HI-DR: Exploiting Health Status-Aware Attention and an EHR Graph+ for Effective Medication Recommendation

Taeri Kim[1], Jiho Heo[1], Hyunjoon Kim[2], Sang-Wook Kim[1]*

[1]Department of Computer Science, Hanyang University, South Korea
[2]Department of Data Science, Hanyang University, South Korea
{taerik, linda0123, hyunjoonkim, wook}@hanyang.ac.kr

Abstract

We focus on the medication recommendation problem aiming to recommend accurate medications for a patient's current visit. Most existing methods for this problem utilize the patient's current health status, medications prescribed at her past visits, and an Electronic Health Records (EHR) graph which represents whether medications have been co-prescribed. However, we point out their two key limitations: (1) they have difficulty in utilizing *only* the medications which have been prescribed in health status *similar* to the patient's current health status, *regardless of* whether they are prescribed at her past visits or at other patients' visits; (2) for two medications that have ever been co-prescribed, their EHR graph does *not* consider the *degree* to which one medication is prescribed together when the other is prescribed. To address these two limitations, we propose a novel medication recommendation framework, named **HI-DR** (pronounced as 'Hi Doctor'), composed of following two core ideas: (Idea 1) **H**ealth status-aware attent**I**on; (Idea 2) an electronic health recor**D**s g**R**aph+. Extensive experiments on real-world datasets demonstrate the *significant superiority* of HI-DR (up to 18.69% higher accuracy than the best competitor) and the effectiveness of two core ideas in HI-DR.

Code — https://github.com/Bigdasgit/HI-DR

1 Introduction

Medication recommendation aims to recommend accurate medications to a patient at her current visit. This field has been actively studied as it can save time and effort of doctors in prescribing medications to patients, and can be useful even in circumstances short of doctors (Choi et al. 2016a; Zheng et al. 2021; Wu et al. 2020, 2022; Kim et al. 2024).

Early medication recommender systems relied on *instance-based* methods (Zhang et al. 2017; Wang et al. 2018; Gong et al. 2021), which recommend medications to a patient by considering the diagnoses and procedures *only* given at her *current* visit. Even if some patients receive the same diagnoses and procedures at their current visit, the appropriate medications to prescribe could be different depending on whether each patient's condition is *acute* or *chronic*. However, instance-based methods are *unable* to distinguish between the two conditions, which can be estimated from each patient's *past* health records (Shang et al. 2019b; Yang et al. 2021; Wu et al. 2022; Kim et al. 2024).

To alleviate this limitation of instance-based methods, *longitudinal-based* methods utilize not only the diagnoses and procedures given at the patient's *current* visit but also her *past* health records (Shang et al. 2019b; Yang et al. 2021; Ma et al. 2022; Wu et al. 2022; Yang et al. 2023; Kim et al. 2024). Longitudinal-based methods usually consist of two components, an *encoder* to obtain patient representation and a *predictor* to determine the medications that align with this patient representation. The encoders in most of these methods aggregate the diagnoses/procedures given at *both* the patient's current and past visits via a Gated Recurrent Unit (Cho et al. 2014) to obtain a *patient representation* indicative of her *current health status considering her acute or chronic condition*. The predictors in many of these methods first obtain *medication representations* by capturing *relations between co-prescribed medications*. For this, they apply Graph Convolutional Network (GCN) (Kipf and Welling 2017) to the *Electronic Health Records (EHR) graph*, which represents medications as nodes and connects two medication nodes with an edge if and only if they have been co-prescribed in *at least one visit* among all visits of all patients. Then, they utilize the patient representation generated by the encoder and these medication representations to obtain **(a)** a representation for *every medication*, reflecting how *much* each medication is relevant to the patient's *current health status*, and **(b)** a vector for the *medications prescribed at the patient's past visits*, reflecting how much the health status at each of her past visits is related to her current one. Finally, they recommend medications to the patient by fusing the patient representation, representation **(a)**, and vector **(b)**. Based on this common framework, many longitudinal-based methods have been developed by improving *one* of the encoder and the predictor in existing ones.

In this paper, we focus on the following two ideas in the *predictor* for effective medication recommendation.

(Idea 1) To recommend *accurate* medications, in generating *vector* **(b)**, the medications prescribed in the health status *dissimilar* to the patient's current one should *not* be considered, even if they were prescribed at *her past visits*; the medications prescribed in the health status *similar* to the patient's

*Corresponding author.

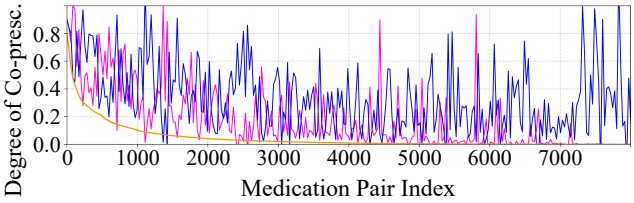

Figure 1: Distribution of the degree of co-prescription for all medication pairs that have ever been co-prescribed in the MIMIC-III dataset. The orange (resp. blue or pink) line represents the degree of co-prescription at the medication pair level (resp. medication level) for each medication pair.

current one may be *beneficial*, even if they were prescribed at *other patients' visits*. These two claims are supported by our empirical findings, which will be detailed in Appendix[1] A. In other words, *only* the visits that have the health status *similar* to the patient's current health status should be used in generating vector **(b)**, *regardless of whose visit they are*.

(Idea 2) To obtain *medication representations* that *accurately* capture relations between *frequently* co-prescribed medications, the *EHR graph* should include the *degree of co-prescription at the medication level*, *i.e.*, the degree to which medication i is prescribed when medication j is prescribed, for two co-prescribed medications i and j. For example, medication j may be *very likely to* be prescribed when medication i is prescribed, while i may be *infrequently* prescribed when j is prescribed. To verify whether this phenomenon truly occurs, we analyze carefully the MIMIC-III (Johnson et al. 2016) dataset, which is commonly utilized in medication recommendation (Shang et al. 2019b; Yang et al. 2021; Wu et al. 2022; Yang et al. 2023; Kim et al. 2024). For every pair of medications that have ever been co-prescribed *at least one visit* among all visits of all patients, Figure 1 shows the distribution of the degree of their co-prescriptions in both *medication pair level* and *medication level*[2]. We can see that, for two medications i and j within a medication pair, the degree to which i is prescribed when j is prescribed and the degree to which j is prescribed when i is prescribed are *quite different* (compare the blue and pink lines in Figure 1). Furthermore, we confirmed that existing methods utilizing the *original* EHR graph obtain *similar* medication representations for *infrequently* co-prescribed medications, thereby leading to the prescription of *unnecessary* medications at the patient's current visit; eventually, this increases the potential for *adverse effects*. Please refer to Appendix A for more detailed information on these findings.

[1]The Appendix for HI-DR can be found at https://github.com/Bigdasgit/HI-DR.

[2]For each pair of medications i and j, their degree of co-prescription at the *medication pair level* is calculated by {(the number of visits where i and j were co-prescribed) / (the total number of visits)}. Their degree of co-prescription at the *medication* i (resp. j) *level* is calculated by {(the number of visits where i and j were co-prescribed) / (the total number of visits where the medication i (resp. j) was prescribed)} (*i.e.*, this involves two different types of degrees of co-prescription for each medication pair).

In this paper, we propose a novel medication recommendation framework, HI-DR, based on (Idea 1) ***H**ealth status-aware attent**I**on* and (Idea 2) an *electronic health recor**D**s (EHR) g**R**aph+*. Via (Idea 1), HI-DR can utilize *only* the medications which have been prescribed in the health status *similar* to the patient's current one; these medications may consist *solely* of those prescribed to other patients *if necessary*. Via (Idea 2), HI-DR can obtain medication representations that *accurately* capture relations between medications which are *frequently* co-prescribed. It is worth mentioning that (1) our goal is to recommend *accurate* medications to each patient based on a set of the medications which have been *actually prescribed by doctors in clinical use*, referred to as the *ground-truth set*; (2) surprisingly, the ground-truth sets have *some degree of adverse effects*; therefore, as accuracy increases, adverse effects *also naturally* increase (Wu et al. 2022; Kim et al. 2024); (3) the significance of our EHR graph+ lies in its ability to reduce these adverse effects *without compromising accuracy*.

Our contributions are summarized as follows:

- **Important Observations**: We observed that existing medication recommender systems overlook two key limitations of (Idea 1) and (Idea 2) in utilizing prescribed medications and the EHR graph in their predictors.
- **Novel Framework**: To overcome these two limitations, we propose a novel framework HI-DR based on the following two core ideas: (Idea 1) health status-aware attention and (Idea 2) an EHR graph+.
- **Extensive Evaluation**: We validate the effectiveness of HI-DR by conducting extensive experiments on real-world datasets. Above all, HI-DR *dramatically outperforms* all state-of-the-art methods, achieving a gain by up to 18.69% in terms of Jaccard over the best competitor within a reasonably *safe range*.

2 Related Work

Aspect of (Idea 1). VITA (Kim et al. 2024) claims that only a patient's past visits whose diagnoses/procedures are relevant to those of her current visit should be used in the *encoder* to obtain her *patient representation* (*i.e.*, *health status*). Note that the encoder of VITA *cannot* utilize other patients' health records in obtaining the patient's health status. On the other hand, MRSC (Wang et al. 2021) attempts to utilize *only* the medications prescribed in the patient's past health status similar to her current one in its predictor, but it does *not* consider the medications prescribed in *other patients*' health status similar to her current one. DAPSNet (Wu et al. 2023) and PROMISE (Wu et al. 2024) attempt to utilize the medications prescribed in other patients' health status similar to the patient's current one in their predictors, but they *also* utilize the medications prescribed at *all of her past visits*. However, (Idea 1) health status-aware attention utilizes the patient's own past visits only for the *minority* of patients (*spec.*, about 7.16%), achieving the *highest accuracy*; this result will be detailed in Section 4.

Aspect of (Idea 2). Unlike most existing methods that do *not* consider the *degree* of co-prescription, *only* MERITS (Zhang et al. 2021) and REFINE (Bhoi et al. 2024) at-

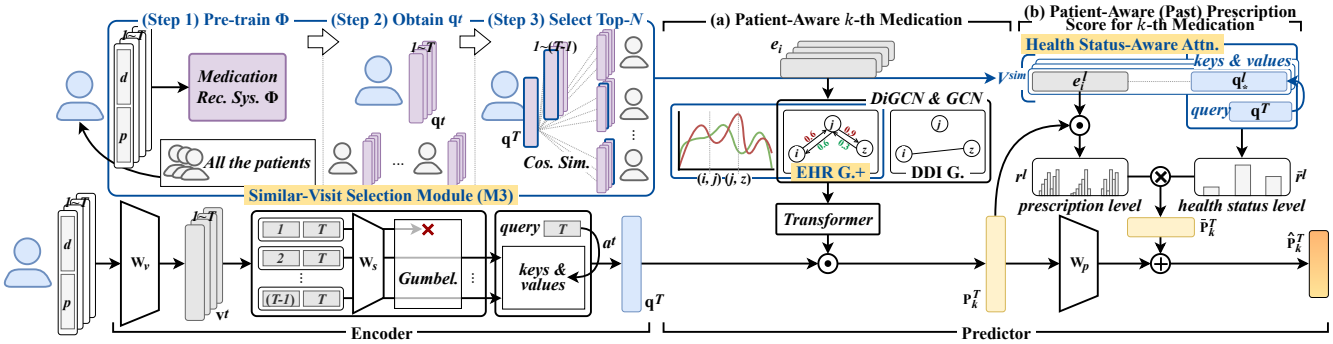

Figure 2: Overview of HI-DR composed of two components: an encoder considering the relevance between the patient's current visit and each of her past ones; a predictor based on (Idea 1) health status-aware attention and (Idea 2) an EHR graph+.

tempt to consider the degree of co-prescription at the *medication pair level* (*i.e.*, the orange line in Figure 1) in the EHR graph via edge weights. However, these efforts *alone* are *not* enough to obtain medication representations that accurately capture the *asymmetric relations* between medications at the *medication level*, which are *quite different* from the symmetric relations at the *medication pair level* (compare the orange and blue (resp. pink) lines in Figure 1). To consider the degree at the medication level, two medications that have ever been co-prescribed can be connected by *two opposite directed edges* in the EHR graph; this inspires the design of (Idea 2) our EHR graph+.

3 Proposed Framework: HI-DR

In this section, we detail a novel medication recommendation framework HI-DR, which enhances its *predictor* via (Idea 1) *health status-aware attention* and (Idea 2) an *EHR graph+*. HI-DR's overall procedure is presented in Figure 2.

3.1 Problem Definition

Definition 1: Patient Health Records. In the MIMIC-III (Johnson et al. 2016) and MIMIC-IV (Johnson et al. 2023) datasets, the health records of each patient x are represented as a sequence of visits $\mathcal{V}_x = [\mathbf{V}_x^1, \cdots, \mathbf{V}_x^{(T-1)}, \mathbf{V}_x^T]$, where $\mathbf{V}_x^{(T-1)}$ denotes the visit $(T-1)$ of patient x. For simplicity, we omit the subscript x indicating a patient, and describe HI-DR by using a single patient case, following (Shang et al. 2019a,b; Wu et al. 2022; Kim et al. 2024). Therefore, the patient's visits are denoted by $\mathcal{V} = [\mathbf{V}^1, \cdots, \mathbf{V}^{(T-1)}, \mathbf{V}^T]$. Each visit $\mathbf{V}^t$ of a patient consists of diagnoses, procedures, and medications given for that patient at t. They are represented as multi-hot vectors $\mathbf{d}^t \in \mathbb{R}^{|\mathcal{D}|}$, $\mathbf{p}^t \in \mathbb{R}^{|\mathcal{P}|}$, and $\mathbf{m}^t \in \mathbb{R}^{|\mathcal{M}|}$ respectively, where $\mathcal{D}$, $\mathcal{P}$, and $\mathcal{M}$ denote sets of all diagnoses, all procedures, and all medications, respectively.

Definition 2: EHR and DDI Graphs. An EHR and a Drug-Drug Interactions (DDI) graphs are denoted by $\mathcal{G}_{EHR} = (\mathcal{M}, \mathcal{E}_{EHR})$ and $\mathcal{G}_{DDI} = (\mathcal{M}, \mathcal{E}_{DDI})$, respectively, where $\mathcal{E}_{EHR}$ and $\mathcal{E}_{DDI}$ represent sets of edges between medications. An edge $(i, j) \in \mathcal{E}_{EHR}$ indicates that two medications i and j have been *co-prescribed at least once* at any visit of *any* patient. An edge $(i, j) \in \mathcal{E}_{DDI}$ indicates that two medications i and j may *cause adverse effects* when they are taken together (Wu et al. 2022; Kim et al. 2024). The same EHR and DDI graphs are shared for all patients.

Medication Recommendation Problem. Given a patient's past health records $[\mathbf{V}^1, \cdots, \mathbf{V}^{(T-1)}]$, the diagnoses $\mathbf{d}^T$ and procedures $\mathbf{p}^T$ at her current visit T, and the EHR and DDI graphs, the goal is to recommend the accurate medications $\hat{\mathcal{M}}^T$ at her current visit. We summarize the key notations used in this paper in Appendix B.

3.2 Key Components

Encoder. The encoder of HI-DR produces a *patient representation* ('patient rep.', in short) $\mathbf{q}^T \in \mathbb{R}^{dim}$, where dim denotes the dimensionality of the representation, indicative of the *patient's current health status* considering *acute or chronic condition*. First, HI-DR concatenates the diagnoses $\mathbf{d}^t$ and procedures $\mathbf{p}^t$ for each visit t of a patient up to her current visit T, and then obtains a representation $\mathbf{v}^t \in \mathbb{R}^{dim}$ at her visit t by feeding the concatenation of two vectors $\mathbf{d}^t$ and $\mathbf{p}^t$ into an embedding layer (formally, $\forall t \in \{1, \cdots, (T-1), T\}$, $\mathbf{v}^t = concat(\mathbf{d}^t, \mathbf{p}^t)\mathbf{W}_v$, where $\mathbf{W}_v \in \mathbb{R}^{(|\mathcal{D}|+|\mathcal{P}|)\times dim}$ denotes a weight matrix of the embedding layer). Next, HI-DR concatenates the representations $\mathbf{v}^t$ of each past visit t and $\mathbf{v}^T$ of the current visit T, and then obtains the probability $s^t \in \mathbb{R}$, which indicates the probability of each past visit t being selected as a past visit *relevant* to the current visit T by feeding the concatenated representation into a Multi-Layer Perceptron (MLP) layer (formally, $\forall t \in \{1, \cdots, (T-1)\}$, $s^t = sigmoid(concat(\mathbf{v}^t, \mathbf{v}^T)\mathbf{W}_s + b_s)$, where $\mathbf{W}_s \in \mathbb{R}^{2dim\times 1}$ and $b_s \in \mathbb{R}$ denote a weight matrix and a bias of the MLP layer, respectively). Subsequently, HI-DR concatenates the probabilities s^t and $(1 - s^t)$ of each past visit t into a two-dimensional vector, and then applies the Gumbel-softmax (Jang, Gu, and Poole 2017; Maddison, Mnih, and Teh 2017) to this vector to determine whether to select each past visit t as a relevant past visit to the current visit T. Given the past visits t selected by the Gumbel-softmax, HI-DR computes the degree of relevance $a^t \in \mathbb{R}$ between the relevant past visits t and the current visit T via the attention

network used in (Kim et al. 2024), which utilizes the representation $\mathbf{v}^T$ of T as the query, and the representations $\mathbf{v}^T$ of the current visit T and $\mathbf{v}^t$ of all the relevant past visits t as keys and values, respectively. Finally, HI-DR obtains the patient rep. $\mathbf{q}^T$ by fusing the representations $\mathbf{v}^t$ and $\mathbf{v}^T$ based on their degree of relevance a^t (formally, $\mathbf{q}^T = \sum_{\mathbf{v}^t \in \{all\_relevant\_past\_visits\} \cup \{\mathbf{v}^T\}} a^t \mathbf{v}^t$) (Kim et al. 2024).

Predictor. Given the patient rep. $\mathbf{q}^T$, the predictor of HI-DR recommends *accurate* medications $\hat{\mathcal{M}}^T$ to the patient via (Idea 1) health status-aware attention and (Idea 2) an EHR graph+. HI-DR predicts the medications to be recommended *one by one*, by initializing $\hat{\mathcal{M}}^T$ as an empty set at the beginning, and repeatedly predicting the k-th ($k \geq 1$) medication until it predicts the <END> token (class), which indicates that all necessary medications have been predicted (Wu et al. 2022; Kim et al. 2024). For simplicity, we describe the predictor of HI-DR via the process of predicting the k-th medication.

Medication Representations $\mathbf{e}_i$. HI-DR first obtains representations $\mathbf{e}_i \in \mathbb{R}^{dim}$ of all medications i, by considering both relations between medications that (1) are *frequently co-prescribed* and (2) have *adverse effects* when taken together (Shang et al. 2019b; Wu et al. 2022; Kim et al. 2024).

EHR Graph+. To *accurately* consider the relations between medications that are *frequently* co-prescribed, HI-DR incorporates the *degree of co-prescription at the medication level* into the original EHR graph $\mathcal{G}_{EHR}$, thereby obtaining an EHR graph+ $\mathcal{G}^+_{EHR} = (\mathcal{M}, \mathcal{E}^+_{EHR})$, where a *directed edge* $(i, j) \in \mathcal{E}^+_{EHR}$ from medication i to medication j *weighted by the degree* to which j is prescribed when i is prescribed (*i.e.*, the blue and pink lines in Figure 1). The same EHR graph+ is shared for all patients.

Then, HI-DR applies a two-layer DiGCN (Tong et al. 2020), which is a GCN applicable to a *directed graph*, to $\mathcal{G}^+_{EHR}$ with randomly initialized medication representations. To consider the relations between medications that have adverse effects when taken together, HI-DR applies a two-layer GCN to the DDI graph $\mathcal{G}_{DDI}$ with randomly initialized medication representations. HI-DR obtains medication representations $\mathbf{e}_i$ by fusing the outputs of the DiGCN and GCN via a subtraction operation per medication, like most existing methods. Medication representations $\mathbf{e}_i$ are obtained once (when $k = 1$) and then reused for predicting the k-th ($k \geq 2$) medication. We will validate whether the EHR graph+ effectively addresses the limitations of EHR graphs used by existing methods in Section 4.

Given the medication representations $\mathbf{e}_i$ and the patient rep. $\mathbf{q}^T$ (obtained from the encoder), HI-DR obtains (a) *representation* $\mathbf{p}^T_k \in \mathbb{R}^{dim}$ and (b) *vector* $\bar{\mathbf{p}}^T_k \in \mathbb{R}^{|\mathcal{M}|}$, and finally predicts the k-th medication as follows:

$$\begin{gathered} argmax_{k \in \{1,2,\cdots,|\mathcal{M}|\}}(\hat{\mathbf{p}}^T_k), \\ \text{where } \hat{\mathbf{p}}^T_k = \lambda_k \{softmax(\mathbf{p}^T_k \mathbf{W}_p + \mathbf{b}_p)\} + (1 - \lambda_k)\bar{\mathbf{p}}^T_k, \end{gathered} \tag{1}$$

where $\hat{\mathbf{p}}^T_k \in \mathbb{R}^{|\mathcal{M}|}$ denotes the score of each medication being predicted as the k-th medication; $\lambda_k \in \mathbb{R}$ denotes a learnable parameter that balances $\mathbf{p}^T_k$ and $\bar{\mathbf{p}}^T_k$; $\mathbf{W}_p \in \mathbb{R}^{dim \times |\mathcal{M}|}$ and $\mathbf{b}_p \in \mathbb{R}^{|\mathcal{M}|}$ denote a weight matrix and a bias vector, respectively. The specific process of obtaining (a) $\mathbf{p}^T_k$ and (b) $\bar{\mathbf{p}}^T_k$ is as follows.

(a) Representation of Patient-Aware k-th Medication $\mathbf{p}^T_k$. The representation $\mathbf{p}^T_k$ denotes the features of the k-th medication necessary regarding the patient rep. $\mathbf{q}^T$. Suppose that we have previously predicted up to the $(k-1)$-th medication.[3] To obtain $\mathbf{p}^T_k$, HI-DR derives features of the medication which is necessary regarding $\mathbf{q}^T$ but has not been predicted yet by considering features of the previous $(k-1)$ medications (Wu et al. 2022; Kim et al. 2024) as follows:

$$\mathbf{p}^T_k = softmax(\frac{transformer(\mathbf{e}^1, \mathbf{e}^2, \cdots, \mathbf{e}^{(k-1)}) \odot \mathbf{q}^T}{\sqrt{dim}})\mathbf{q}^T, \tag{2}$$

where $transformer(\cdot)$ denotes a transformer-based model used in (Wu et al. 2022; Kim et al. 2024); $\mathbf{e}^{(k-1)}$ denotes a medication representation of the $(k-1)$-th predicted medication; $\odot$ denotes a dot-product.

(b) Patient-Aware (Past) Prescription Score Vector $\bar{\mathbf{p}}^T_k$ for k-th Medication. Suppose that we selected the top-N patient reps. $\mathbf{q}^l_*$ (where $l \in \{1, 2, \cdots, N\}$) *similar* to patient rep. $\mathbf{q}^T$, where the subscript $*$ denotes that $\mathbf{q}^l_*$ can be the patient rep. of either the patient *herself* or *another* patient. Let $\mathcal{V}^{sim}$ denote a set of visits $\mathbf{V}^l_*$ for these patient reps $\mathbf{q}^l_*$. The vector $\bar{\mathbf{p}}^T_k$ provides the probability of each medication prescribed at similar visits $\mathbf{V}^l_* \in \mathcal{V}^{sim}$ to be recommended as the k-th medication.

Health Status-Aware Attention. To obtain $\bar{\mathbf{p}}^T_k$, HI-DR *flexibly* selects $\mathbf{V}^l_*$ *regardless of whose visit they are*, and then utilizes the medications prescribed at $\mathbf{V}^l_*$. Then, how can we effectively obtain $\mathbf{V}^l_*$ regardless of patients?

To answer this question, we might consider the following candidates for similar-visit selection modules: **(M1)** utilizing patient rep. $\mathbf{q}^t$: for every visit t of all patients in the training set, compute the cosine similarity between (1) $\mathbf{q}^T$ and (2) patient rep. $\mathbf{q}^t$ of the visit t, and select the top-N similar visits in *real-time* during training and inference; **(M2)** utilizing multi-hot vector $concat(\mathbf{d}^t, \mathbf{p}^t)$: for diagnoses $\mathbf{d}^t$ and procedures $\mathbf{p}^t$ of every visit t of all patients in the training set, compute the Jaccard similarity between the multi-hot vectors (1) $concat(\mathbf{d}^T, \mathbf{p}^T)$ of a target patient's current visit T and (2) $concat(\mathbf{d}^t, \mathbf{p}^t)$ of the visit t, and select the top-N similar visits for the current visit of all patients *in advance*.

However, there are some *limitations* in the two modules above: module (M1) results in significantly *prolonged* training and inference times since it needs to additionally obtain patient reps. $\mathbf{q}^t$ at all visits of the target patient and all other patients in the training set from the encoder, whenever predicting medications per patient. We highlight that MRSC (Wang et al. 2021), DAPSNet (Wu et al. 2023), and PROMISE (Wu et al. 2024) utilize the methods *similar to module* (M1) in obtaining their vectors (b), *without* addressing any time efficiency concerns; module (M2) does *not* consider past health records of each visit t, so it does *not* distinguish whether diagnoses $\mathbf{d}^t$ and procedures $\mathbf{p}^t$ at each visit t are given for the *acute or chronic conditions*.

[3] When $k = 1$, a randomly initialized <START> token (representation) is used (Wu et al. 2022; Kim et al. 2024).

Therefore, we design a novel similar-visit selection module, *mitigating* the limitations of two modules (M1) and (M2) and *taking their advantages*: **(M3)** utilizing *patient rep.* $\mathbf{q}^t$ obtained from a *pre-trained medication recommendation model* Φ. In (M3), (Step 1) trains model Φ on the training set, and then tunes hyperparameters of model Φ on the validation set; (Step 2) obtains patient reps. $\mathbf{q}^t$ at all visits of all patients from the model Φ fixed in (Step 1); (Step 3) computes the cosine similarity between (1) $\mathbf{q}^T$ and (2) patient rep. $\mathbf{q}^t$ at each of the target patient's past visits and other patients' visits in the training set, and obtains $\mathcal{V}^{sim}$ for the current visit of all patients *in advance*.[4] We will demonstrate the effectiveness of module (M3) in Section 4.

Ultimately, HI-DR is equipped with our a health status-aware attention based on *module* (M3) to obtain the patient-aware (past) prescription score vector $\bar{\mathbf{p}}_k^T$ for the k-th medication. Specifically, HI-DR first obtains $\mathcal{V}^{sim}$ via module (M3). Then, for each medication i prescribed at these similar visits $\mathbf{V}_*^l$, HI-DR computes the probability of the medication i being recommended as the k-th medication, considering the degree of relevance between the patient's current visit and similar visits $\mathbf{V}_*^l$ at two levels (*spec.*, *prescription level* $r_i^l \in \mathbb{R}$ and *health status level* $\bar{r}^l \in \mathbb{R}$) (Wu et al. 2022; Kim et al. 2024). The relevance r_i^l at the prescription level denotes how much each medication i prescribed at similar visits $\mathbf{V}_*^l$ is relevant to (a) the representation of patient-aware k-th medication $\mathbf{p}_k^T$. In other words, $\forall i \in \{medications(\text{`}meds.\text{'}, \text{ in short})\_prescribed\_at\ \mathbf{V}_*^l\}, \forall l = \{1, 2, \cdots, N\}$,

$$r_i^l = \frac{exp((\mathbf{e}_i^l \odot \mathbf{p}_k^T)\sqrt{dim})}{\sum_{j \in \{meds.\_prescribed\_at\ \mathbf{V}_*^l\}} exp((\mathbf{e}_j^l \odot \mathbf{p}_k^T)/\sqrt{dim})}, \quad (3)$$

where $\mathbf{e}_i^l$ denotes the medication representation of medication i prescribed at visit $\mathbf{V}_*^l$. The relevance $\bar{r}^l$ at the health status level indicates how much the health status (*i.e.*, $\mathbf{q}_*^l$) at visits $\mathbf{V}_*^l$ is relevant to that (*i.e.*, $\mathbf{q}^T$) of the patient's current visit T. To achieve this, HI-DR first obtains the patient reps. $\mathbf{q}_*^l$ at visits $\mathbf{V}_*^l$ via the encoder of HI-DR. Then, HI-DR obtains the relevance $\bar{r}^l$ at the health status level via an attention network that utilizes the patient rep. $\mathbf{q}^T$ as a query, and the patient reps. $\mathbf{q}_*^l$ for all visits $\mathbf{V}_*^l$ as keys and values. In other words, $\forall \mathbf{q}_*^l, \forall l = \{1, 2, \cdots, N\}$,

$$\bar{r}^l = \frac{exp((\mathbf{q}^T \mathbf{W}_{\bar{r}} \mathbf{q}_*^{l\intercal})/\sqrt{dim})}{\sum_{f=1}^{N} exp((\mathbf{q}^T \mathbf{W}_{\bar{r}} \mathbf{q}_*^{f\intercal})/\sqrt{dim})}, \quad (4)$$

where $\mathbf{W}_{\bar{r}} \in \mathbb{R}^{dim \times dim}$ denotes a weight matrix of the attention network. After obtaining two relevances r_i^l at a prescription level and $\bar{r}^l$ at a health status level, HI-DR obtains (b) the patient-aware (past) prescription score vector $\bar{\mathbf{p}}_k^T$ for the k-th medication by fusing the two relevances as follows:

$$\bar{\mathbf{p}}_k^T = \sum_{l=1}^{N} \bar{r}^l \mathbf{r}^l \ \ (\mathbf{r}^l \in \mathbb{R}^{|\mathcal{M}|}),$$
$$\text{where } \forall i \in \mathcal{M},\ \mathbf{r}^l = \begin{cases} r_i^l, & if\ i \in \{meds.\_prescribed\_at\ \mathbf{V}_*^l\}, \\ 0, & otherwise. \end{cases} \quad (5)$$

As mentioned earlier, HI-DR finally predicts the k-th medication by fusing (a) representation $\mathbf{p}_k^T$ and (b) vector $\bar{\mathbf{p}}_k^T$ via Eq. (1).

3.3 Training

As in (Wu et al. 2022; Kim et al. 2024), we learn the medication representations $\mathbf{e}_i$ and other learnable parameters of HI-DR by utilizing the cross-entropy loss function for the medications predicted by HI-DR as follows:

$$\mathcal{L} = -\sum_{t=1}^{T} \sum_{i=1}^{|\mathcal{M}|} m_i^t log(\hat{\mathbf{p}}_{k,i}^t), \quad (6)$$

where $m_i^t \in \mathbb{R}$ represents 1 if medication i is (actually) prescribed at visit t, and 0 otherwise.

4 Evaluation

To evaluate the effectiveness of HI-DR, we designed experiments to answer the following key research questions (RQs):

- **RQ1**: Does HI-DR provide *better accuracy* than existing state-of-the-art medication recommender systems and *better safety* than the ground-truth sets?
- **RQ2**: Are HI-DR's two core ideas, *i.e.*, (Idea 1) health status-aware attention and (Idea 2) an EHR graph+, effective in improving accuracy or safety?
- **RQ3**: In HI-DR, how many patients have their own past visits included within the set of similar visits?
- **RQ4**: Does equipping the two core ideas of HI-DR with *existing methods* improve their accuracy and/or safety?

4.1 Experimental Settings

Datasets. We used not only the real-world MIMIC-III (Johnson et al. 2016) dataset, which is commonly utilized in most medication recommendation research, but also its follow-up MIMIC-IV (Johnson et al. 2023) dataset *larger* than MIMIC-III; we filtered out, from these two datasets, the patients who visited only once and the diagnoses, procedures, and medications related only to them, following (Sun et al. 2022; Wu et al. 2022; Yang et al. 2023; Kim et al. 2024). We present statistics of two datasets in Appendix C.

Competitors. We compared HI-DR with the following *ten* competitors: one basic classifier (*spec.*, Nearest as in (Shang et al. 2019b)) and nine state-of-the-art medication recommender systems (*spec.*, RETAIN (Choi et al. 2016b), LEAP (Zhang et al. 2017), GAMENet (Shang et al. 2019b), MRSC (Wang et al. 2021), SafeDrug (Yang et al. 2021), DrugRec (Sun et al. 2022), COGNet (Wu et al. 2022), MoleRec (Yang et al. 2023), and VITA (Kim et al. 2024)).

[4] Here, *any* medication recommendation method that obtains patient reps. can be utilized as model Φ. HI-DR utilizes VITA (Kim et al. 2024), which focuses on obtaining accurate patient reps., as model Φ; however this VITA is equipped with our EHR graph+ instead of its original EHR graph.

Datasets	MIMIC-III				MIMIC-IV			
Metrics	**Jaccard**	**PRAUC**	**F1**	**DDI rate**	**Jaccard**	**PRAUC**	**F1**	**DDI rate**
Nearest	0.3917	0.3813	0.5474	0.0787	0.4523	0.4461	0.6045	0.0913
RETAIN (NeurIPS'16)	0.4701	0.7498	0.6359	0.0844	0.4591	0.7022	0.6004	0.0712
LEAP (KDD'17)	0.4306	0.6379	0.5941	0.0771	0.4273	0.5949	0.5836	0.0693
GAMENet (AAAI'19)	0.4350	0.6754	0.5902	0.0866	0.4600	0.7029	0.6106	0.0791
MRSC (CIKM'21)	0.4856	0.7413	0.6451	0.0665	0.4487	0.6907	0.6063	0.0951
SafeDrug (IJCAI'21)	0.5060	0.7537	0.6630	0.0661	0.4731	0.6976	0.6251	0.0643
DrugRec (NeurIPS'22)	0.5240	0.7735	0.6861	**0.0616**	0.4380	0.6761	0.5824	**0.0516**
COGNet (WWW'22)	0.5087	0.7567	0.6666	0.0824	0.4943	0.7069	0.6473	0.0914
MoleRec (WWW'23)	0.5292	0.7740	0.6835	0.0749	0.4901	0.7252	0.6431	0.0741
VITA (AAAI'24)	0.5282	0.7673	0.6815	0.0803	0.5218	0.7148	0.6685	0.0910
HI-DR	**0.6281**	**0.8101**	**0.7418**	0.0802	**0.6119**	**0.7800**	**0.7215**	0.0912

Table 1: Accuracies of ten competitors and HI-DR. The ground-truth DDI rate is 0.0808 on MIMIC-III and 0.0909 on MIMIC-IV, respectively.

In contrast, we were *not* able to conduct experiments on MERITS (Zhang et al. 2021), REFINE (Bhoi et al. 2024), DAPSNet (Wu et al. 2023), and PROMISE (Wu et al. 2024), because MERITS requires *private information* (*spec.*, time-series data of blood glucose measurements), DAPSNet includes a GitHub link but the code does *not* exist, and REFINE and PROMISE do *not* provide any link or code. Note, however, we applied the *concepts* of these methods to HI-DR and presented the comparative results in Table 2.

Evaluation Protocols. We randomly split all patients in each dataset into training $(4/6)$, validation $(1/6)$, and test $(1/6)$ sets; we measured *accuracy* of each method in terms of Jaccard, PRAUC, and F1; we measured *safety* with the DDI rate, as in (Wu et al. 2022; Kim et al. 2024). The *lower* the score of the DDI rate, the *safer* (*i.e.*, *lower* potential for adverse effects) the recommended medications. We conducted five independent evaluations and reported the mean and standard deviation of their results in the case of tables; in Tables 1 and 2, all improvements are *statistically significant* with a p-value ≤ 0.001. We carefully tuned the hyperparameters of all methods via extensive grid search; please refer to Appendix C for specific hyperparameter values.

4.2 Results and Analysis

Due to space limitations, the results for RQ2 and RQ3 on MIMIC-IV are shown in Appendix D, since these results showed *similar* tendencies as those on MIMIC-III. The best and the second-best results in each column (*i.e.*, metric) in the following tables are in **bold** and underline, respectively.

RQ1: Comparison with Ten Competitors. The results are shown in Table 1. Our findings are summarized as follows.

Accuracy (*i.e.*, Jaccard, PRAUC, and F1). HI-DR *significantly* outperforms *all* competitors on *all* datasets for *all* metrics. Specifically, on MIMIC-III and MIMIC-IV, HI-DR outperforms the best competitors, *i.e.*, MoleRec and VITA, by up to 18.69% and 17.26% respectively in terms of Jaccard. These are *dramatic improvements* in the sense that on MIMIC-III and MIMIC-IV, existing state-of-the-art methods outperform their best competitors by a maximum of up to only 3.73% and 5.67% in terms of Jaccard in their original papers (Wang et al. 2021; Yang et al. 2021; Sun et al. 2022; Wu et al. 2022; Yang et al. 2023; Kim et al. 2024).

Safety (*i.e.*, the DDI rate). As mentioned in Section 1, to recommend the *ground-truth sets* to patients, the DDI rate is *inevitably increased*. For all the ground-truth sets, the average DDI rate is 0.0808 (resp. 0.0909) on MIMIC-III (resp. MIMIC-IV). HI-DR's average DDI rate is 0.0802 (resp. 0.0912) on MIMIC-III (resp. MIMIC-IV), which is *lower* than (resp. *comparable* to) that of the ground-truth set. That is, HI-DR is *significant* in that it *maintains its DDI rate within the range of the ground-truth sets*, thanks to our EHR graph+, while *significantly improving accuracy*.

RQ2: Ablation Study. To verify the effectiveness of HI-DR's two ideas, (Idea 1) *health status-aware attention* and (Idea 2) an *EHR graph+*, in terms of accuracy and safety, respectively, we compared HI-DR and its nine variants.

HI-DR's variants w.r.t. (Idea 1). (1) $-$HA utilizes *only* the patient's *all* past visits like most existing methods, *instead of* using health status-aware attention. (2) $-\text{HA}_{\text{others top-}N}$ selects the top-N similar visits only from *other patients*' visits in the health status-aware attention, and utilizes the selected visits *together* with medications prescribed at the patient's *all* past visits (*i.e.*, *similar* to DAPSNet and PROMISE). (3) $-\text{HA}_{\text{own top-}N}$ selects the top-N similar visits only from the *patient's own* past visits in the health status-aware attention; here, the number of the similar visits can be fewer than N if a patient has fewer than N past visits. (4) $-\text{HA}_{\text{(M1) rep.}}$ (resp. (5) $-\text{HA}_{\text{(M2) mul.}}$) selects the top-$N$ similar visits based on the *similar-visit selection module* (M1) (resp. (M2)).

HI-DR's variants w.r.t. (Idea 2). (6) $-$G does *not* utilize the EHR graph+ *at all* (*spec.*, applies a zero-layer DiGCN to the EHR graph+). (7) $-$G+ replaces the EHR graph+ with the *original* EHR graph. (8) $-\text{G+}_{\text{pair}}$ replaces the EHR graph+ with a *weighted undirected* EHR graph, which assigns each edge with a weight based on the degree of co-prescription at the *medication pair level* (*i.e.*, the orange line in Figure 1 *similar* to MERITS and REFINE). (9) $-\text{G+}_{\text{GAT}}$ utilizes a two-layer Graph Attention Network (Velickovic et al. 2018) to the *original* EHR graph.

For HI-DR and all its nine variants, we set $N = 3$, which showed the best accuracy on HI-DR. Table 2 shows their accuracies and safeties. Below, we summarize the findings in Table 2.

Metrics	Jaccard	PRAUC	F1	DDI rate
HI-DR	0.6281	0.8101	0.7418	0.0802
−HA	0.5298	0.7725	0.6838	**0.0748**
−HA$_{\text{others top-}N}$	0.5368	0.7608	0.6874	0.0792
−HA$_{\text{own top-}N}$	0.5338	0.7680	0.6780	0.0798
−HA$_{\text{(M1) rep.}}$	-	-	-	-
−HA$_{\text{(M2) mul.}}$	0.5361	0.7531	0.6858	0.0749
−G	**0.6312**	**0.8116**	**0.7430**	0.0845
−G+	0.6148	0.8026	0.7322	0.0855
−G+$_{\text{pair}}$	0.6233	0.8054	0.7383	0.0835
−G+$_{\text{GAT}}$	0.5199	0.7584	0.6729	0.0802

Table 2: The effects of health status-aware attention and an EHR graph+. The ground-truth DDI rate is 0.0808.

Effectiveness of (Idea 1). We observed that HI-DR *consistently* outperforms *all* its variants for *all* accuracy metrics. This result supports the effectiveness of health status-aware attention that (i) selects top-N similar visits (ii) regardless of whose visit they are. Note that, despite (i) selecting top-N similar visits, both $-\text{HA}_{\text{others top-}N}$ and $-\text{HA}_{\text{own top-}N}$ show *comparable* to or *even lower* accuracy than $-$HA for all accuracy metrics. This is because $-\text{HA}_{\text{others top-}N}$ may utilize medications prescribed in the patient's past health status *dissimilar* to her current one; $-\text{HA}_{\text{own top-}N}$ is *not* able to fully benefit from the advantages of selecting top-N similar visits, due to the *limited number* of a patient's *own* visit (*spec.*, 2.37 on average as presented in the table of statistics on MIMIC-III in Appendix C). Although $-\text{HA}_{\text{(M2) mul.}}$ (i) selects top-N similar visits (ii) regardless of whose visit they are, it mostly shows accuracy *lower* than or *comparable* to $-$HA, $-\text{HA}_{\text{others top-}N}$, and $-\text{HA}_{\text{own top-}N}$. Because the patients' acute/chronic condition is *very important* due to its influence on medication prescription, selecting top-N similar visits based *only* on the diagnoses and procedures given at *each* visit in $-\text{HA}_{\text{(M2) mul.}}$ makes it difficult to provide the best prescription. Meanwhile, we could *not* show the result for $-\text{HA}_{\text{(M1) rep.}}$, as its training does not finish within reasonable time (*spec.*, about 259,200 seconds per iteration), which is utilized by MRSC, DAPSNet, and PROMISE.

Effectiveness of (Idea 2). We observed that $-$G+ shows the *lowest* safety (*i.e.*, the *highest* DDI rate), followed by an improvement in safety in the order of $-$G, $-\text{G+}_{\text{pair}}$, and HI-DR. These results demonstrate that, in recommending *safe* medications, (i) it is preferable *not to use the original EHR graph*, despite its common use as a *de facto standard*; however, (ii) an EHR graph that includes the degree of co-prescription has a *positive effect*; moreover, (iii) this effectiveness is *further enhanced* when the degree of co-prescription is considered at the *medication level*, rather than at the medication pair level. Additionally, we verify whether learning the degree of co-prescription via a deep-learning model, by considering even the latent aspects, allows for a more-accurate reflection of these degrees, *i.e.*, by comparing HI-DR and $-\text{G+}_{\text{GAT}}$. As shown in Table 2, the safeties of HI-DR and $-\text{G+}_{\text{GAT}}$ are the same (0.0802), achieving the *highest* safety (*i.e.*, the *lowest* DDI rate) among variants w.r.t. (Idea 2) an EHR graph+; note, however, the accuracy of $-\text{G+}_{\text{GAT}}$ is *significantly lower* than HI-DR. These results demonstrate the ability of our EHR graph+ to reduce adverse effects *without compromising accuracy*.

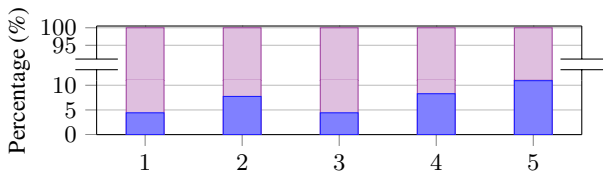

Figure 3: The percentage of patients whose own past visits are included (resp. not included) in the top-N similar visits, indicated by the blue (resp. purple) bar.

RQ3: Analysis of Similar Visits. We investigate how many patients have *their own past visits* included among the top-N similar visits. To do this, for every patient, we first computed the cosine similarity between the patient rep. (*i.e.*, $\mathbf{q}^T$) for the patient's current visit and that (*i.e.*, $\mathbf{q}^t$) for every past visit of hers, and got the average over all these similarities. Then, we sorted all the patients in ascending order of their average similarities, and divided them into five intervals, each containing an equal number of patients. Lastly, for each interval, we computed the percentage of patients whose own past visits are included in the top-N similar visits, which are shown in Figure 3. As expected, interval 5 (a group of patients whose past visits are *most similar* to their current ones) shows the *highest* percentage of patients whose past visits are included in the top-N similar visits. However, note that the health status-aware attention does *not* select *any* of the patient's own past visits for the *majority* of patients (*spec.*, about 92.84%); this further supports our claim that selecting similar visits should *not* be limited *solely* to the patient's *own* past visits like MRSC, *nor* should it necessarily include *all* such visits like DAPSNet and PROMISE, as drawn from the superior performance of HI-DR.

RQ4: Compatibility of HI-DR's Two Core Ideas. HI-DR's two core ideas, *i.e.*, health status-aware attention and an EHR graph+, can be *orthogonally combined with most existing methods*, thereby *improving* their accuracy and/or safety. Unfortunately, due to space limitations, we were not able to include the results for RQ4 in this subsection. Please refer to Appendix D for the detailed results and findings.

5 Conclusions

We proposed a novel framework of medication recommendation, HI-DR, based on (Idea 1) *health status-aware attention*, which *flexibly* selects the visits with the health status similar to the patient's current one, *regardless of* whether they are the *patient's own* or *others'*, and then utilizes only the medications prescribed in these visits, and (Idea 2) an *EHR graph+*, which tackles the original EHR graph, despite its common use as a *de facto standard*, and considers the *degree* to which one medication is prescribed when the other is prescribed, for two co-prescribed medications. Extensive experiments demonstrated that (1) HI-DR is *significantly* more accurate than ten state-of-the-art competitors within a reasonably *safe range*; (2) each of our core ideas is effective in improving accuracy and safety.

Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2022-0-00352 and No. RS-2022-00155586).

References

Bhoi, S.; Lee, M. L.; Hsu, W.; and Tan, N. C. 2024. REFINE: a fine-grained medication recommendation system using deep learning and personalized drug interaction modeling. *Advances in Neural Information Processing Systems*, 36.

Cho, K.; van Merriënboer, B.; Bahdanau, D.; and Bengio, Y. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In *Proceedings of the SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111.

Choi, E.; Bahadori, M. T.; Schuetz, A.; Stewart, W. F.; and Sun, J. 2016a. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. In *Machine Learning for Healthcare Conference*, 301–318.

Choi, E.; Bahadori, M. T.; Sun, J.; Kulas, J.; Schuetz, A.; and Stewart, W. F. 2016b. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism. *Advances in neural information processing systems*, 29.

Gong, F.; Wang, M.; Wang, H.; Wang, S.; and Liu, M. 2021. SMR: Medical Knowledge Graph Embedding for Safe Medicine Recommendation. *Big Data Research*, 23: 100174.

Jang, E.; Gu, S.; and Poole, B. 2017. Categorical Reparameterization with Gumbel-Softmax. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Johnson, A.; Bulgarelli, L.; Pollard, T.; Horng, S.; Celi, L. A.; and Mark, R. 2023. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific data*, 10(1): 1.

Johnson, A. E.; Pollard, T. J.; Shen, L.; Lehman, L.-w. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1): 1–9.

Kim, T.; Heo, J.; Kim, H.; Shin, K.; and Kim, S.-W. 2024. VITA: 'Carefully Chosen and Weighted Less' Is Better in Medication Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 8600–8607.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Ma, X.; Wang, Y.; Chu, X.; Ma, L.; Tang, W.; Zhao, J.; Yuan, Y.; and Wang, G. 2022. Patient Health Representation Learning via Correlational Sparse Prior of Medical Features. *IEEE Transactions on Knowledge and Data Engineering*, 35(11): 11769–11783.

Maddison, C. J.; Mnih, A.; and Teh, Y. W. 2017. The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Shang, J.; Ma, T.; Xiao, C.; and Sun, J. 2019a. Pre-training of Graph Augmented Transformers for Medication Recommendation. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 5953–5959.

Shang, J.; Xiao, C.; Ma, T.; Li, H.; and Sun, J. 2019b. GAMENet: Graph Augmented MEmory Networks for Recommending Medication Combination. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 1126–1133.

Sun, H.; Xie, S.; Li, S.; Chen, Y.; Wen, J.-R.; and Yan, R. 2022. Debiased, Longitudinal and Coordinated Drug Recommendation through Multi-Visit Clinic Records. *Advances in Neural Information Processing Systems*, 35: 27837–27849.

Tong, Z.; Liang, Y.; Sun, C.; Li, X.; Rosenblum, D.; and Lim, A. 2020. Digraph Inception Convolutional Networks. *Advances in Neural Information Processing Systems*, 33: 17907–17918.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph Attention Networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Wang, L.; Zhang, W.; He, X.; and Zha, H. 2018. Personalized Prescription for Comorbidity. In *Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA)*, 3–19.

Wang, Y.; Chen, W.; Pi, D.; Yue, L.; Xu, M.; and Li, X. 2021. Multi-hop Reading on Memory Neural Network with Selective Coverage for Medication Recommendation. In *Proceedings of the ACM International Conference on Information and Knowledge Management (ACM CIKM)*, 2020–2029.

Wu, J.; Dong, Y.; Gao, Z.; Gong, T.; and Li, C. 2023. Dual attention and patient similarity network for drug recommendation. *Bioinformatics*, 39(1): btad003.

Wu, J.; Yu, X.; He, K.; Gao, Z.; and Gong, T. 2024. PROMISE: A pre-trained knowledge-infused multimodal representation learning framework for medication recommendation. *Information Processing & Management*, 61(4): 103758.

Wu, L.; Yuliang, S.; Lin, C.; Yongqing, Z.; and Zhongmin, Y. 2020. Predicting Prescriptions via DSCA-Dual Sequences with Cross Attention Network. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 615–622.

Wu, R.; Qiu, Z.; Jiang, J.; Qi, G.; and Wu, X. 2022. Conditional Generation Net for Medication Recommendation. In *Proceedings of the ACM Web Conference (ACM WWW)*, 935–945.

Yang, C.; Xiao, C.; Ma, F.; Glass, L.; and Sun, J. 2021. SafeDrug: Dual Molecular Graph Encoders for Recommending Effective and Safe Drug Combinations. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 3735–3741.

Yang, N.; Zeng, K.; Wu, Q.; and Yan, J. 2023. MoleRec: Combinatorial drug recommendation with substructure-aware molecular representation learning. In *Proceedings of the ACM Web Conference (ACM WWW)*, 4075–4085.

Zhang, S.; Li, J.; Zhou, H.; Zhu, Q.; Zhang, S.; and Wang, D. 2021. MERITS: medication recommendation for chronic disease with irregular time-series. In *IEEE International Conference on Data Mining (ICDM)*, 1481–1486.

Zhang, Y.; Chen, R.; Tang, J.; Stewart, W. F.; and Sun, J. 2017. LEAP: learning to prescribe effective and safe treatment combinations for multimorbidity. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD)*, 1315–1324.

Zheng, Z.; Wang, C.; Xu, T.; Shen, D.; Qin, P.; Huai, B.; Liu, T.; and Chen, E. 2021. Drug Package Recommendation via Interaction-aware Graph Induction. In *Proceedings of the ACM Web Conference (ACM WWW)*, 1284–1295.