FastRx: Exploring Fastformer and Memory-Augmented Graph Neural Networks for Personalized Medication Recommendations

NGUYEN MINH THAO PHAN, Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu City, Taiwan and College of Information and Communication Technology, Can Tho University, Can Tho City, Vietnam
LING CHEN, Institute of Hospital & Health Care Administration, National Yang Ming Chiao Tung University, Taipei City, Taiwan
CHUN-HUNG CHEN, School of Medicine, National Yang Ming Chiao Tung University, Taipei City, Taiwan
WEN-CHIH PENG, Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu City, Taiwan

Personalized medication recommendations aim to suggest a set of medications based on the clinical conditions of a patient. Not only should the patient's diagnosis, procedure, and medication history be considered, but drug-drug interactions (DDIs) must also be taken into account to prevent adverse drug reactions. Although recent studies on medication recommendation have considered DDIs and patient history, personalized disease progression and prescription have not been explicitly modeled. In this work, we proposed FastRx, a Fastformer-based medication recommendation model to capture longitudinality in patient history, in combination with Graph Convolutional Networks (GCNs) to handle DDIs and co-prescribed medications in Electronic Health Records (EHRs). Our extensive experiments on the MIMIC-III dataset demonstrated superior performance of the proposed FastRx over existing state-of-the-art models for medication recommendation. The source code and data used in the experiments are available at https://github.com/pnmthaoct/FastRx.

CCS Concepts: • **Information systems** → **Data mining**; • **Applied computing** → **Health informatics**;

Additional Key Words and Phrases: Medication Recommendation, Electronic Health Records, Graph Convolutional Networks, Attention Mechanism

ACM Reference format:
Nguyen Minh Thao Phan, Ling Chen, Chun-Hung Chen, and Wen-Chih Peng. 2024. FastRx: Exploring Fastformer and Memory-Augmented Graph Neural Networks for Personalized Medication Recommendations. *ACM Trans. Intell. Syst. Technol.* 15, 6, Article 128 (December 2024), 21 pages.
https://doi.org/10.1145/3696111

Authors' Contact Information: Nguyen Minh Thao Phan, Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu City, Taiwan and College of Information and Communication Technology, Can Tho University, Can Tho City, Vietnam; e-mail: pnmthaoct@gmail.com; Ling Chen, Institute of Hospital & Health Care Administration, National Yang Ming Chiao Tung University, Taipei City, Taiwan; e-mail: ling.chen@nycu.edu.tw; Chun-Hung Chen, School of Medicine, National Yang Ming Chiao Tung University, Taipei City, Taiwan; e-mail: jimchen1551.y@nycu.edu.tw; Wen-Chih Peng (corresponding author), Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu City, Taiwan; e-mail: wcpengcs@nycu.edu.tw.

ACM 2157-6912/2024/12-ART128
https://doi.org/10.1145/3696111

1 Introduction

In recent years, **Deep Learning (DL)** has shown great success in medical applications [25]. Specifically, many DL algorithms were developed for medication recommendation based on **Electronic Health Records (EHRs)** to assist effective clinical decision-making. The challenge of manual medication prescribing lies in the extensive volume of a patient's medical history that needs to be considered, as well as the potential **drug–drug interactions (DDIs)** between their current prescriptions [26] and any existing drugs the patient has been taking [15]. To tackle this challenge, medication recommendation models tried to predict medication recommendation according to a patient's specific clinical conditions while avoiding adverse drug interactions [36].

Figure 1 gives an example of EHRs of a patient. The patient visited the hospital three times in total. Each time, a number of diagnoses in **International Classification of Diseases, Ninth Revision (ICD-9)** codes and prescriptions in **Anatomical Therapeutic Chemical (ATC)** codes were recorded by clinical professionals. Repeat diagnoses and medications are marked in red, and related diseases are indicated by arrows. The visualization reveals the interconnections between the patient's clinical conditions and prescribed medications. Note that there may be diseases in common across different visits or diseases related to previously diagnosed conditions. For example, a patient may have a chronic condition that requires ongoing treatment and then develop a new condition related to the original condition. In this case, the doctor may prescribe the same medication for the chronic condition and add a new medication to treat the related condition.

Medication recommendation has drawn increasing attention due to its practical value. DL-based drug recommendation algorithms generally fall into three categories, namely, rule-based, instance-based, and longitudinal approaches. Rule-based and instance-based models [27, 43] utilized a patient's current diagnosis and treatments to make recommendations without considering the progression of the disease over time. Longitudinal models [8, 19, 29, 36, 39, 40] were developed to take advantage of the longitudinal patient history and capture temporal relationships for more accurate recommendations. Existing longitudinal models typically follow a two-stage approach: first, they aggregate available data into a patient-level representation, and second, they make medication recommendations based on that representation.

However, an important gap in existing studies was that they failed to capture the longitudinal relationships amongst the medications of a patient. For instance, patients with long-term diseases needed to take the same drug for a lifetime. Wu et al. [36] performed a statistical analysis on the MIMIC-III dataset and found that most visits were associated with similar drug prescriptions. This forced us to reevaluate how we should use historical data from a medication recommendation perspective. The challenge was to figure out if a past medicine was still helpful. Yang et al. [39] compared consecutive visits in MIMIC-III and found more similarities in medications than in diagnoses. This finding suggested that it may be more important to predict the change in medication.

Furthermore, DDIs should be considered along with clinical conditions in recommending medications [10]. A high DDIs rate may have adverse side effects or interactions between medications [22]. Although doctors have tried to avoid combined medications that might interact, having a model that recommends a list of medicines based on a patient's diagnosis and procedures can help minimize negative drug interactions. This would be a valuable tool for physicians in making decisions. However, within the largest benchmark dataset available to the public, MIMIC-III [14], most patients have around 20 medications on average. Eliminating all potential drug-to-drug interactions remains a challenge. Therefore, it was crucial to consider the relationship between patient conditions and prescribed drugs to provide safe combined recommendations.

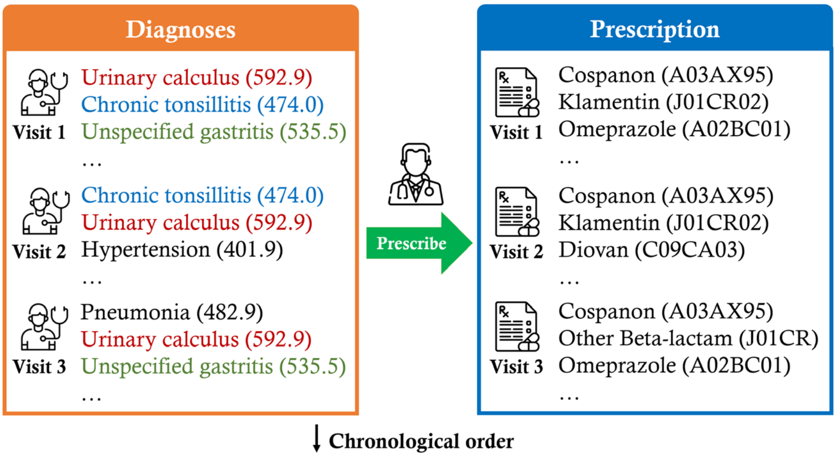

Fig. 1. An example of EHRs of a patient with multiple visits and the associated diagnoses (ICD-9 codes) and prescriptions (ATC codes). ATC, Anatomical Therapeutic Chemical; ICD-9, International Classification of Diseases, Ninth Revision.

To capture the relationships for medical objects, graph representation has been used to model EHRs and DDIs [34, 44]. There has been a significant focus on **Graph Convolutional Networks (GCNs)** [17], which has shown great success in analyzing complex graph structures [6, 42]. Due to their ability to generate meaningful representations of nodes and edges in graphs, these networks are proving to be adequate for tasks involving abundant relational information between various entities, as evidenced by studies [6, 13, 17, 21, 23, 37, 42]. GCN operates on a graph's adjacency matrix to learn relationships between nodes and has shown promising results in medication recommendation [24].

In this article, we present FastRx, a novel personalized medication recommendation model that explores current and historical medical data. Our model comprises three components, including a patient representation encoder employing **Convolutional Neural Networks (CNNs)**, a Fastformer with additive attention to transform diagnosis and procedure codes into patient representations, and a medication graph encoder that employs GCNs to model complex relationships among patient visits to capture the dynamics of drug interactions. Finally, the integrated medication recommender synthesizes a combination of medications, leveraging the patient's longitudinal representation and past medication history.

The key contributions of our work are as follows:

- We designed a cutting-edge DL model, called FastRx, for personalized medication recommendation which is able to capture DDIs and patient similarities in EHRs using GCNs and handle longitudinal patient history using a **one-dimensional (1D)**-CNN and additive attention mechanism.
- We evaluated our method against state-of-the-art models, highlighting FastRx's superior performance using standard medication recommendation evaluation metrics, such as the DDI rate, Jaccard, F1, and PRAUC scores.
- We conducted an ablation study to justify the efficacy of each component in FastRx and demonstrated its practical utility with a case study.

The subsequent sections of this article are organized as follows. Section 2 provides a concise overview of current state-of-the-art approaches and the application of DL in the recommendation of medications. The problem is formulated in Section 3 before exploring our proposed approach in Section 4. To demonstrate the effectiveness of our model, we conducted extensive experiments in Section 5. Finally, we summarize the findings and implications of this work in Section 6.

2 Related Work

Medication recommendation is a crucial aspect of patient care and has gained increasing attention in recent years. Existing studies were classified into rule-based, instance-based, or longitudinal approaches.

Rule-based approach [1, 7, 8, 12, 18, 27] relied on a pre-established set of rules or guidelines to make medication recommendations. They were commonly used in healthcare settings as they provided a structured and systematic way to guide medication decisions. For example, Lakkaraju et al. [18] constructed a set of if-then-else rules for medication decision-making. However, rule-based models did not take into account the unique characteristics of individual patients and were unable to capture a patient's condition or treatment plan over time.

Instance-based approach [11, 43] learned patient characteristics from data to make recommendations, rather than relying on predetermined rules or guidelines. Although this approach allowed for more personalized and flexible medication recommendations, its models did not consider the longitudinal aspect of patient conditions.

Longitudinal approach [4, 29, 33, 34, 36, 38–40] considered a patient's medical history over time. These methods attempted to monitor and assess changes in a patient's health or treatment plan to provide more informed and dynamic recommendations. For a diverse range of clinical prediction tasks, a few studies [4, 8], such as RETAIN, represented longitudinal patient history using **Recurrent Neural Networks (RNNs)**. GAMENet [29] and DMNP [19] designed memory-based networks using RNNs to address dependencies among longitudinal medical codes. MICRON [39] was the first to specifically model patient conditions and predict medication changes, but did not consider the relationships between disease and medications, nor among medications themselves. SafeDrug [40] proposed to use a DDI metric to capture DDIs explicitly and combine it with a molecular graph for medication recommendation. COGNet [36] explicitly models the relationship between medication recommendations for the same patient, based on an EHR graph, a DDI graph, and patient conditions.

The key difference between the proposed model and the existing methods was that our approach explicitly modeled the patient's history for personalized medication recommendations. In fact, we observed a significant influence from recent visits to the medication prescribed on our dataset.

3 Problem Formulation

3.1 EHRs

EHRs are digital versions of a patient's health records that may contain a patient's medical history, diagnoses, medications, treatment plans, allergies, radiology images, and laboratory and test results. These data could provide valuable resources for healthcare professionals to make informed decisions about a patient's care. In this study, we used the diagnosis, procedure, and medication information found in a patient's EHR.

Let a collection of EHRs of patient i be $\mathbf{X}_i = \{\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}, ..., \mathbf{x}_i^{(N_i)}\}$, where N_i is the number of visits for patient i. For each $\mathbf{x}_i^{(t)}$ is a set of $\{\mathbf{d}_i^{(t)}, \mathbf{p}_i^{(t)}, \mathbf{m}_i^{(t)}\}$, where $\mathbf{d}_i^{(t)}, \mathbf{p}_i^{(t)}, \mathbf{m}_i^{(t)}$ refer to the diagnosis, procedure, and medication of the tth visit of patient i, respectively. Let $\mathcal{D} = \{\mathbf{d}_i^{(1)}, \mathbf{d}_i^{(2)}, ..., \mathbf{d}_i^{(N_D)}\}$ be a set of diagnoses of size N_D, $\mathcal{P} = \{\mathbf{p}_i^{(1)}, \mathbf{p}_i^{(2)}, ..., \mathbf{p}_i^{(N_P)}\}$ be a set of procedures of size N_P, and $\mathcal{M} = \{\mathbf{m}_i^{(1)}, \mathbf{m}_i^{(2)}, ..., \mathbf{m}_i^{(N_M)}\}$ be a set of medications of size N_M.

Table 1. Description of Notations Used in FastRx

Notation	Explanation
$\mathbf{X}_i$	The clinical documentaries of patient i
$\mathbf{x}_i^{(n)}$	The visit n of patient i
N_i	Number of visits of patient i
$\mathcal{D}, \mathcal{P}, \mathcal{M}$	Diagnoses, Procedures, and Medication set
$\mathbf{c}_i$	The patient feature i
$\mathbf{h}_i^{(n)}$	The patient representation i of visit of patient n
$\mathbf{G}_e, \mathbf{G}_d$	EHR Graph and DDI Graph, respectively
$\mathcal{E}_e, \mathcal{E}_d$	Edge set of graph $\mathbf{G}_e, \mathbf{G}_d$, respectively
$\mathbf{A}_e, \mathbf{A}_d$	Adjacency matrix of $\mathbf{G}_e, \mathbf{G}_d$, respectively
$\hat{\mathbf{O}}^{(i)}$	The model output
$\hat{\mathbf{M}}_i$	The recommended medication list
$\mathbf{M}_i$	The ground truth medication list

3.2 EHR and DDI Graphs

A graph $\mathbf{G} = \{\mathcal{V}, \mathcal{E}\}$ can be formally defined by a set of vertices $\mathcal{V}$ and edges $\mathcal{E}$. Here, we define an EHR graph $\mathbf{G}_e = \{\mathcal{M}, \mathcal{E}_e\}$, where an adjacency matrix $\mathbf{A}_e \in \mathbb{R}^{|\mathcal{M}|\times|\mathcal{M}|}$ and $\mathbf{A}_e[i, j] = 1$ indicates medication m_i and m_j were prescribed at the same visit; otherwise, $\mathbf{A}_e[i, j] = 0$.

Let $\mathbf{G}_d = \{\mathcal{M}, \mathcal{E}_d\}$ be a DDI graph, where an adjacency matrix $\mathbf{A}_d \in \mathbb{R}^{|\mathcal{M}|\times|\mathcal{M}|}$ and $\mathbf{A}_d[i, j] = 1$ indicates that the medication m_i and m_j had a DDI; otherwise, $\mathbf{A}_d[i, j] = 0$. Both the EHR and DDI graphs are the same for all patients.

3.3 Medication Recommendation Problem

Given a patient's current and historical visits $\mathbf{X}_i = \{\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}, ..., \mathbf{x}_i^{(t)}, ..., \mathbf{x}_i^{(N_i)}\}$, where $\mathbf{x}_i^{(t)} = \{\mathbf{d}_i^{(t)}, \mathbf{p}_i^{(t)}, \mathbf{m}_i^{(t)}\}$, $\mathbf{d}_i^{(t)} \in \mathcal{D}$, $\mathbf{p}_i^{(t)} \in \mathcal{P}$, $\mathbf{m}_i^{(t)} \in \mathcal{M}$, an EHR graph $\mathbf{G}_e$ and a DDI graph $\mathbf{G}_d$, the goal is to train a model that can recommend the proper medication combination $\mathbf{m}_i^{N_i+1}$ for the patient i.

4 Methodology

4.1 Model Architecture

Our proposed model, FastRx (Figure 2), comprises three modules: a patient representation encoder, a medication graph encoder, and an integrated medication recommender. The patient representation encoder captures longitudinal changes in patient conditions from patient diagnoses and procedures from consecutive visits. We propose a hierarchical dependence learning method to capture dependence between patients and medications at both the global and local levels. Specifically, on the one hand, the model learns the local information of the patient record using a sliding window in 1D-CNN. On the other hand, we use an adapted Fastformer-based learning module [35] to learn the global representation of the entire patient record through an additive attention mechanism. By learning the global and local hierarchical dependencies mentioned above, we can learn accurate information about patients. The medication graph encoder, on the other hand, encodes drug combinations from the EHR and DDI using GCNs. Finally, the integrated medication recommender integrates the outputs from the patient representation encoder and medication graph encoder and generates the final recommendation on drug combinations. Detailed explanations for each module follow in this section. Table 1 summarizes the notations used in FastRx.

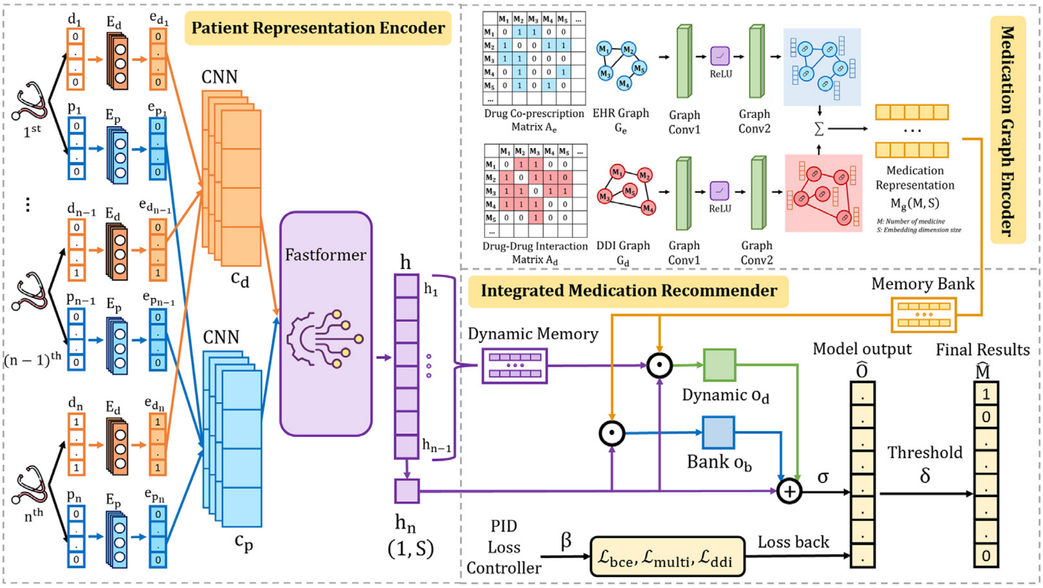

Fig. 2. The proposed model architecture, FastRx, includes three components: patient representation encoder (1), medication graph encoder (2), and integrated medication recommender (3).

4.2 Patient Representation Encoder

To encode the longitudinal representation of the patient, we first extracted diagnosis and procedure codes for each patient's visit. Given multi-hot diagnosis and procedure vectors $\mathbf{d}^{(t)} \in \{0, 1\}^{|\mathcal{D}|+1}$, $\mathbf{p}^{(t)} \in \{0, 1\}^{|\mathcal{P}|+1}$, the diagnosis and procedure codes of a visit t were projected into a corresponding embedding space using the vector-matrix dot product. Following previous studies [29, 43], we adopted the mapping matrices to transform the diagnosis and procedure vectors into embedding spaces. Specifically, we designed mapping matrices, $\mathbf{E}_d \in \mathbb{R}^{(|\mathcal{D}|+1)\times s}$, and $\mathbf{E}_p \in \mathbb{R}^{(|\mathcal{P}|+1)\times s}$, where each row records an embedding vector for a specific diagnosis or procedure, and s indicates the dimension of the embedding space. $\mathbf{E}_d$ and $\mathbf{E}_p$ are learnable and shared between each visit and patient during training. These diagnostic and procedure embedding vectors from a single visit were then concatenated into a combined patient input representation for the visit.

$$\mathbf{e}_{d\,i} = \left\langle \mathbf{d}^{(t)} \mathbf{E}_d \right\rangle \tag{1}$$

$$\mathbf{e}_{p_i} = \left\langle \mathbf{p}^{(t)} \mathbf{E}_p \right\rangle \tag{2}$$

where $\mathbf{e}_{d\,i}$, $\mathbf{e}_{p_i} \in \mathbb{R}^s$ are averaged using the operator $\langle\cdot\rangle$.

We note that clinicians generally consider a patient's historical diagnoses, procedures, drug interactions, and prescriptions from past visits, in addition to the current visit's diagnosis and procedures, when determining the suitable medications for the present visit. Moreover, the significance of this information may vary based on its relevance to the current visit's diagnosis and procedures.

Unlike the medical prediction models used RNNs [5, 8], we applied 1D-CNN to learn representations by local context of the diagnosis and procedure codes of a patient's visit n. This decision is based on the observation that doctors often prioritize recent visits over those from a distant past when caring for a patient in the hospital. For example, during a patient visit i, the diagnoses are as follows: "6930 27652 9912 3893," indicating that the patient suffers from dermatitis as drugs taken internally (6930) and hypovolemia (27652) and needs to be done immunization for allergy (9912) and venous catheterization, not elsewhere classified (3893). Local contexts of "27652" such as "6930" and "9912 3893" are significant in modeling the relationships among these codes in the medical context. Therefore, we employed 1D-CNN to obtain the contextual representations. This layer identifies the interdependencies among consecutive visits by processing the multivariate sequence of diagnoses and procedures embedding, $[\mathbf{e}_{d1}, \mathbf{e}_{d2}, ..., \mathbf{e}_{dN}]$, $[\mathbf{e}_{p_1}, \mathbf{e}_{p_2}, ..., \mathbf{e}_{p_N}] \in \mathbb{R}^{N \times s}$, respectively, where N is the number of visits per patient. The process of embedding these dependencies unfolds as follows:

$$\mathbf{c}_d = ReLU(CNN_{1D}([\mathbf{e}_{d1}, \mathbf{e}_{d2}, ..., \mathbf{e}_{dN}])) \tag{3}$$

$$\mathbf{c}_p = ReLU(CNN_{1D}([\mathbf{e}_{p_1}, \mathbf{e}_{p_2}, ..., \mathbf{e}_{p_N}])) \tag{4}$$

where $\mathbf{c}_d, \mathbf{c}_p \in \mathbb{R}^{N \times s}$ are the output of the hidden layer of the 1D-CNN network.

Each patient visit does not adhere to a specific order in diagnosis and medication codes. Although conventional order-based models such as RNNs [3] are not suitable, the Transformer model enables unordered input and interoperability through a multihead self-attention mechanism, allowing intuitive assessment of interactions between distinct medical records. These self-attention mechanisms enable the dynamic weighing of the significance of various relationships within a sequence with respect to each other. It is crucial to identify correlations between various occurrences of diagnosis and treatment.

Inspired by the remarkable achievements of the Transformer-based model [28], we adopted the foundational Transformer architecture and employed a pre-trained Transformer-based module. An additive attention mechanism is applied to model global contexts and then further transform each token representation based on its interaction with global context representations. We chose the Fastformer, an efficient Transformer model that can accomplish effective context modeling with linear complexity. It first transforms the input embedding matrix, which is extracted by the 1D-CNN layer, into query, key, and value sequences. Fastformer, building upon the Transformer's self-attention mechanism, innovatively refines the attention process to achieve a more efficient computation, particularly for the extensive sequences typical in patient records.

In this way, the input data for each patient can be represented by $\mathbf{c}_d$ and $\mathbf{c}_p$. We used a Fastformer network to learn the interactions among medical ontologies with concatenation operation (||) as follows:

$$\mathbf{h} = Fastformer(\mathbf{c}_d \,||\, \mathbf{c}_p) \tag{5}$$

where $\mathbf{h} \in \mathbb{R}^{N \times s'}$, with s' as the dimension of the Fastformer output embeddings, represents the latent space embedding of the input data, capturing the global dependencies within and between the sequences.

In contrast to the traditional attention mechanism, where the complexity scales quadratically with the sequence length, the Fastformer introduces a linear complexity alternative. It achieves this by first computing a global query and key that distill the input sequence into a more manageable representation:

$$Q_g = \sum_{i=1}^{n} \text{softmax}(W_q \cdot x_i) \odot x_i \tag{6}$$

$$K_g = \sum_{i=1}^{n} \text{softmax}(W_k \cdot x_i) \odot x_i \tag{7}$$

Here, W_q and W_k are trainable weight matrices for the query and key, respectively, and x_i denotes the embedding of the ith token in the sequence. The element-wise multiplication ($\odot$) with the softmax-normalized weights allows the model to emphasize the most significant tokens, effectively summarizing the entire sequence into global query and key vectors. These vectors capture the essence of the sequence, enabling the model to process long inputs more efficiently.

Subsequently, the Fastformer applies an element-wise addition between the global query and the individual token embeddings, followed by a scaled dot-product with the global key, to produce the final attention output:

$$A_i = \text{softmax}\left(\frac{(Q_g + x_i) \cdot K_g}{\sqrt{s'}}\right) V \tag{8}$$

where A_i is the attention score for the ith token, V is the value vector, and s' is the scaling factor equivalent to the dimension of the token embeddings. This additive attention mechanism allows the Fastformer to attend to global information without the computational burden of traditional self-attention.

By utilizing the Fastformer's efficient attention mechanism, our approach can capture global dependencies within the patient's medical history without being hindered by the sequence length, thereby addressing the Transformer's bottleneck in handling long sequences. This latent representation, $\mathbf{h}$, encapsulates the comprehensive interaction patterns among diagnoses and procedures, forming a foundational element for the subsequent medication recommendation process.

4.3 Medication Graph Encoder

To encode the relationships among the medications found in the EHRs, we adopted the GCN approach of [29, 36]. Two types of relationship were modeled using graph representation, namely the EHR graph and the DDI graph, as defined in Section 3.2. The EHR graph encoded drug combinations that coexist in the same prescription. It was called the EHR graph, since they were built based on consecutive pairs of visits of all patient prescriptions extracted from the EHR data. For example, Amoxicillin/Clavulanic acid and Diclofenac may be prescribed together to treat acute tonsillitis. The DDI graph captures DDIs, namely, known adverse interactions between medications that should not be used together.

Given the input medication features $\mathbf{X} \in \mathbb{R}^{|\mathcal{M}| \times s'}$ and the medication graph adjacency matrix $\mathbf{A}_* \in \mathbb{R}^{|\mathcal{M}| \times |\mathcal{M}|}$, the $GCN(\cdot,\cdot) \in \mathbb{R}^{|\mathcal{M}| \times s'}$ produces the new medication representations as follows:

$$GCN(\mathbf{X}, \mathbf{A}) = \sigma(\hat{\mathbf{D}}^{-\frac{1}{2}}(\mathbf{A} + \mathbf{I}\,)\hat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X}) \tag{9}$$

where $\hat{\mathbf{D}}$ is a diagonal matrix such that $\hat{\mathbf{D}}_{i,i} = \sum_j \mathbf{A}_{i,j}$ and $\mathbf{I}$ is the identity matrix.

On each graph, we then used a two-layer GCN to advance better embeddings for the EHR and DDI, $\mathbf{G}_e$ and $\mathbf{G}_d$, respectively.

$$\mathbf{G}_e = GCN(ReLU(GCN(\mathbf{E}_m, \mathbf{A}_e)) \odot \mathbf{W}^g_e, \mathbf{A}_e) \tag{10}$$

$$\mathbf{G}_d = GCN(ReLU(GCN(\mathbf{E}_m, \mathbf{A}_d)) \odot \mathbf{W}^g_d, \mathbf{A}_d) \tag{11}$$

where $\mathbf{W}^g_* \in \mathbb{R}^{|\mathcal{M}| \times s'}$ stores the learnable parameters, $\mathbf{E}_m \in \mathbb{R}^{|\mathcal{M}| \times s'}$ is embedding matrices for medication representations, and $\odot$ is denoted as the element-wise multiplication operator.

Finally, we obtained the interaction-aware medication representations $\mathbf{M}_g \in \mathbb{R}^{|\mathcal{M}| \times s'}$ by compacting $\mathbf{G}_e$ & $\mathbf{G}_d$

$$\mathbf{M}_g = \mathbf{G}_e - \lambda \mathbf{G}_d \tag{12}$$

where λ is a learnable parameter.

The subtraction between $\mathbf{G}_e$ and $\mathbf{G}_d$ was to obtain a representation that captured the differences between a target medication and a candidate medication. It was defined as the element-wise subtraction between their respective medication representations. By using subtraction to obtain a different representation, the model was able to capture the unique features of each medication and compare them in a more fine-grained manner.

4.4 Integrated Medication Recommender

Inspired by [29], we adapted the **Memory Bank (MB)** and **Dynamic Memory (DM)** mechanism to integrate the outputs from the GCNs of the medication graph encoder and the patient representation encoder. MB consisted of a set of bank memories $\mathbf{o}_b^t$, which are pre-learned embeddings of medications, and DM included a DM $\mathbf{o}_d^t$, a learned embedding of the patient's medical history.

We fused multiple graphs by combining different node embeddings as MB $\mathbf{M}_g^t \in \mathbb{R}^{|M| \times s'}$, where λ is a weighting variable. Simultaneously, using DM $\mathbf{M}_d^t$ to integrate historical patient information, we further included patent history representation $\mathbf{h}^t \in \mathbb{R}^s (1 \leq t < N)$ (keys) obtained from the patient representation encoder module and matching multi-hot medication vector $\mathbf{m}^t$ (values) as key-value forms to capture data from various viewpoints adequately. This type of design makes it possible to find the most comparable patient representation over time and to retrieve the appropriate set of weighted medications.

$$\mathbf{M}_d^t = \{\mathbf{h}^t : \mathbf{m}^t\}_1^{t-1} \tag{13}$$

The memory representation output $\mathbf{o}_d^t$, $\mathbf{o}_b^t \in \mathbb{R}^s$ are computed as follows given the patient representation $\mathbf{h}_n^t$ and the current memory state $\mathbf{M}_d^t$, $\mathbf{M}_g^t$:

$$\mathbf{o}_d^t = \mathbf{M}_g^t(\mathbf{M}_{d,v}^t)\mathrm{Softmax}(\mathbf{M}_{d,k}^t \mathbf{h}_n^t) \tag{14}$$

$$\mathbf{o}_b^t = \mathbf{M}_g^t \mathrm{Softmax}(\mathbf{M}_g^t \mathbf{h}_n^t) \tag{15}$$

where $\mathbf{M}_{d,k}^t = [\mathbf{h}^1, \mathbf{h}^2, ..., \mathbf{h}^{t-1}]$ denotes key vectors $\in \mathbb{R}^{|t-1| \times s'}$ and $\mathbf{M}_{d,v}^t = [\mathbf{m}^1, \mathbf{m}^2, ..., \mathbf{m}^{t-1}]$ denotes value vectors $\in \mathbb{R}^{|t-1| \times |M_N|}$.

During the recommendations, the model used DM to query MB, transforming it into a query vector. This vector computed the attention weights between the query and each bank memory, facilitating the retrieval of pertinent historical information. MB effectively integrates the patient's medical history into the recommendation process, enhancing the precision of medication recommendations.

Finally, the output of the integrated medication recommender module $\hat{\mathbf{O}}$ by combining the latest patient representation $\mathbf{h}_n^t$ with the output memory representations $\mathbf{o}_b^t$ and $\mathbf{o}_d^t$, followed by a sigmoid function σ for scaling.

$$\hat{\mathbf{O}} = \sigma(\mathbf{h}_n^t \oplus \mathbf{o}_b^t \oplus \mathbf{o}_d^t) \tag{16}$$

The recommended medication combination $\hat{\mathbf{M}}^t$ was then derived by dichotomizing the model output $\hat{\mathbf{O}}$ to 0 or 1 with a threshold δ.

4.5 Model Training

In this work, we used two loss functions to assess the accuracy of the proposed model, including the binary cross-entropy loss function and multi-label hinge loss function. Furthermore, we evaluated the effectiveness of the DDI recommendation results using the DDI loss function. Finally, we employed an overall loss function that combined these three losses together.

Binary cross-entropy (BCE) *loss* was used to assess the classification effectiveness of our model. This loss measured the difference between the predicted output and the actual value. The BCE formula is:

$$L_{bce} = -\sum_{i=1}^{|M|} \mathbf{M}_i^{(t)} log(\hat{\mathbf{O}}_i^{(t)}) + (1 - \mathbf{M}_i^{(t)}) log(1 - \hat{\mathbf{O}}_i^{(t)}) \tag{17}$$

Multi-label hinge loss was used to assess the accuracy of our multilabel classification model. This loss function quantifies the difference between the predicted and actual labels. The formula for this loss is:

$$L_{multi} = \sum_{i,j:\mathbf{M}_i^{(t)}=1, \mathbf{M}_j^{(t)}=0} \frac{max(0, 1 - (\hat{\mathbf{O}}_i^{(t)} - \hat{\mathbf{O}}_j^{(t)}))}{|\mathbf{M}|} \tag{18}$$

DDI loss was determined as the DDI of the combination of recommendations generated by our model. This was a measure of the potential negative interactions between different drugs when taken simultaneously. Based on the DDI of our recommendations, we assessed the potential impact on model performance and identified any potential risks to the patient. The formula below allowed us to quantify the DDI of the medication recommendation. This allowed us to better understand the potential risks and benefits of the recommendations.

$$L_{ddi}^{(t)} = \sum_{i=1} \sum_{j=1} \mathbf{I}_{ij} \cdot \hat{\mathbf{O}}_i^{(t)} \cdot \hat{\mathbf{O}}_j^{(t)} \tag{19}$$

Overall Loss Function. A common method for training with multiple loss functions involves calculating the weighted sum of the terms measuring the loss [9]. The overall loss function was a linear combination of the three loss functions defined previously:

$$L = (\alpha L_{bce} + (1 - \alpha) L_{multi}) + \beta L_{ddi} \tag{20}$$

where α and β were hyper-parameters.

Two prediction losses L_{bce} and L_{multi} are compatible, leading us to select a threshold from the validation set. Hence, α controlled the tradeoff between L_{bce} and L_{multi} experimentally set as 0.95.

We noted that adverse DDIs can occur in the dataset, where doctors might mistakenly prescribe interacting drugs. Therefore, training with ground truth labels might inadvertently increase DDI occurrences. Inspired by [2], we adjusted the threshold during training using a Proportional-Integral-Derivative controller. Therefore, training with ground truth labels might inadvertently increase DDI occurrences. We aim to balance prediction loss and DDI loss by incorporating only the proportional error signal as negative feedback when the DDI rate of the recommended drugs (denoted as DDI in Equation (21)) surpasses certain thresholds. If the patient-level DDI falls below the threshold, denoted as γ, our focus is solely on maximizing prediction accuracy. Conversely, if the DDI exceeds the threshold, it will dynamically adjust to decrease DDI occurrences.

Therefore, β controlled the impact of DDI loss, determined by the formula below:

$$\beta = \begin{cases} 0, & DDI \leq \gamma \\ 1 - min\left(0, 1 + \frac{\gamma - DDI}{K_p}\right), & otherwise \end{cases} \tag{21}$$

where γ is a predefined DDI acceptance rate of 0.06 and K_p is the coefficient of the proportional signal of 0.05.

The inference phase of the model essentially mirrors the training pipeline. We utilize a threshold, denoted as $\delta = 0.5$, on the output drug representation in Equation (16). We then identify the drugs corresponding to the entries whose value exceeds δ as the final recommendations.

Algorithm 1: One Training Epoch of **FastRx**

```
Input: a training set $X^{train}$, hyperparameters $\alpha$ and $\beta$, pre-defined DDI acceptance rate $\gamma$,
       the DDI matrix $\mathbf{D}$;
Initialize parameters: $E_d$, $E_p$, 1D-CNN, Fastformer, $\{W^{(i)}\}$;
Construct GCN models $G_e$ and $G_d$ for EHR and DDI, respectively;
Obtain the interaction-aware medication representations $M_g$ by fusing $G_e$ & $G_d$
 $\Longrightarrow M_g = G_e - \lambda G_d$;
for $patient\ i \leftarrow 1$ to $|X^{train}|$ do
    Get patient $i$'s history, $X_i$;
    ### Patient Representation Encoder ###
    for $visit\ t \leftarrow 1$ to $|X_i|$ do
        Select the $t^{th}$ visit of patient $i$, $x_i^{(t)}$;
        Generate embeddings $e_d^{(t)}$ & $e_p^{(t)}$;
    end
    $c_d^{(i)} \leftarrow ReLU(CNN_{1D}([e_d^{(1)}, e_d^{(2)}, ..., e_d^{|X_i|}]))$;
    $c_p^{(i)} \leftarrow ReLU(CNN_{1D}([e_p^{(1)}, e_p^{(2)}, ..., e_p^{|X_i|}]))$;
    $h^{(i)} \leftarrow Concat([c_d^{(i)}, c_p^{(i)}])$;
    Pass $h^{(i)}$ into the $Fastformer$ layer;
    ### Medication Graph Encoder ###
    Generate global memory bank vector $M_g^{(i)}$;
    Generate global dynamic memory vector $M_d^{(i)}$;
    Generate output memory representations $o_b^i$ and $o_d^i$;
    ### Integrated Drug Recommender ###
    Generate drug recommendation $\hat{O}_l^{(i)}$ and obtain multi-hot drug vector $\hat{M}_l^{(i)}$;
    Accumulate $L_{bce}$, $L_{multi}$, $L_{ddi}$ ;                        /* Losses */
end
$L \leftarrow (\alpha L_{bce} + (1-\alpha)L_{multi}) + \beta L_{ddi}$;
Optimize parameters based on $L$;
```

The comprehensive algorithmic framework is elaborated in Algorithm 1.

5 Experimental Study

In this section, we evaluate our proposed FastRx on MIMIC-III dataset. The experiments are conducted to address the following questions.

- *RQ1:* Could the proposed model outperform state-of-the-art approaches on medication recommendation?
- *RQ2:* What is the contribution of each proposed component to the overall recommendation performance?
- *RQ3:* Could the proposed model outperform state-of-the-art approaches on computational complexity analysis (e.g., training speed, inference times)?
- *RQ4:* How do the δ and γ parameters impact the method's performance?
- *RQ5:* How to explain the recommendation results?

Table 2. Summary of Processed MIMIC-III Dataset Statistics

Items	**Number**
Number of visits/Number of patients	15,032/6,350
Diagnoses/Procedure/Medication space size	1,958/1,430/112
Avg./Max. Number of visits	2.37/29
Avg./Max. Number of diagnoses per visit	13.63/39
Avg./Max. Number of procedures per visit	4.54/32
Avg./Max. Number of medications per visit	19.57/52
Total Number of DDI pairs	337

5.1 Dataset

We used MIMIC III [14] dataset, available at PhysioNet,[1] which contains 58,976 hospital admissions of 46,520 patients, from 2001 to 2012. We acquire DDI relationships from TWOSIDES [32] and transform drug coding from National Drug Code to ATC third level to facilitate integration with the dataset. Following [40], we extract text information for diagnoses and procedures by mapping their ICD codes to text descriptions in dictionary tables. Table 2 shows the summary statistics of the processed dataset.

5.2 Implementation Details

In this section, we describe the experimental settings, data processing, model configurations, and parameters and elaborate on the sampling methodology used during the testing phase. All computations utilized a Linux workstation with 32 GB RAM, 20 CPU cores, and a 12 GB NVIDIA GeForce RTX 3060 GPU. For a fair comparison, we follow the data processing procedure in [40]. Data were randomly divided into training, validation, and test sets ($\frac{2}{3} : \frac{1}{6} : \frac{1}{6}$). In terms of the configurations and parameters of the model, the embedding size for tables $\mathbf{E}_d$ and $\mathbf{E}_p$ is set to 64. FastRx hyper-parameters chosen from the validation set: $\delta = 0.5$, weights $\alpha = 0.95$, β determined by Equation (21) from DDI loss, $K_p = 0.05$, and acceptance rate = 0.06. All models were compared, implemented in PyTorch 1.4.0, and trained with Adam optimizer [16] with 1e-4 learning rate, 1e-6 weight decay for 100 epochs, and dropout parameters set to 0.2. We used a single Fastformer [35] layer with 4 attention heads and incorporated 2 hidden layers, each with 256 hidden states. The output size for each layer was set to 64 and a Cyclical Learning Rate scheduler with a triangular training policy was used. Due to the limited availability of publicly accessible EHRs data, we utilize bootstrapping sampling during this phase, following the recommendation in [40]. The models were trained in the training set, with hyper-parameter selection on the validation set. Bootstrapping sampled 80% of test set data points in each of the 10 runs for evaluation. Performance calculations were based on these results.

5.3 Evaluation Metrics

To evaluate the effectiveness of our proposed model against the baselines, we used five metrics for medication recommendations [29, 36, 39, 40], namely, **Jaccard Similarity Score (Jaccard), Average F1 (F1)**, **Precision-Recall AUC (PRAUC)**, **DDI Rate (DDI)**, and **Average number of drugs (Avg.# of Drugs)**, as defined below.

— *Jaccard* was a statistical method that quantifies the similarity between two finite sample sets. It was used to compare the overlap between the model-recommended medications and the

[1] https://mimic.physionet.org/

ground-truth medications, as defined below:

$$\text{Jaccard} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\mathcal{M}_i \cap \hat{\mathcal{M}}_i\right|}{\left|\mathcal{M}_i \cup \hat{\mathcal{M}}_i\right|} \tag{22}$$

where T is the number of patients, $\mathcal{M}_i$ is the combination of ground-truth medications and $\hat{\mathcal{M}}_i$ is the predicted result. A higher Jaccard similarity coefficient indicates a greater overlap.

- *F1* the harmonic mean of precision and recall for evaluating binary classification models, defined below.

$$\text{Precision}_i = \frac{\left|\mathcal{M}_i \cap \hat{\mathcal{M}}_i\right|}{\left|\hat{\mathcal{M}}_i\right|} \tag{23}$$

$$\text{Recall}_i = \frac{\left|\mathcal{M}_i \cap \hat{\mathcal{M}}_i\right|}{\left|\mathcal{M}_i\right|} \tag{24}$$

$$F_1 = \frac{1}{N}\sum_{i=1}^{N} 2 * \frac{\text{Precision}_i * \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{25}$$

- *PRAUC* calculated the area under the precision-recall curve, defined below.

$$\text{PRAUC} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{|\mathcal{M}|}\text{Precision}(k)_i \bigtriangleup \text{Recall}(k)_i \tag{26}$$

$$\bigtriangleup\text{Recall}(k)_i = \text{Recall}(k)_i - \text{Recall}(k-1)_i \tag{27}$$

where k is the rank in the sequence of drugs, $|M|$ is the length of drug set. $\text{Precision}(k)_i$ is the precision at cut-off k in ordered retrieval list, and $\bigtriangleup\text{Recall}(k)_i$ is the change in recall when deriving the kth drug.

- *DDI* refers to the interactions that can occur between multiple medications when taken concurrently. In the realm of medication recommendation, the DDI was quantified using the following formula:

$$\text{DDI} = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{\left|\hat{\mathcal{M}}_i\right|}\sum_{k=j+1}^{\left|\hat{\mathcal{M}}_k\right|} 1\left\{A_d[\mathcal{M}_i^j, \mathcal{M}_i^k] = 1\right\}}{\sum_{j=1}^{\left|\hat{\mathcal{M}}_i\right|}\sum_{k=j+1}^{\left|\hat{\mathcal{M}}_k\right|} 1} \tag{28}$$

where A_d is the adjacency matrix of the DDI graph, $\hat{\mathcal{M}}_i^{(j)}$ denotes the jth recommended medication and $1\{\cdot\}$ is a function that returns 1 when the expression in $\{\cdot\}$ is true, otherwise 0.

- *Avg.# of Drugs* refers to the average number of medications included in each recommendation. The significance of this metric is to assess the complexity of the combination of medications provided by the recommender system. A higher metric means that each recommended regimen contains more medications, which can increase the complexity of the medication and the risk of adverse effects for patients. Conversely, a lower indicator means that medication combinations may be easier to manage and reduce unnecessary medication use,

$$\text{Avg.\# of Drugs} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{\mathcal{M}}_i\right| \tag{29}$$

where N represents the total number of visits for patient i and $\left|\hat{\mathcal{M}}_i\right|$ denotes the number of predicted medications in visit t of patient i.

5.4 Baseline Models

To investigate the effectiveness of our proposed FastRx, we perform a comparison with recent and competitive state-of-the-art approaches listed below.

- *Logistic Regression (LR)* with a L2-norm was used as a basic classifier. It is a binary classification linear model that transforms into a probability space through a linear combination of feature weights.
- *Ensemble Classifier Chain (ECC)* [27] is an ensemble of multiple support vector machine classifiers to make multi-label predictions.
- *LEAP* [43] is a Long Short Term Memory-based model that uses an attention mechanism and incorporates external medical knowledge to improve its recommendations.
- *DMNP* [19] is a memory-augmented neural network that makes predictions using differentiable neural computers.
- *RETAIN* [8] uses patient history data to simulate a diagnostic process and employs attention and gate mechanisms to improve prediction interpretation.
- *GAMENet* [29] utilizes a combination of a memory network and a graph neural network to recommend medication combinations.
- *4SDrug* [31] is designed to recommend small drug sets to ensure fewer DDIs.
- *MICRON* [39] uses a recurrent residual learning model to improve the learning of medication changes and recommend future medication from multiple data sources.
- *SafeDrug* [40] uses a drug molecular graph and a DDI graph to predict a combination of safe medications.
- *DrugRec* [30] uses a causal graphical model to address recommendation bias, extends the model for multivisit scenarios, and introduces a novel algorithm to coordinate multiple drugs considering DDI.
- *MoleRec* [41] enhances the precision of medication outcome predictions and interactions through a more accurate modeling of the correlation between distinct molecular substructures present in a medication molecule and the patient's health condition.
- *COGNet* [36] introduced a copy-or-predict mechanism for medication recommendation based on patient procedure and diagnosis with a DDI graph and an EHR graph.
- *StratMed* [20] develops an innovative dual-graph structure to represent the safety and precision of drugs on the same level.

5.5 Performance Comparison (RQ1)

The performance (mean ± SD) of the models was compared using the standard metrics in Table 3. The top and second best results are highlighted in bold and underlined, respectively. The notation ↓ indicates better performance for lower metric values, while ↑ suggests better performance for higher metric values. Our proposed FastRx outperformed the other models compared, achieving a superior 0.5443 Jaccard score, 0.6963 F1 score, and 0.7882 PRAUC. In contrast, instance-based models, including LR, ECC, and LEAP, were suboptimal because they relied solely on current patient conditions.

Methods that considered the patient's history, such as RETAIN, DMNC, GAMENet, SafeDrug, MICRON, and COGNet, generally performed better due to incorporating long-term medical patient information into their analysis. RETAIN and DMNC only relied on medical history and did not introduce external knowledge into the models. Although GAMENet used additional data, such as

Table 3. Performance Comparison of Different Methods on MIMIC-III Dataset

Method	DDI ↓	Jaccard ↑	F1 ↑	PRAUC ↑	Avg.# of Drugs
LR	0.0829 ± 0.0009	0.4865 ± 0.0021	0.6434 ± 0.0019	0.7509 ± 0.0018	16.1773 ± 0.0942
ECC	0.0846 ± 0.0018	0.4996 ± 0.0049	0.6569 ± 0.0044	0.6844 ± 0.0038	18.0722 ± 0.1914
LEAP	0.0731 ± 0.0008	0.4521 ± 0.0024	0.6138 ± 0.0026	0.6549 ± 0.0033	18.7138 ± 0.0666
DMNC	0.0842 ± 0.0011	0.4864 ± 0.0025	0.6529 ± 0.0030	0.7580 ± 0.0039	20.0000 ± 0.0000
RETAIN	0.0835 ± 0.0020	0.4887 ± 0.0028	0.6481 ± 0.0027	0.7556 ± 0.0033	20.4051 ± 0.2832
GAMENet	0.0864 ± 0.0006	0.5067 ± 0.0025	0.6626 ± 0.0025	0.7631 ± 0.0030	27.2145 ± 0.1141
4SDrug	0.0703 ± 0.0011	0.4800 ± 0.0027	0.6404 ± 0.0024	0.7611 ± 0.0026	16.1684 ± 0.1280
MICRON	0.0641 ± 0.0007	0.5100 ± 0.0033	0.6654 ± 0.0031	0.7631 ± 0.0026	17.9267 ± 0.2172
SafeDrug	**0.0589 ± 0.0005**	0.5213 ± 0.0030	0.6768 ± 0.0027	0.7647 ± 0.0025	19.9178 ± 0.1604
DrugRec	0.0597 ± 0.0006	0.5220 ± 0.0034	0.6771 ± 0.0031	0.7720 ± 0.0036	22.0006 ± 0.1604
MoleRec	0.0756 ± 0.0006	0.5301 ± 0.0025	0.6841 ± 0.0022	0.7748 ± 0.0022	22.2239 ± 0.1661
COGNet	0.0858 ± 0.0008	0.5316 ± 0.0020	0.6644 ± 0.0018	0.7707 ± 0.0021	27.6279 ± 0.0802
StratMed	0.0642 ± 0.0005	0.5321 ± 0.0035	0.6861 ± 0.0034	0.7779 ± 0.0043	20.5318 ± 0.1681
FastRx	0.0669 ± 0.0007	**0.5443 ± 0.0037**	**0.6963 ± 0.0032**	**0.7882 ± 0.0037**	23.0349 ± 0.1909

graph information and drug molecular structures, contributing to better performance, it produced a high DDI rate. SafeDrug modeled drug molecular structure and implementation of a Local Bipartite Encoder, resulting in the lowest DDI rate among best baselines. In contrast to the models mentioned earlier, MICRON was the first to model changes in a patient's condition over time; its recurrent residual method failed to consider the correlations between different medications and the relationship between medications and the underlying disease being treated. COGNet introduced an attention mechanism to analyze the patient's history and graphs to model medication relationships, but maintains a high DDI rate because of no DDI constraint compared to our proposed FastRx (0.0852 vs. 0.0669). FastRx outperformed existing approaches in Jaccard, F1, and PRAUC by modeling hierarchical dependent learning from the patient's information and investigating how their previous core health condition influenced their current condition.

Generally speaking, a lower number of drugs can reduce the risks associated with polypharmacy, such as adverse drug reactions and interactions, while a higher number of medications may sometimes be necessary for patients with multiple comorbidities or more complex health conditions. As Table 3 shows, our proposed FastRx recommends nearly three more drugs on average than StratMed, the strongest baseline. However, when looking at DDI, FastRx has on average only 0.0027 higher DDI than StratMed, which means the nearly three more drugs recommended may not all contribute to DDI. On the other hand, FastRx demonstrates slightly more than 1% improvement in performance for Jaccard, F1 and PRAUC metrics, compared to StratMed. We would like to argue that the slight increase in the number of drugs and DDIs is offset by the performance gain in this tradeoff.

Compared with all baselines, our FastRx demonstrates superior prediction accuracy when operating under DDI constraints. The results of our experiments demonstrate that our model effectively maintains a balance between prediction accuracy and safety by utilizing the patient representation encoder and medication graph encoder module. Furthermore, it successfully predicts drug combinations that are both safer and more effective compared to current methods.

Table 4. Ablation Study for FastRx on MIMIC-III Dataset

Method	DDI ↓	Jaccard ↑	F1 ↑	PRAUC ↑	Avg.# of Drugs
FastRx w/o $\mathcal{D}$	0.0673 ± 0.0008	0.5115 ± 0.0041	0.6676 ± 0.0038	0.7647 ± 0.0037	22.1290 ± 0.0156
FastRx w/o $\mathcal{P}$	0.0682 ± 0.0008	0.5204 ± 0.0037	0.6766 ± 0.0032	0.7761 ± 0.0035	23.1795 ± 0.2054
FastRx w/o GCN	0.0685 ± 0.0008	0.5405 ± 0.0041	0.6932 ± 0.0036	0.7881 ± 0.0037	23.1122 ± 0.1861
FastRx w/o 1D-CNN	0.0691 ± 0.0007	0.5411 ± 0.0034	0.6938 ± 0.0030	0.7877 ± 0.0032	22.8025 ± 0.2047
FastRx w/ RNN	0.0585 ± 0.0005	0.5250 ± 0.0044	0.6798 ± 0.0039	0.7761 ± 0.0038	21.6724 ± 0.2034
FastRx w/ Transformer	0.0685 ± 0.0008	0.5294 ± 0.0051	0.6831 ± 0.0045	0.7802 ± 0.0038	23.6528 ± 0.2228
FastRx	**0.0669 ± 0.0007**	**0.5443 ± 0.0037**	**0.6963 ± 0.0032**	**0.7882 ± 0.0037**	23.0349 ± 0.1909

The best results are highlighted in bold.

5.6 Ablation Study (RQ2)

We further experimented with a number of configurations for an ablation study, to verify the effectiveness of each component within the proposed FastRx. The configurations are listed below.

- FastRx *w/o* $\mathcal{D}$: without diagnosis $\mathcal{D}$ from the EHRs.
- FastRx *w/o* $\mathcal{P}$: without procedure $\mathcal{P}$ from the EHRs.
- FastRx *w/o* $\mathcal{GCN}$: without the GCN module, along with the EHR and DDI graphs.
- FastRx *w/o* 1D-CNN: without the 1D CNN module.
- FastRx *w/* RNN: with the RNN module.
- FastRx *w/* Transformer: a version that replaced Fastformer with a Transformer layer of 4 attention heads, 2 hidden layers, 256 hidden states, and an output size of 64 for each layer. These settings are the same as those of the Fastformer version for fair comparison.

Table 4 below shows the results of the ablation experiments. It is clear that the absence of diagnoses $\mathcal{D}$ or procedures $\mathcal{P}$ resulted in much poorer model performance in all metrics, demonstrating the importance of including both $\mathcal{D}$ and $\mathcal{P}$. Additionally, removing the module $\mathcal{GCN}$ with the EHR and DDI graphs and the 1D-CNN module caused a considerable decrease in model efficacy in all metrics. Furthermore, we conducted an experiment by substituting 1D-CNN with RNN. The findings indicated that 1D-CNN proved to be more efficient and appropriate in prioritizing diagnosis codes and procedures in recent examinations. Meanwhile, the comparison between the Transformer version (FastRx w/ Transformer) and the proposed Fastformer version (FastRx) demonstrated that the proposed Fastformer version of FastRx outperformed the Transformer version in medication recommendation. Therefore, the above results illustrate the necessity of all key components included in the proposed FastRx.

5.7 Computational Complexity Analysis (RQ3)

Table 5 illustrates the model complexity evaluation. We compare the model complexity of FastRx alongside other baselines based on the number of parameters (# of Params), training time per epoch in seconds (Training Time Epoch(s)), and inference time in seconds (Inference Time(s)). These metrics are crucial indicators of a model's computational demands during both training and deployment phases. Notably, our comparison focuses on improvements over GAMENet, COGNet, MoleRec and others except that RETAIN is not specialized for drug recommendation. FastRx exhibits significantly lower space and time complexity compared to other baselines. Specifically, FastRx proves to be much more efficient during inference, offering significant advantages over its counterparts. LEAP and DMNC, characterized by sequential modeling, tend to recommend drugs individually, resulting in time-consuming processes. Meanwhile, GAMENet is significantly dependent on a large MB, necessitating greater space requirements. In contrast, FastRx emerges as

Table 5. Comparison of Model Complexity Metrics for FastRx and DL Baselines

Model	**# of Params**	**Training Time/Epoch(s)**	**Inference Time(s)**
LEAP	177.395	284.02	4.82
DMNC	521.886	3,439.49	191.81
RETAIN	285.489	23.95	4.82
GAMENet	444.209	92.51	7.34
MICRON	275.395	54.49	9.56
SafeDrug	336.122	138.58	6.62
MoleRec	507.060	419.52	19.98
COGNet	1.357.560	159.82	120.12
StratMed	572.624	313.4	110.34
FastRx (Transformer)	1.352.373	105.8	7.49
FastRx (Fastformer)	598.389	32.77	4.11

Table 6. Performance under Various Threshold Values (δ)

δ	**DDI ↓**	**Jaccard ↑**	**F1-Score ↑**	**PRAUC ↑**	**Avg.# of Drugs**
0.3	0.0663 ± 0.0007	0.5434 ± 0.0035	0.6956 ± 0.0030	0.7878 ± 0.0033	23.0638 ± 0.1714
0.4	0.0671 ± 0.0008	0.5439 ± 0.0042	0.6962 ± 0.0036	0.7878 ± 0.0035	22.8984 ± 0.1995
0.5	0.0669 ± 0.0007	0.5443 ± 0.0037	0.6963 ± 0.0032	0.7882 ± 0.0037	23.0349 ± 0.1909
0.6	0.0674 ± 0.0007	0.5437 ± 0.0034	0.6959 ± 0.0030	0.7897 ± 0.0035	23.1785 ± 0.1950
0.7	0.0689 ± 0.0008	0.5438 ± 0.0037	0.6962 ± 0.0033	0.7885 ± 0.0034	23.1310 ± 0.1997

a more efficient and flexible solution suitable for real-world deployment scenarios. Overall, our findings suggest that FastRx represents a promising option for drug recommendation tasks, offering efficiency and adaptability for practical applications.

5.8 Experiments on δ and γ Parameters (RQ4)

After thoroughly analyzing the performance metrics across various threshold values (δ), we have determined that a threshold of 0.5 is optimal for our medication recommendation system, as illustrated in Table 6. Our method exhibits a balanced performance across all key metrics at this threshold, effectively minimizing adverse drug interactions while maintaining high accuracy. Furthermore, selecting a threshold of 0.5 aligns with common standards, facilitating fair and standardized comparisons with other baseline approaches.

Table 7 illustrates DDI controllability using the threshold γ, with a ground truth DDI rate of 0.0808 in the MIMIC-III dataset. By testing γ values from 0 to 0.08 across 10 experiments each, we found that DDI rates were effectively controlled and capped by γ. Higher γ values allowed more drugs per combination, enhancing recommendation accuracy, while very low γ values (< 0.02) slightly reduced accuracy. For a fair comparison with other baselines, we selected a value of 0.06. The parameter γ provided a way for doctors to control the tradeoff between DDI rates and accuracy in recommendations. This analysis provided evidence that our model could effectively balance DDI rates and accuracy, offering a flexible tool for doctors to manage the safety and efficacy of drug recommendations.

5.9 Case Study (RQ5)

We present an example of patient visits from the MIMIC-III dataset to demonstrate FastRx's recommended medications. The patient had three hospital visits. On the first visit, the patient suffered

Table 7. Performance under Acceptance DDI Rate (γ)

γ	DDI ↓	Jaccard ↑	F1-Score ↑	PRAUC ↑	Avg.# of Drugs
0.00	0.0627 ± 0.0007	0.5422 ± 0.0032	0.6945 ± 0.0028	0.7864 ± 0.0036	22.8162 ± 0.1921
0.01	0.0636 ± 0.0007	0.5410 ± 0.0034	0.6936 ± 0.0030	0.7878 ± 0.0035	22.5945 ± 0.1886
0.02	0.0641 ± 0.0006	0.5419 ± 0.0038	0.6944 ± 0.0033	0.7870 ± 0.0034	22.5834 ± 0.1963
0.03	0.0623 ± 0.0008	0.5436 ± 0.0043	0.6957 ± 0.0038	0.7886 ± 0.0038	22.7116 ± 0.1876
0.04	0.0646 ± 0.0007	0.5429 ± 0.0034	0.6953 ± 0.0030	0.7881 ± 0.0035	22.4777 ± 0.1935
0.05	0.0666 ± 0.0006	0.5436 ± 0.0038	0.6958 ± 0.0033	0.7882 ± 0.0038	22.8220 ± 0.1990
0.06	0.0669 ± 0.0007	0.5443 ± 0.0037	0.6963 ± 0.0032	0.7882 ± 0.0037	23.0349 ± 0.1909
0.07	0.0683 ± 0.0006	0.5426 ± 0.0036	0.6951 ± 0.0032	0.7881 ± 0.0035	23.2687 ± 0.1953
0.08	0.0710 ± 0.0007	0.5425 ± 0.0037	0.6949 ± 0.0033	0.7896 ± 0.0034	23.3301 ± 0.1959

Table 8. The Example for the Case Study

	Diagnoses	Medications
Visit 1	9962, 1919, 3485, 34290, 4590	**N02B**, **A07A**, **A06A**, **B01A**, **A02B**, **A12C**, **A01A**, **N03A**, N05B, N02A, **A12B**, **C07A**, **A12A**, **N01A**
Visit 2	1913, 34510, 3485, 30000	**B05C**, **A12C**, **N02B**, A12A, **A02B**, **N03A**, **A06A**, **C02D**, **A12B**, **A01A**, **A04A**, **N02A**, J01M, D06A, **A07A**, **C07A**, **B01A**, **J01D**
Visit 3	1913, 34290, 34590	**A07A**, **A12C**, **A01A**, **N03A**, **A06A**, **C02D**, **A12B**, **A04A**, **A02B**, J01M, **N02A**, D06A, **N02B**, C07A, **B05C**, **B01A**, **J01D**

mechanical complications of the nervous system device and graft (9962), malignant neoplasm of the brain (1919), cerebral edema (3485), hemiplegia (34290), and epilepsy (34590). On the second visit, the patient was diagnosed with a malignant neoplasm of the parietal lobe (1913), generalized convulsive epilepsy (34510), cerebral edema (3485), and anxiety state (30000). On the third visit, the patient was again diagnosed with malignant neoplasm of the parietal lobe (1913), epilepsy (34590), and hemiplegia (34290).

Table 8 lists the diagnoses of these three visits in the ICD-9 codes and the corresponding medications in the ATC codes. The bold ATC codes in the table indicate the medications predicted by our model, with the green color indicating the true positives and the red color indicating the false positives, for example, B01A and J01D in Visit 2. The remaining black ones were false negatives, for example, J01M, D06A, and C07A in Visit 3. For diseases that recurred during different visits, the same drugs were recommended. For example, the model outputs N03A (antiepileptics) and N02B (analgesics) in each recommendation. We observed that A12C, a recommended mineral supplement, suggested that the model recommended not only medication to treat primary diseases but also adjuvant therapy, which doctors recommended three to four on average per patient visit. Therefore, our model could recommend both primary medications and adjuvant therapies.

5.10 Discussions

Graph Update. The current design of our proposed FastRx relies on static DDI graphs, which were constructed based on the most comprehensive and up-to-date interactions available at the time of development. It allows the model to provide reliable recommendations based on established interactions. On the other hand, FastRx also incorporates an EHR graph based on personalized prescription data derived from EHRs to accommodate patient-specific variations. The model employs an attention mechanism that dynamically weights the relevance of different nodes and meta-paths within the EHR graph. In light of the dynamic nature of medical knowledge and individual patient responses, these graphs can be potentially updated periodically or in real-time by automatically checking new DDI information and incorporating new EHRs. Future iterations of our model may include exploring real-time learning and reinforcement learning algorithms and establishing real-time DDI databases with standardized reporting protocols, to further enhance adaptability by allowing the graph to adjust based on ongoing clinical practices and patient feedback.

EHR Quality. In real-world settings, EHRs can contain incomplete, erroneous, or inconsistent entries. Such data quality issues can potentially affect the performance of the proposed model. Our model leverages graph neural networks and attention mechanisms to manage incomplete and noisy data inherently. To further address data incompleteness, we will employ in our future work advanced imputation techniques, such as multiple imputation by chained equations, pharmacological context-aware imputation, and temporal imputation, to fill in missing data effectively. Additionally, data augmentation methods, including Generative Adversarial Networks, can be used to generate realistic synthetic data, maintaining clinical relevance. To address potential data noise and error, anomaly detection algorithms or rule-based systems can be used in our future work to identify and correct errors in EHR entries. Besides, real-time quality checks and feedback loops with clinicians can be used for information from external sources such as prescription databases and clinical trial records for continuous data quality monitoring. These measures should enhance the model's robustness to noisy data, ensuring reliable performance in varied and less controlled clinical environments.

6 Conclusion

In this research, we presented FastRx, a Fastformer-based model for personalized medication recommendation. The proposed model captured changes in a patient's current and historical visits based on 1D-CNN combined with Fastformer and integrated DDIs and combination medication in prescriptions using GCNs. The proposed FastRx outperformed the state-of-the-art models compared on the MIMIC-III dataset. Further ablation experiments demonstrated the effectiveness of individual components of FastRx. In future research, we plan to focus on reducing the DDI rate of recommended medication combinations while improving the performance of the model. Future work includes incorporating lab reports and medical notes, which are rich sources of information for additional insights such as patient allergies.

References

[1] Daniel Almirall, Scott N. Compton, Meredith Gunlicks-Stoessel, Naihua Duan, and Susan A. Murphy. 2012. Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. *Statistics in Medicine* 31, 17 (2012), 1887–1902. DOI: https://doi.org/10.1002/sim.4512

[2] Karl Johan Åström and Tore Hägglund. 1995. *PID Controllers: Theory, Design, and Tuning*. ISA- The Instrumentation, Systems and Automation Society, Pennsylvania, United States.

[3] Tian Bai, Shanshan Zhang, Brian L. Egleston, and Slobodan Vucetic. 2018. Interpretable representation learning for healthcare via capturing disease progression through time. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. ACM, New York, NY, 43–51. DOI: https://doi.org/10.1145/3219819.3219904

[4] Jacek M. Bajor and Thomas A. Lasko. 2017. Predicting medications from diagnostic codes with recurrent neural networks. In *Proceedings of the 5th International Conference on Learning Representations (ICLR '17),* 19 pages. Retrieved from https://openreview.net/forum?id=rJEgeXFex

[5] Inci M. Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K. Jain, and Jiayu Zhou. 2017. Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*. ACM, New York, NY, 65–74. DOI: https://doi.org/10.1145/3097983.3097997

[6] Jie Chen, Tengfei Ma, and Cao Xiao. 2018. Fastgcn: Fast learning with graph convolutional networks via importance sampling. arXiv:1801.10247. Retrieved from https://arxiv.org/abs/1801.10247

[7] Zhuo Chen, Kyle Marple, Elmer Salazar, Gopal Gupta, and Lakshman Tamil. 2016. A physician advisory system for chronic heart failure management based on knowledge patterns. *Theory and Practice of Logic Programming* 16, 5–6 (2016), 604–618. DOI: https://doi.org/10.1017/s1471068416000429

[8] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. 2016. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In *Advances in Neural Information Processing Systems*, Vol. 29. DOI: https://doi.org/10.5555/3157382.3157490

[9] Alexey Dosovitskiy and Josip Djolonga. 2020. You only train once: Loss-conditional training of deep networks. In *Proceedings of the International Conference on Learning Representations*. OpenReview.net, 17 pages. Retrieved from https://api.semanticscholar.org/CorpusID:214278158

[10] I. Ralph Edwards and Jeffrey K. Aronson. 2000. Adverse drug reactions: Definitions, diagnosis, and management. *The Lancet* 356, 9237 (2000), 1255–1259. DOI: https://doi.org/10.1016/s0140-6736(00)02799-9

[11] Fan Gong, Meng Wang, Haofen Wang, Sen Wang, and Mengyue Liu. 2021. SMR: Medical knowledge graph embedding for safe medicine recommendation. *Big Data Research* 23 (2021), 100174. DOI: https://doi.org/10.1016/j.bdr.2020.100174

[12] Meredith Gunlicks-Stoessel, Laura Mufson, Ana Westervelt, Daniel Almirall, and Susan Murphy. 2016. A pilot SMART for developing an adaptive treatment strategy for adolescent depression. *Journal of Clinical Child & Adolescent Psychology* 45, 4 (2016), 480–494. DOI: https://doi.org/10.1080/15374416.2015.1015133

[13] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, Vol. 30, 11 pages.

[14] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data* 3, 1 (May 2016). DOI: https://doi.org/10.1038/sdata.2016.35

[15] David N. Juurlink, Muhammad Mamdani, Alexander Kopp, Andreas Laupacis, and Donald A. Redelmeier. 2003. Drug-drug interactions among elderly patients hospitalized for drug toxicity. *JAMA* 289, 13 (2003), 1652–1658. DOI: https://doi.org/10.1001/jama.289.13.1652

[16] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. arXiv:1412.6980. Retrieved from http://arxiv.org/abs/1412.6980

[17] Thomas N. Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907, 14 pages. Retrieved from https://openreview.net/forum?id=SJU4ayYgl

[18] Himabindu Lakkaraju and Cynthia Rudin. 2017. Learning cost-effective and interpretable treatment regimes. In *Artificial Intelligence and Statistics*. PMLR, 166–175.

[19] Hung Le, Truyen Tran, and Svetha Venkatesh. 2018. Dual memory neural computer for asynchronous two-view sequential learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1637–1645. DOI: https://doi.org/10.1145/3219819.3219981

[20] Xiang Li, Shunpan Liang, Yulei Hou, and Tengfei Ma. 2024. StratMed: Relevance stratification between biomedical entities for sparsity on medication recommendation. *Knowledge-Based Systems* 284 (2024), 111239.

[21] Yifu Li, Ran Jin, and Yuan Luo. 2019. Classifying relations in clinical narratives using segment graph convolutional and recurrent neural networks (Seg-GCRNs). *Journal of the American Medical Informatics Association* 26, 3 (2019), 262–268. DOI: https://doi.org/10.1093/jamia/ocy157

[22] Leape L. L. 1995. Systems analysis of adverse drug events. *JAMA* 274 (1995), 35–43. DOI: https://doi.org/10.1001/jama.1995.03530010049034

[23] Chengsheng Mao, Liang Yao, and Yuan Luo. 2022. Imagegcn: Multi-relational image graph convolutional networks for disease identification with chest x-rays. *IEEE Transactions on Medical Imaging* 41, 8 (2022), 1990–2003.

[24] Chengsheng Mao, Liang Yao, and Yuan Luo. 2022. MedGCN: Medication recommendation and lab test imputation via graph convolutional networks. *Journal of Biomedical Informatics* 127 (2022), 104000. DOI: https://doi.org/10.1016/j.jbi.2022.104000

[25] Ziad Obermeyer and Ezekiel J. Emanuel. 2016. Predicting the future—Big data, machine learning, and clinical medicine. *The New England Journal of Medicine* 375, 13 (2016), 1216. DOI: https://doi.org/10.1056/nejmp1606181

[26] Maria Panagioti, Jonathan Stokes, Aneez Esmail, Peter Coventry, Sudeh Cheraghi-Sohi, Rahul Alam, and Peter Bower. 2015. Multimorbidity and patient safety incidents in primary care: A systematic review and meta-analysis. *PloS One* 10, 8 (2015), e0135947. DOI: https://doi.org/10.1371/journal.pone.0135947

[27] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. 2011. Classifier chains for multi-label classification. *Machine Learning* 85 (2011), 333–359. DOI: https://doi.org/10.1007/978-3-642-04174-7_17

[28] Junyuan Shang, Tengfei Ma, Cao Xiao, and Jimeng Sun. 2019. Pre-training of graph augmented transformers for medication recommendation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI '19)*. ijcai.org, China, 5953–5959. DOI: https://doi.org/10.24963/ijcai.2019/825

[29] Junyuan Shang, Cao Xiao, Tengfei Ma, Hongyan Li, and Jimeng Sun. 2019. Gamenet: Graph augmented memory networks for recommending medication combination. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, 1126–1133. DOI: https://doi.org/10.1609/aaai.v33i01.33011126

[30] Hongda Sun, Shufang Xie, Shuqi Li, Yuhan Chen, Ji-Rong Wen, and Rui Yan. 2022. Debiased, longitudinal and coordinated drug recommendation through multi-visit clinic records. In *Advances in Neural Information Processing Systems*, Vol. 35, 27837–27849.

[31] Yanchao Tan, Chengjun Kong, Leisheng Yu, Pan Li, Chaochao Chen, Xiaolin Zheng, Vicki S. Hertzberg, and Carl Yang. 2022. 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 (KDD '22)*. ACM, New York, NY, 3970–3980. DOI: https://doi.org/10.1145/3534678.3539089

[32] Nicholas P. Tatonetti, Patrick Ye, Roxana Daneshjou, and Russ B. Altman. 2012. Data-driven prediction of drug effects and interactions. *Science Translational Medicine* 4 (2012), 125ra31. Retrieved from https://api.semanticscholar.org/CorpusID:2716426

[33] Lu Wang, Wei Zhang, Xiaofeng He, and Hongyuan Zha. 2018. Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2447–2456. DOI: https://doi.org/10.1145/3219819.3219961

[34] Shanshan Wang, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Jun Ma, and Maarten de Rijke. 2019. Order-free medicine combination prediction with graph convolutional reinforcement learning. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 1623–1632. DOI: https://doi.org/10.1145/3357384.3357965

[35] Chuhan Wu, Fangzhao Wu, Tao Qi, Yongfeng Huang, and Xing Xie. 2021. Fastformer: Additive attention can be all you need. arXiv:2108.09084, 11 pages. Retrieved from https://arxiv.org/abs/2108.09084

[36] Rui Wu, Zhaopeng Qiu, Jiacheng Jiang, Guilin Qi, and Xian Wu. 2022. Conditional generation net for medication recommendation. In *Proceedings of the ACM Web Conference*. ACM, 935–945. DOI: https://doi.org/10.1145/3485447.3511936

[37] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S. Yu Philip. 2020. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems* 32, 1 (2020), 4–24. Retrieved from https://api.semanticscholar.org/CorpusID:57375753

[38] Cao Xiao, Edward Choi, and Jimeng Sun. 2018. Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review. *Journal of the American Medical Informatics Association* 25, 10 (2018), 1419–1428. DOI: https://doi.org/10.1093/jamia/ocy068

[39] Chaoqi Yang, Cao Xiao, Lucas Glass, and Jimeng Sun. 2021. Change matters: Medication change prediction with recurrent residual networks. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, 3728–3734. DOI: https://doi.org/10.24963/ijcai.2021/513

[40] Chaoqi Yang, Cao Xiao, Fenglong Ma, Lucas Glass, and Jimeng Sun. 2021. Safedrug: Dual molecular graph encoders for recommending effective and safe drug combinations. arXiv:2105.02711. Retrieved from https://arxiv.org/abs/2105.02711

[41] Nianzu Yang, Kaipeng Zeng, Qitian Wu, and Junchi Yan. 2023. MoleRec: Combinatorial drug recommendation with substructure-aware molecular representation learning. In *Proceedings of the ACM Web Conference (WWW '23)*. ACM, New York, NY, 4075–4085. DOI: https://doi.org/10.1145/3543507.3583872

[42] Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, 7370–7377. DOI: https://doi.org/10.1609/aaai.v33i01.33017370

[43] Yutao Zhang, Robert Chen, Jie Tang, Walter F. Stewart, and Jimeng Sun. 2017. 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*, 1315–1324. DOI: https://doi.org/10.1145/3097983.3098109

[44] Chenyi Zhuang and Qiang Ma. 2018. Dual graph convolutional networks for graph-based semi-supervised classification. In *Proceedings of the 2018 World Wide Web Conference*, 499–508. DOI: https://doi.org/10.1145/3178876.3186116

Received 3 March 2024; revised 12 August 2024; accepted 21 August 2024