Classifier Chain (ECC) employ traditional instance-based learning approaches for prediction tasks. RETAIN, DMNC, GAMENet, MICRON, and RASNet all leverage RNNs or memory mechanisms. RETAIN employs an attention and gating mechanism to improve interpretability, DMNC uses a memory-augmented neural network, and GAMENet leverages graph convolutional networks with dynamic memory. MICRON and RASNet focus on recurrent networks for patient records and drug change prediction. COGNet and MoleRec use advanced representation learning techniques. COGNet is an encoder-decoder generative network, while MoleRec employs substructure-aware molecular representation learning for drug recommendation.

Parameter Settings. We randomly divide each dataset into training, validation, and test sets in a $2/3 : 1/6 : 1/6$ ratio. Hyperparameters are selected based on performance on the validation set. The model parameters are optimized using the Adam optimizer with an initial learning rate of $1e^{-4}$. Each baseline method is configured using default settings in the original paper or our fine-tuned parameters leading to the best performance.

5.2 Results of Comparative Study

Table 1 presents the performance of different methods. It is evident that our proposed model, HHGCN-DrugRec, outperforms all baseline methods in most cases. Specifically, HHGCN-DrugRec achieves the highest Jaccard index, PRAUC, and F1-score, while maintaining a competitive DDI rate across the four evaluated departments.

Traditional learning-based models like LR and ECC demonstrate relatively poor performance, as they struggle to capture the non-linear relationships and interactions among medical variables, leading to an oversimplification of the underlying data patterns. Models like RETAIN and DMNC overly rely on predefined sequence structures, reducing their flexibility and adaptability when applied to diverse and complex visit records. This limitation is reflected in their lower PRAUC and F1-scores across all datasets, indicating that these models fail to capture the full scope of information contained in EHR data. The traditional graph-based approach GAMENet is constrained by the limited number of nodes connected in conventional graphs. While convolutional operations allow the model to learn multi-hop information, its capacity remains restricted,

Table 1: Performance of different methods. The best and the runner-up results are highlighted in bold and underline respectively.

Methods	MIMIC-III-MED				MIMIC-III-SURG			
	Jaccard ↑	DDI ↓	PRAUC ↑	F1 ↑	Jaccard ↑	DDI ↓	PRAUC ↑	F1 ↑
LR	0.3254	0.0959	0.6055	0.4646	0.3626	0.0754	0.6215	0.5055
ECC	0.2801	0.0916	0.6014	0.4027	0.3221	0.0757	0.6184	0.4544
RETAIN	0.3180	0.0956	0.5863	0.4696	0.3491	0.0729	0.5766	0.5024
DMNC	0.2847	0.0903	0.5526	0.4124	0.3156	0.0734	0.5323	0.4482
GAMENet	0.3737	0.1130	0.6198	0.5250	0.4247	0.0717	0.6666	0.5776
COGNet	0.3765	0.1169	0.6028	0.5274	0.3954	0.1044	0.6406	0.5452
MICRON	0.3481	0.0666	0.6203	0.4928	0.3525	0.0773	0.6149	0.4942
MoleRec	0.3843	0.0737	0.6205	0.5355	0.4164	0.0652	0.6691	0.5783
RASNet	0.3846	0.0673	0.6259	0.5361	0.4274	0.0667	0.6694	0.5704
HHGCN-DrugRec	**0.3940**	**0.0613**	**0.6289**	**0.5423**	**0.4305**	**0.0649**	**0.6746**	**0.5819**
Methods	MIMIC-IV-GU				MIMIC-IV-OMED			
	Jaccard ↑	DDI ↓	PRAUC ↑	F1 ↑	Jaccard ↑	DDI ↓	PRAUC ↑	F1 ↑
LR	0.2722	0.2720	0.5573	0.3555	0.2982	0.0689	0.5811	0.3893
ECC	0.2929	0.2523	0.5536	0.3778	0.3170	0.0835	0.5951	0.4119
RETAIN	0.2516	0.2481	0.4849	0.3542	0.3292	0.1492	0.5980	0.4188
DMNC	0.3171	0.2540	0.4134	0.4219	0.3404	0.0802	0.5219	0.4412
GAMENet	0.3264	0.2790	0.5566	0.4480	0.3515	0.0595	0.5989	0.4762
COGNet	0.3193	**0.1925**	0.4993	0.4183	0.3191	0.0804	0.5164	0.4017
MICRON	0.2335	0.3146	0.5227	0.3107	0.3072	0.0555	0.5768	0.4022
MoleRec	0.3433	0.2870	0.5675	0.4569	0.3415	0.0591	0.5924	0.4644
RASNet	0.2594	0.2127	0.4597	0.3516	0.2882	0.0985	0.5154	0.3812
HHGCN-DrugRec	**0.3568**	0.2059	**0.5764**	**0.4682**	**0.3618**	**0.0523**	**0.6091**	**0.4802**

hindering the ability to capture more extensive and intricate information. Although models such as COGNet, MICRON, MoleRec and RASNet are specifically designed to process EHR data, they fall short in capturing the intricate relationships within the EHR data, resulting in less satisfactory performance.

Conversely, the enhanced modeling capabilities of HHGCN-DrugRec enable it to more fully utilize the available information in EHRs, thereby achieving superior results across all metrics. The use of hypergraph modeling allows HHGCN-DrugRec to effectively represent complex interactions, leading to significant improvements in both accuracy and safety of drug combination recommendation.

5.3 Ablation Study

We conduct a series of ablation studies to evaluate the contribution of different components on the performance of HHGCN-DrugRec. The study includes the following five variants: (a) **w/o capsule network** excludes the capsule network module, which is designed to capture the multifaceted nature of patient diagnoses and procedures, (b) **w/o visit layer** removes the hypergraph convolution module for each visit, which is designed to capture the complex interactions among diagnoses, procedures, and drugs, (c) **w/o patient layer** excludes the module

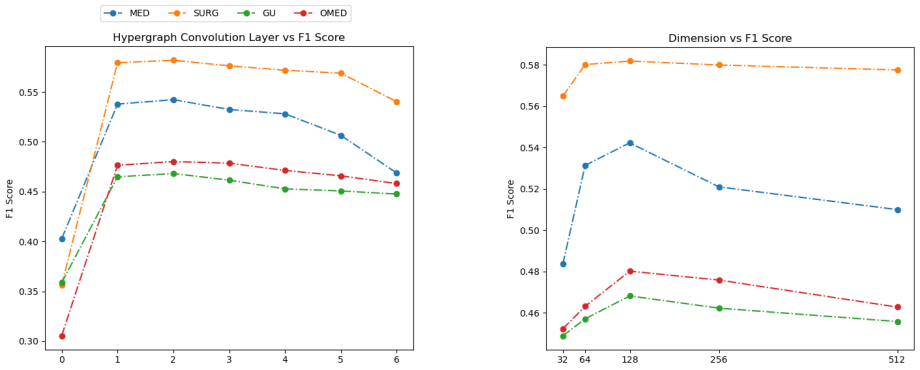
Table 2: Results of ablation study.

Methods	MIMIC-III-MED				MIMIC-III-SURG			
	Jaccard ↑	DDI ↓	PRAUC ↑	F1 ↑	Jaccard ↑	DDI ↓	PRAUC ↑	F1 ↑
w/o capsule	0.3941	0.0582	0.6276	0.5365	0.3978	0.0653	0.6582	0.5625
w/o visit	0.3824	0.0673	0.6052	0.5275	0.4169	0.0735	0.6466	0.5573
w/o patient	0.3817	0.0654	0.6139	0.5279	0.4055	0.0798	0.6531	0.5542
w/GCN visit	0.3846	0.0593	0.6122	0.5341	0.4195	0.0726	0.6554	0.5697
w/GCN patient	0.3901	0.0585	0.6154	0.5367	0.4203	0.0698	0.6614	0.5723
HHGCN-DrugRec	**0.3940**	**0.0613**	**0.6289**	**0.5423**	**0.4305**	**0.0649**	**0.6746**	**0.5819**
Methods	MIMIC-IV-GU				MIMIC-IV-OMED			
	Jaccard ↑	DDI ↓	PRAUC ↑	F1 ↑	Jaccard ↑	DDI ↓	PRAUC ↑	F1 ↑
w/o capsule	0.3415	0.2641	0.5717	0.4636	0.3610	0.0567	0.5989	0.4763
w/o visit	0.3403	0.2700	0.5686	0.4541	0.3528	0.0538	0.5862	0.4688
w/o patient	0.3397	0.2715	0.5623	0.4518	0.3586	0.0557	0.5817	0.4670
w/GCN visit	0.3452	0.2659	0.5697	0.4614	0.3587	0.0551	0.5914	0.4698
w/GCN patient	0.3403	0.2392	0.5628	0.4625	0.3601	0.0554	0.5873	0.4706
HHGCN-DrugRec	**0.3568**	**0.2059**	**0.5764**	**0.4682**	**0.3618**	**0.0523**	**0.6091**	**0.4802**

for historical sequences, which integrates and enhances the representation of longitudinal patient data, (d) **w/GCN visit layer** replaces the hypergraph with traditional graph and substitutes the hypergraph convolution with a standard graph convolution, and (e) **w/GCN patient layer** replaces hypergraph convolution with graph convolution in patient layer. Both (d) and (e) aim to evaluate the effectiveness of the hypergraph designed in this paper. By comparing these variants with the whole model, we can determine the specific contribution of each component to the overall performance of the HHGCN-DrugRec model.

Table 2 presents the results of the ablation studies. We observe that removing either the hypergraph convolution visit layer or the patient layer from the HHGCN-DrugRec model leads to a significant decrease in accuracy. Replacing the hypergraph with a traditional graph improves the results compared to removing the layer entirely, but still performs worse than the hypergraph-based approach. This highlights the importance of hypergraph convolutions in capturing the complex, higher-order relationships within EHR data. By integrating diverse types of medical information, hypergraph convolutions enhance both the model's representational capacity and its predictive performance.

The capsule network module also plays a vital role in the overall performance of the model. The ablation study shows that removing the capsule network leads to a reduction in model accuracy. The synergistic interaction between these components allows the model to fully leverage the wealth of information embedded in EHR data, capturing the complexity and variability of visit records to deliver more accurate and reliable predictions.



(a) Hypergraph convolution layer l (b) Feature dimension d

Fig. 3: The impact of hypergraph convolution layer l and feature dimension d.

5.4 Hyperparameter Analysis

We conduct a series of experiments to evaluate the effects of hypergraph convolution layer l and feature dimension d. The results are shown in Figure 3.

- **Impact of hypergraph convolution layer l.** The experimental results in Fig.3a indicate that the optimal performance is achieved with two convolutional layers. With zero layers, the performance is poor due to insufficient feature extraction. However, increasing the number of layers beyond two results in a gradual decline in performance. This decline is likely due to the over-smoothing effect, where node representations become too similar, reducing their discriminative power, as well as an increased risk of overfitting, which impairs the model's ability to generalize.
- **Impact of dimension d.** By comparing the performance of HHGCN-DrugRec with varying feature dimension d, we observe that the F1 score reaches its optimal value across all datasets when $d = 128$ as shown in Fig.3b. In lower dimensions, such as 32 or 64, the model may not have enough capacity to effectively capture the complex relationships inherent in the medical entities. On the other hand, as the dimension increases to larger values, such as 256 and 512, the feature introduces significant redundancy, making the model more prone to overfitting the training data. Therefore, an intermediate dimension like $d = 128$ provides a good balance between sufficient expressive power and effective generalization, resulting in the best performance.

6 Conclusion

We have proposed a novel hierarchical hypergraph convolution network for drug combination recommendation, named HHGCN-DrugRec. This model employs hypergraph convolution to adeptly capture the intricate interactions within EHR

data. Moreover, we have integrated a capsule network module to capture the multifaceted nature of patient diagnoses and procedures. Extensive experiments conducted on four departments(MED, SURG, GU, and OMED) of the MIMIC-III and MIMIC-IV datasets validate the effectiveness of our approach, demonstrating superior performance in terms of both accuracy and safety across all evaluated departments.

In the future, we will explore the application of hypergraph structures to model drug molecular graphs and leverage the hierarchical nature of diagnostic and procedural codes to enhance patient representation. These efforts are expected to further enhance the accuracy and safety of our drug combination recommendation model.

Acknowledgments. This work was supported in part by the S&T Program of Hebei (Grant No. 246W7703D) and the National Key Research and Development Program of China (Grant No. 2022YFB2703301).

References

1. Chen, Q., Li, X., Geng, K., Wang, M.: Context-aware safe medication recommendations with molecular graph and DDI graph embedding. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 7053-7060 (2023)
2. Bhoi, S., Lee, M.L., Hsu, W., Tan, N.C.: REFINE: a fine-grained medication recommendation system using deep learning and personalized drug interaction modeling. In: Proceedings of the 37th International Conference on Neural Information Processing Systems, pp. 24013-24024 (2023)
3. Cowie, M.R., et al.: Electronic health records to facilitate clinical research. In: Clinical Research in Cardiology, vol. 106, pp. 1–9 (2017). https://doi.org/10.1007/s00392-016-1025-6
4. Yao, Z., Liu, B., Wang, F., Sow, D., Liu, Y.: Ontology-aware prescription recommendation in treatment pathways using multi-evidence healthcare data. In: ACM Transactions on Information Systems (TOIS), vol. 41, pp. 1-29 (2023). https://doi.org/10.1145/3579994
5. Li, X., et al.: CausalMed: causality-based personalized medication recommendation centered on patient health state. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, pp. 1276–1285 (2024)
6. Wu, R., Qiu, Z., Jiang, J., Qi, G., Wu, X.: Conditional generation net for medication recommendation. In: Proceedings of the ACM Web Conference 2022, pp. 935–945 (2022)
7. Bhoi, S., Lee, M.L., Hsu, W., Fang, H.S.A., Tan, N.C.: Personalizing medication recommendation with a graph-based approach. In: ACM Transactions on Information Systems (TOIS), vol. 40, pp. 1-23 (2021). https://doi.org/10.1145/3488668
8. Zhang, Y., Chen, R., Tang, J., Stewart, W.F., Sun, J.: LEAP: learning to prescribe effective and safe treatment combinations for multimorbidity. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1315–1324 (2017)
9. Shickel, B., Tighe, P.J., Bihorac, A., Rashidi, P.: Deep EHR: a survey of recent advances in deep learning techniques for Electronic Health Record (EHR) analysis. In: IEEE Journal of Biomedical and Health Informatics, vol. 22, pp. 1589-1604 (2018). 10.1109/JBHI.2017.2767063

10. Chen, Q., Li, X., Geng, K., Wang, M.: Context-aware safe medication recommendations with molecular graph and DDI graph embedding. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 7053–7060 (2023)
11. He, Y., Wang, C., Li, N., Zeng, Z.: Attention and memory-augmented networks for dual-view sequential learning. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 125-134 (2020)
12. Yang, C., Xiao, C., Ma, F., Glass, L., Sun, J.: Safedrug: dual molecular graph encoders for recommending effective and safe drug combinations. In: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, pp. 3735-3741 (2021)
13. Tan, Y., et al.: 4SDrug: symptom-based Set-to-set Small and Safe Drug Recommendation. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 3970–3980 (2022)
14. Li, X., Liang, S., Hou, Y., Ma, T.: StratMed: relevance stratification between biomedical entities for sparsity on medication recommendation. In: Knowledge-Based Systems, vol. 284 (2024). https://doi.org/10.1016/j.knosys.2023.111239
15. Gao, C., Wang, X., He, X., Li, Y.: Graph neural networks for recommender system. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, pp. 1623–1625 (2022)
16. Chang, J., et al.: Sequential recommendation with graph neural networks. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 378–387 (2021)
17. Shang, J., Xiao, C., Ma, T., Li, H., Sun, J.: Gamenet: graph augmented memory networks for recommending medication combination. In: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, pp. 1126-1133 (2019)
18. Jin, Y., Zhang, W., He, X., Wang, X., Wang, X.: Syndrome-aware herb recommendation with multi-graph convolution network. In: 2020 IEEE 36th International Conference on Data Engineering (ICDE), pp. 145-156 (2020)
19. Johnson, A.E., et al.: MIMIC-III, a freely accessible critical care database. In: Scientific data, vol. 3, pp. 1–9 (2016). https://doi.org/10.1038/sdata.2016.35
20. Johnson, A.E., et al.: MIMIC-IV, a freely accessible electronic health record dataset. In: Scientific data, vol. 10 (2023). https://doi.org/10.1038/s41597-022-01899-x
21. Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-label classification. In: Machine learning, vol. 85, pp. 333-359 (2011). https://doi.org/10.1007/s10994-011-5256-5
22. Choi, E., et al.: Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. In: Advances in Neural Information Processing Systems, pp. 3504-3512 (2016)
23. Le, H., Tran, T., Venkatesh, S.: Dual memory neural computer for asynchronous two-view sequential learning. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1637–1645 (2018)
24. Yang, C., Xiao, C., Glass, L., Sun, J.: Change matters: medication change prediction with recurrent residual networks. In: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, pp. 3728–3734 (2021)
25. Zhu, Q., et al.: RASNet: recurrent aggregation neural network for safe and efficient drug recommendation. In: Knowledge-Based Systems, vol. 299 (2024). https://doi.org/10.1016/j.knosys.2024.112055
26. Yang, N., Zeng, K., Wu, Q., Yan, J.: MoleRec: combinatorial drug recommendation with substructure-aware molecular representation learning. In: Proceedings of the ACM Web Conference 2023, pp. 4075–4085 (2023)