SSPNet: Leveraging Robust Medication Recommendation with History and Knowledge

Haodi Zhang[1,2] , **Jiawei Wen**[1] , **Jiahong Li**[1] , **Yuanfeng Song**[3,4] , **Liang-Jie Zhang**[1] , **Lin Ma**[5]
[1]College of Computer and Software Engineering, Shenzhen University, Shenzhen, China
[2]Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
[3]Hong Kong University of Science and Technology, Hong Kong, China
[4]AI Group, WeBank Co., Ltd, Shenzhen, China
[5]School of Environment Nanjing University
hdzhang@szu.edu.cn, wenjiawei2023@email.szu.edu.cn, 2200271013@email.szu.edu.cn,
yfsong@webank.com, zhanglj@szu.edu.cn, malin1979@nju.edu.cn

Abstract

Automated medication recommendation is a crucial task within the domain of artificial intelligence in healthcare, where recommender systems are supposed to deliver precise, personalized drug combinations tailored to the evolving health states of patients. Existing approaches often treat clinical records (e.g., diagnoses, procedures) as isolated or unified entities, neglecting the inherent set-structured nature of medical data and the need to model interdependencies among clinical elements. To address the gap, we propose SSPNet , a novel end-to-end framework designed to process complete clinical record sets and directly generate optimal medication sets. SSPNet employs a set-based encoder to effectively capture and represent a patient's health condition from the electronic health records (EHRs), while a permutation-consistent decoder predicts the entire medication combination as a set. In addition, we introduce a novel personalized representation mechanism to capture the drugs previously used by individual patients. Extensive experiments on MIMIC-III and MIMIC-IV data sets reveal that SSPNet surpasses existing state-of-the-art methods in the accuracy of medication recommendations.

1 Introduction

In the healthcare domain, medication recommendation plays a vital role in determining the optimal combination of drugs tailored to the specific health conditions of a patient. The complexity increases significantly in cases of multimorbidity [Skou *et al.*, 2022], where patients suffer from multiple co-existing conditions, making medical decision-making particularly challenging. The current approach to prescribing medications is predominantly manual, which poses a significant challenge in achieving a balance between therapeutic benefits and minimizing associated risks, such as side effects and potential drug interactions. The crafting of an optimal medication regimen is a complex endeavor, even for experienced healthcare professionals, due to the multitude of factors and interactions that must be considered. The extensive adoption of digital health records in modern healthcare settings has significantly enriched electronic health records (EHRs) [Cowie *et al.*, 2017; Evans, 2016], providing detailed and comprehensive patient data. This abundance of healthcare data presents a promising opportunity to develop advanced predictive models that can enhance clinical decision-making processes.

Due to the clinical importance of drug recommendation, numerous promising methods based on deep learning have been proposed. However, these methods often overlook the set-based nature of drug recommendation [Tan *et al.*, 2022]. The predominant existing methods obtain representations of patients' health conditions by aggregating their diagnoses and procedures information into a unified entity [Wang *et al.*, 2021; Mi *et al.*, 2024]. Yet, these methods can lead to a loss of detailed patient health information. Additionally, some methods [Zhang *et al.*, 2017; Wu *et al.*, 2022] treat drug recommendation as a sequential decision-making process rather than the prediction of a set of drugs. These methods implicitly or explicitly establishes an order in the drug recommendation process. However, the process of actual medication recommendation is not governed by a predetermined sequence.

In this paper, we propose Set-to-Set Prediction Net (SSPNet), an encoder-decoder-based framework that predicts appropriate medication combinations in a set-to-set manner. Our model recommends a set of drugs to patients based on the set of their health status. Concurrently, our model is capable of capturing previously used medications, thereby facilitating the provision of personalized drug recommendations. SSPNet[1] consists of a permutation-consistent encoder-decoder-based network and a personalized drug representation module. Two encoders model the dependencies of diagnoses and procedures to obtain a patient's health condition representation set. The personalized drug representation module scales the representation of medications based on the correlation between the patient's health status and historical medication records. The decoder takes into the set of patient health status and drug representations to predict the target set of drugs. Our contributions can be summarized as follows.

[1]https://github.com/ResearchGroupHdZhang/SSPNet.git

- We introduce SSPNet, a novel set-to-set prediction framework for automated medication recommendation that handles the complexity of predicting sets of medications. SSPNet encodes patient health condition set from EHRs into robust high-dimensional representations and directly outputs appropriate set of medications. We highlight the robustness of our framework with respect to different orders of the inputs.
- We develop a personalized drug representation module (PDRM) to capture the historical use of medications in patients. This module precisely models the correlation between current health condition and historical medication information. Moreover, the PDRM module also has excellent compatibility and can benefit other baseline approaches.
- We conducted extensive experiments on MIMIC-III, and the results show that SSPNet outperforms current state-of-the-art methods. The effectiveness of SSPNet is validated through both quantitative metrics and qualitative case studies, confirming its practical utility and reliability in real-world clinical settings.

2 Related Work

In this section, we present a comprehensive review of the relevant methodologies, categorized according to the types of information utilized.

2.1 Multi-label Classification-based Medication Recommendation

Existing predominant approaches uniformly regard drug recommendation as a multi-label classification task [Sun *et al.*, 2022; Zhang *et al.*, 2023; Chen *et al.*, 2023; Shang *et al.*, 2019a; Bhoi *et al.*, 2021]. For example, SafeDrug [Yang *et al.*, 2021] and MoleRec [Yang *et al.*, 2023] models a patient's health condition using a Dual-RNN and enhances the efficacy and safety of drug recommendation considering the molecular structures of the drugs. Recently, 4SDrug [Tan *et al.*, 2022] proposed medication combinations by employing a set-oriented approach to measure the similarity between symptom sets and individual drugs. These methods assume independence among medications, or equal contribution of all diseases to the medication combinations. Given the complex relationships among drugs, the assumptions are ineffective for medication recommendation. In this paper, inspired by 4SDrug [Tan *et al.*, 2022], we aim to model the dependencies among medications while recommending drugs in parallel.

2.2 Sequential Decision-making in Medication Recommendation

There exist models that regard the drug recommendation process as a series of sequential decision-making processes [Wang *et al.*, 2024]. An early work, known as LEAP [Zhang *et al.*, 2017], uses a recurrent decoder to capture the drug-disease relationship for the current health condition of the patient, allowing sequential drug recommendation. COGNet [Wu *et al.*, 2022] and VITA [Kim *et al.*, 2024] employ an encoder-decoder generative network to predict medicines individually. These methods also mine the relationship between current patient health and past visits via a copy module. However, these methods implicitly or explicitly impose an order on the medication sets [Li *et al.*, 2023], which may lead to sub-optimal recommendations. In this paper, we aim to produce medication combinations with a set prediction method, which models the relationship between diseases and medications while alleviating the ordering restrictions.

2.3 Set Prediction

For a considerable duration, approaches predicting sets from feature vectors have consistently neglected the inherent unordered quality of sets. Recently, several frameworks [Locatello *et al.*, 2020; Kosiorek *et al.*, 2020; Carion *et al.*, 2020] have incorporated permutation-consistent set generators to address this challenge, since they do not require a specific ordering. DSPN [Zhang *et al.*, 2019], a permutation-consistent set generator, employs backpropagation to decode a set. However, generating a set using gradient descent from an initially guessed set can be computationally expensive. The Transformer [Vaswani *et al.*, 2017], which is permutation-consistent, learns to update the elements of the initial set in a joint manner. Therefore, both TSPN [Kosiorek *et al.*, 2020] and DETR [Carion *et al.*, 2020] leverage Transformer architecture to predict the final set in parallel. Inspired by these methods, we design a set recommendation framework based on Transformer architecture, ensuring permutation-consistency for medication recommendations.

3 Problem Formulation

In this section, we formulate the core task of automated medication recommendation.

3.1 Electrical Health Records (EHR)

Denote D the set of diagnoses, P the set of procedures and M the set of medications. For a patient p, the healthcare information from p's EHR is a sequence V of all historical visits,

$$V = [v^{(1)}, v^{(2)}, ..., v^{(T)}], \tag{1}$$

where $v^{(i)}$ is the i-th visit and T is the total number of visits for the patient. Each visit $v^{(i)} \in V$ is a triplet that contains the diagnosis, procedure, and medication records,

$$v^{(i)} = (v_d^{(i)}, v_p^{(i)}, v_m^{(i)}), \tag{2}$$

where $v_d^{(i)} \subseteq D$, $v_p^{(i)} \subseteq P$ and $v_m^{(i)} \subseteq M$. We denote all EHRs by $\mathcal{V}$, and all patient by $\mathcal{P}$. Note that $|\mathcal{V}| = |\mathcal{P}|$. To enhance the accuracy and interpretability of the recommendation, we integrate domain knowledge from the EHR graph and the drug-drug interaction (DDI) graph into the recommendation model. The EHR graph is an undirected graph $G_c = \langle M, \varepsilon_c \rangle$, where the nodes are all drug M, and the edges are defined as:

$$\varepsilon_c = \{(m_1, m_2) \mid \exists V \in \mathcal{V}.\ \exists v^{(i)} \in V.\ m_1 \in v_m^{(i)} \wedge m_2 \in v_m^{(i)}\}, \tag{3}$$

If two drugs in D appear in the same medication record of some visit of some patient, they are connected by an EHR graph edge. The DDI graph $G_d = \langle M, \varepsilon_c \rangle$ is the induced subgraph of the DDI information from database *Twosides* [Tatonetti *et al.*, 2012] with respect to the given medication space M. Then, we denote $\boldsymbol{A}_c, \boldsymbol{A}_d \in \{0,1\}^{|M|\times|M|}$ the adjacency matrices of the graphs G_c and G_d, respectively.

3.2 Medication Recommendation Problem

For some visit $v^{(t)} \in V$ of a given patient $p \in \mathcal{P}$, suppose that the diagnosis record $v_d^{(t)} = \{d_1, d_2, ..., d_k\}$, and the procedure record $v_p^{(t)} = \{p_1, p_2, ..., p_l\}$. With EHR and DDI graphs G_c and G_d as domain knowledge, we aim to learning a recommendation predication function $f(\cdot) : 2^D \times 2^P \mapsto 2^M$, that generates a proper medication combinations for the visit,

$$\hat{\mathbf{m}}^{(t)} = f(v_d^{(t)}, v_p^{(t)} | G_c, G_d). \tag{4}$$

where $\hat{\mathbf{m}}^{(t)} \subseteq M$ is encoded into multi-hot vector in the rest of the paper.

4 Methodology

Our framework SSPNet consists of four main modules, as shown in Figure 1. The details of these modules are presented in the following subsections.

4.1 Medication Representation Module

The co-occurrence and DDI of certain medications in a single prescription is a common practice in human-expert medication recommendation. We fuse the medication's feature vectors with co-occurrence and DDI information in the EHR graph. We first convert all medications into vectors via an embedding matrix $\mathbf{E}_m \in \mathbb{R}^{|M| \times h}$, where h is embedding size. Given the drug features $\boldsymbol{M} \in \mathbb{R}^{|M| \times h}$ and the co-occurrence adjacency matrix $\boldsymbol{A}_c$, we can fuse the co-occurrence information to medication representations by a dual-layer convolutional network (GCN) [Kipf and Welling, 2016] as:

$$\begin{aligned} \boldsymbol{G}_c &= GCN(ReLU(GCN(\boldsymbol{M}, \boldsymbol{A}_c))\boldsymbol{W_c}, \boldsymbol{A_c}), \\ where GCN(\boldsymbol{M}, \boldsymbol{A}_c) &= \sigma(\hat{\boldsymbol{O}}^{-\frac{1}{2}} \hat{\boldsymbol{A}}_c \hat{\boldsymbol{O}}^{-\frac{1}{2}} \boldsymbol{M}), \end{aligned} \tag{5}$$

where $\boldsymbol{W_c}$ is the learnable parameter. The $\hat{\boldsymbol{A}}_c = \boldsymbol{A}_c + \boldsymbol{I}$, and $\boldsymbol{I}$ represents the identity matrix. The diagonal node degree matrix associated with $\hat{\boldsymbol{A}}_c$ is denoted by $\hat{\boldsymbol{O}}$. Similarly, we can capture the DDI relations of medications:

$$\boldsymbol{G}_d = GCN(ReLU(GCN(\boldsymbol{M}, \boldsymbol{A}_d))\boldsymbol{W_d}, \boldsymbol{A_d}), \tag{6}$$

The above two parts are aggregated to obtain the representation of medications.

$$\boldsymbol{Z} = \boldsymbol{G}_c - \lambda \boldsymbol{G}_d. \tag{7}$$

where the λ is a learnable parameter.

4.2 Patient Representation Module

Patient's healthcare information includes diagnose records and procedure records. We design two embedding tables, $\mathbf{E}_d \in \mathbb{R}^{|D| \times h}$, $\mathbf{E}_p \in \mathbb{R}^{|P| \times h}$, where each row corresponds to the embedding vectors of the diagnose and the procedure, respectively. Given the diagnose and procedure codes of i-th visit, $v_d^{(i)}$, $v_p^{(i)}$, each diagnose and each procedure are converted into h-dimensional vectors $\mathbf{e}_d^i$ and $\mathbf{e}_p^i$ using the embedding matrices $\mathbf{E}_b$ and $\mathbf{E}_p$, respectively. They constitute the patient's diagnose and procedure embedding matrices $\boldsymbol{D}^{(i)} \in \mathbb{R}^{|k_i| \times h}$ and $\boldsymbol{P}^{(i)} \in \mathbb{R}^{|l_i| \times h}$. Two encoders based on set attention block (SAB) [Lee *et al.*, 2019; Kosiorek *et al.*, 2020] are utilized to obtain the patient's representation. In the following, we present the details of SABs.

Set Attention Blocks

Set attention blocks aim to model relationships among diagnoses and procedures within the same visit while ensuring permutation-consistency, respectively. Given a query matrix $\mathbf{Q} \in \mathbb{R}^{n \times h}$ and key-value pairs $\mathbf{K} \in \mathbb{R}^{n \times h}$ and $\mathbf{V} \in \mathbb{R}^{n \times h}$, we can formulate the attention function as follows:

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Softmax(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{h}})\mathbf{V}, \tag{8}$$

To efficiently capture the interaction information spanning multiple views, SAB leverages multi-head attention by projecting the **Q**, **K**, **V** onto different representation subspaces. Next, a linear layer is used to transform the concatenation of all attention outputs:

$$\begin{aligned} Multihead(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= concat(\mathbf{O}_1, \mathbf{O}_2 ..., \mathbf{O}_g)\mathbf{W}^{(O)}, \\ where\, \mathbf{O}_i &= Attention(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V), \end{aligned} \tag{9}$$

Given matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n \times h}$, the Multihead Attention Block (MAB) encodes them based on encoder block of the Transformer [Vaswani *et al.*, 2017], without positional encoding:

$$MAB(\mathbf{X}, \mathbf{Y}) = LayerNorm(\mathbf{H} + rFF(\mathbf{H})), \tag{10}$$

$$\mathbf{H} = LayerNorm(\mathbf{X} + Multihead(\mathbf{X}, \mathbf{Y}, \mathbf{Y}; \omega)), \tag{11}$$

Where rFF denotes a feed-forward layer and $LayerNorm$ refers to layer normalization [Ba *et al.*, 2016]. The ω is a parameter. Subsequently, the SAB is capable of being defined in conjunction with the MAB:

$$SAB(\mathbf{X}) := MAB(\mathbf{X}, \mathbf{X}). \tag{12}$$

To sum up, SAB takes a set as input and models the interactions between each element through self-attention, and ultimately outputs a set of the same dimensionality.

Patient's current health condition encoder

We denote the patient's current visit record as the t-th visit. After converting the diagnose and procedure codes of patient's current visit record into embedding matrices $\boldsymbol{D}^{(t)}$ and $\boldsymbol{P}^t$, we use two SABs, which defined in (12), to encode the diagnoses and procedures:

$$\hat{\boldsymbol{D}^{(t)}} = SAB_d(\boldsymbol{D}^{(t)}), \quad \hat{\boldsymbol{P}^{(t)}} = SAB_p(\boldsymbol{P}^{(t)}), \tag{13}$$

where $\hat{\boldsymbol{D}^{(t)}}$ and $\hat{\boldsymbol{P}^{(t)}}$ are the patient's current health condition representations encoded by SABs, respectively.

4.3 Personalized Drug Representations Module

The historical drug treatments can lead to variations in the final prescriptions, even when patients present with similar current health conditions [Shang *et al.*, 2019b; Zheng *et al.*, 2023]. Thus, we have developed a personalized drug representations module. This module leverages historical drug usage information to obtain personalized drug representations.

Pooling by Multihead Attention

The pooling by multihead attention (PMA) [Lee *et al.*, 2019] able to aggregate features while taking into account the importance of different features. Given a set of features $\mathbf{X} \in \mathbb{R}^{n \times h}$, the PMA can be defined as:

$$PMA(\mathbf{X}) = MAB(\mathbf{S}, rFF(\mathbf{X})). \tag{14}$$

where $\mathbf{S} \in \mathbb{R}^{1 \times h}$ is a learnable vector.

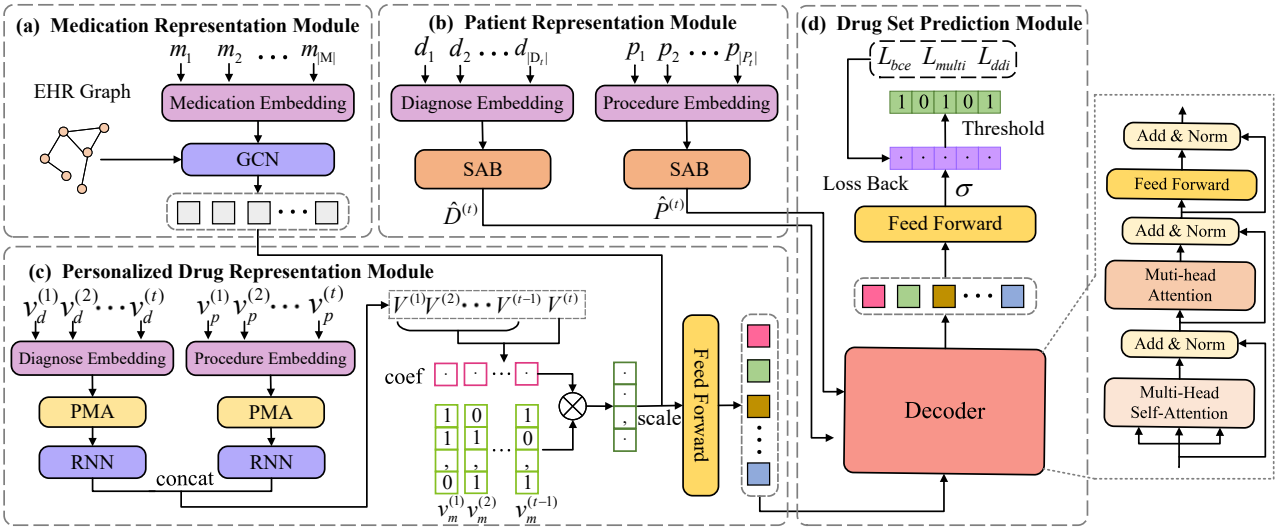


Figure 1: Overview of SSPNet Framework, (a): the medication representation module aggregates EHR information to obtain the representations of medications, (b): the patient representation module employs two SAB blocks to model the patient's current diagnoses and procedures to obtain patient heath representations, (c): the personalized drug representations module scales the medication representations based on the relevance between the patients' current and historical health conditions, and (d): the drug set prediction module utilizes a decoder to take into account the patient health representations and personalized drug representations, predicting the final prescription.

Scale Drug Representations

We use two PMAs to aggregate diagnosis and procedure representations of patient's all visits.

$$\boldsymbol{v}_d^{(i)} = PMA(\boldsymbol{D}^{(i)}), \quad \boldsymbol{v}_p^{(i)} = PMA(\boldsymbol{P}^{(i)}), \tag{15}$$

We model the sequential relationship between patient diagnoses and procedures with a Dual-RNN:

$$\begin{aligned} \boldsymbol{h}_{\boldsymbol{d}}^{(\boldsymbol{i})} &= RNN_d(\boldsymbol{v}_{\boldsymbol{d}}^{(\boldsymbol{i})}, \boldsymbol{h}_d^{(i-1)}), \\ \boldsymbol{h}_{\boldsymbol{p}}^{(\boldsymbol{i})} &= RNN_p(\boldsymbol{v}_{\boldsymbol{p}}^{(\boldsymbol{i})}, \boldsymbol{h}_p^{(i-1)}), \end{aligned} \tag{16}$$

Then we concatenate the $\boldsymbol{h}_d^{(i)}$ and the $\boldsymbol{h}_p^{(i)}$ to obtain each visit's representation $\boldsymbol{V}^{(i)}$.

$$\boldsymbol{V}^{(i)} = CONCAT[\boldsymbol{h}_d^{(i)}, \boldsymbol{h}_p^{(i)}], \tag{17}$$

We employ an attention mechanism to assess the correlation between patients' historical medical records and their previous health conditions.

$$a^{(i)} = Softmax(\frac{\boldsymbol{V}^{(i)}\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{V}^{(t)}}{\sqrt{h}}), \tag{18}$$

Denote the multi-hot vector $\boldsymbol{m}^{(i)} \in \{0,1\}^{|M|}$ as the i-th historical medications, where $\boldsymbol{m}_j^{(i)} = 1$ represent the patient has been treated with the j-th drug, and $\boldsymbol{m}_j^{(i)} = 0$ signifies that the j-th drug has not been used. Then we aggregate the historical medication records by $a^{(i)}$:

$$\boldsymbol{q} = \sum_{i=1}^{t-1} a^{(i)} \boldsymbol{m}^{(i)}, \tag{19}$$

where the $\boldsymbol{q}$ represents the correlation between the patient's current health condition and the medications previously used.

Then, we scale the drug representations based on the $\boldsymbol{q}$ and use a eed-forward network to obtain the personalized drug representations $\boldsymbol{Z}^{'}$:

$$\boldsymbol{Z}^{'} = FF_1(\boldsymbol{Z} + \boldsymbol{q}^T\boldsymbol{Z}). \tag{20}$$

4.4 Drug Set Prediction Module

The drug set prediction module is designed to recommend appropriate sets of medications based on the patient's health condition and personalized drug representations.

Drug Set Prediction

Inspired by previous works [Zaheer *et al.*, 2017; Carion *et al.*, 2020], we first designed a permutation-consistent decoder. In contrast to sequence generation models [Wu *et al.*, 2022; Kim *et al.*, 2024] that generate drug representations sequentially, the decoder takes as input all candidate drugs and predicts a targeted drug combination, without imposing any irrelevant sequential information. Given the personalized drug representations set $\boldsymbol{Z}^{'}$, a multi-head self-attention mechanism defined in (9) is utilized to model the interactions among the medication combinations,

$$\boldsymbol{Z}^{''} = layerNorm(\boldsymbol{Z}^{'} + Multihead(\boldsymbol{Z}^{'}, \boldsymbol{Z}^{'}, \boldsymbol{Z}^{'})), \tag{21}$$

Conditioned on the diagnose representations $\hat{\boldsymbol{D}}^{(t)}$ and procedure representations $\hat{\boldsymbol{P}}^{(t)}$ of the patient's current visit, we further refine the medication representations,

$$\begin{aligned} \hat{\boldsymbol{Z}} = layerNorm\big(&\boldsymbol{Z}^{''} + Multihead(\boldsymbol{Z}^{''}, \hat{\boldsymbol{D}}_t, \hat{\boldsymbol{D}}_t) \\ &+ Multihead(\boldsymbol{Z}^{''}, \hat{\boldsymbol{P}}_t, \hat{\boldsymbol{P}}_t)\big), \end{aligned} \tag{22}$$

Then, a fully connected feed-forward network (with a sigmoid function) transforms the representation $\hat{\boldsymbol{Z}}$ into predicted

medication representation $\hat{\boldsymbol{M}}$,

$$\hat{\boldsymbol{M}} = \sigma(FF_2(\hat{\boldsymbol{Z}})), \tag{23}$$

Subsequently, we use a feed-forward network $FF_2 : \mathbb{R}^h \rightarrow \mathbb{R}$ with a sigmoid activation function σ to derive the probability of each drug being recommended,

$$\hat{\boldsymbol{m}} = \sigma(FF_3(\hat{\boldsymbol{M}})). \tag{24}$$

Finally, we select entries from $\hat{\boldsymbol{m}}$ where values exceed predefined threshold δ, to obtain a multihot prediction vector $\hat{\boldsymbol{o}}$.

Loss Function Design

We first optimize all learnable parameters of SSPNet in training phase. As stated in the previous section, SSPNet gets the predicted medication representation and picks out the predicted medications with drug set prediction module. We consider it as a multi-label binary classification problem. Following [Yang *et al.*, 2021; Yang *et al.*, 2023], we introduce the binary cross-entropy loss $\mathcal{L}_{bec}$, the multi-label margin loss $\mathcal{L}_{multi}$ and DDI loss as our loss functions. The details are presented in the Appendix A.

5 Experiments

In this section, we conduct a series of extensive experiments to evaluate the performance of SSPNet, by comparing it with several competing models.

5.1 Dataset

We conducted experiments on the MIMIC-III [Johnson *et al.*, 2016] and MIMIC-IV [Johnson *et al.*, 2023]. Following the setting in [Yang *et al.*, 2021], we split the datasets into training, validation and test as $4 : 1 : 1$. The statistical information of the preprocessed dataset is shown in Table 1. Details of data pre-processing are presented in the Appendix B.

5.2 Setting

We implement our method using PyTorch 1.9.0, which is built on Python 3.8.16. We conducted all experiments on an Intel Xeon Platinum 8260 server, comprising 24 CPU cores, 188G memory and a 12GB NVIDIA TITAN V GPU. In our models, we set $h = 64$ for $\boldsymbol{E}_d$, $\boldsymbol{E}_p$ and $\boldsymbol{E}_m$. The number of initial medications $n = 131$. The feed-forward network FF_1, FF_1 and FF_3 are designed as one linear layer. We determined the best hyperparameters based on their validation set performance, setting the threshold $\delta = 0.5$, weight $\beta = 0.95$ and $K_p = 0.05$. The Adam optimizer [Kingma and Ba, 2014] is applied for model training, with a learning rate of 1×10^{-4}.

5.3 Baselines and Metrics

We assess the performance of SSPNet through a comparative analysis with the following models: the instance-based standard Logistic Regression (LR), the chain-structured LR classifier model: Ensemble Classifier Chain (ECC) [Read *et al.*, 2011], the dynamic graph-augmented memory-based approach: GAMENet [Shang *et al.*, 2019b], the drug molecule structure-based approaches: SafeDrug [Yang *et al.*, 2021] and MoleRec [Yang *et al.*, 2023]. the sequential decision making approach LEAP: [Zhang *et al.*, 2017], and the conditional generation network-based approaches: COGNet [Wu *et al.*, 2022] and VITA [Kim *et al.*, 2024],

Items	MIMIC-III	MIMIC-IV
# patients	6,350	75752
# clinical events	14,995	197,522
# diseases	1,958	2,000
# procedures	1,430	1,500
# medications	131	131
avg. # of visits	2.37	2.61
avg. # of medications	11.44	6.18

Table 1: Statistics of processed MIMIC-III and MIMIC-IV.

Following previous research [Yang *et al.*, 2021; Wu *et al.*, 2022; Yang *et al.*, 2023] on medication recommendations, we employ four widely-used metrics: Jaccard similarity, F1 score, PRAUC and DDI rate, to verify the accuracy and effectiveness of these models. More details of baselines and metrics are demonstrated in the Appendix C.

5.4 Main Result

The Table 2 presents a comparison of our method's performance with all baseline models. Overall, our present SSPNet outperforms the baseline models in Jaccard, F1 score and PRAUC. Our proposed SSPNet recommends drugs in a set-to-set manner. Hence, SSPNet can better represent the patient's health condition. Predicting drugs in a set form can avoid the influence of the order of predicted drugs. Furthermore, our PDRM module is capable of capturing the longitudinal historical medical records of patients. Therefore, our method can perform better than other methods.

Traditional classifiers, such as LR and ECC, which do not leverage deep learning, cannot model the complex medication combination recommendation, demonstrate poor performance. The multi-label classification-based medication recommendation methods, like GAMENet, neglect the importance of DDI. Consequently, SafeDrug and MoleRec incorporates the drug molecule structures and control DDI through a loss function, which resulting in improved performance. However, these methods aggregate the health status of patients, leading to the loss of patient health information, and moreover, these methods overlook the drugs previously used by the patients. Considering medication combination recommendation as a sequential decision-making task, LEAP difficult to model long medication decision-making. Using a Transformer, COGNet and VITA show improved performance. However, these methods impose an order on the medications, leading to unsatisfactory performance. VITA relies on patients' historical medical records, which makes it challenging to generalize to situations involving single visits.

5.5 Ablation Study

We conducted the following ablation experiments to validate the effectiveness of each module in SSPNet. The result of different variants of SSPNet is shown in Table 3.

Model	MIMIC-III					MIMIC-IV				
	Jaccard↑	**F1↑**	**PRAUC↑**	**DDI↓**	**Avg.#Med**	**Jaccard↑**	**F1↑**	**PRAUC↑**	**DDI↓**	**Avg.#Med**
LR	0.4920	0.6491	0.7552	0.0784	16.4293	0.4395	0.5850	0.7224	0.0764	8.2226
ECC	0.4868	0.6428	0.7382	0.0805	16.0100	0.4172	0.5575	0.7233	0.0754	7.4800
GAMENet	0.5110	0.6668	0.7642	0.0796	25.4674	0.4480	0.5995	0.7099	0.0838	16.2131
SafeDrug	0.5120	0.6685	0.7640	**0.0619**	20.4573	0.4457	0.5972	0.6917	**0.0504**	12.3890
MoleRec	0.5296	0.6837	0.7767	0.0726	21.2022	0.4562	0.6065	0.6929	0.0687	12.9925
LEAP	0.4441	0.6068	0.6475	0.0730	18.8987	0.4085	0.5584	0.5306	0.0673	10.6090
COGNet	0.5290	0.6823	0.7691	0.0813	27.8482	0.4682	0.6165	0.6883	0.0713	18.4972
VITA	0.5263	0.6786	0.7623	0.0765	29.2662	0.4716	0.6139	0.6792	0.0796	18.4972
SSPNet	**0.5517**	**0.7057**	**0.7947**	0.0773	21.6921	**0.4948**	**0.6423**	**0.7386**	0.0713	12.4286

Table 2: Performance Comparison on MIMIC-III and MIMIC-IV. The best results are highlighted in bold and runner-up are underline. ↑ indicates that higher values are preferable, while ↓ arrow signifies that lower values are more desirable. Avg.#Med denote the average number of drugs recommended per visit.

Model	MIMIC-III					MIMIC-IV				
	Jaccard↑	**F1↑**	**PRAUC↑**	**DDI↓**	**Avg.#Med**	**Jaccard↑**	**F1↑**	**PRAUC↑**	**DDI↓**	**Avg.#Med**
SSPNet *w/o* G_c	0.5504	0.7020	0.7925	0.0754	20.3192	0.4917	0.6392	0.7340	0.0739	12.3095
SSPNet *w/o* D	0.5243	0.6776	0.7712	**0.0725**	21.8232	0.4393	0.5885	0.6787	0.0707	11.9674
SSPNet *w/o* P	0.5308	0.6845	0.7780	0.0756	23.0812	0.4685	0.6189	0.6897	0.0705	13.1828
SAB $\longrightarrow$ Dual-RNN	0.5384	0.6909	0.7841	0.0785	22.0585	0.4560	0.6058	0.6937	**0.0685**	13.9377
SSPNet *w/o* PDRM	0.5414	0.6939	0.7830	0.0750	21.9193	0.4767	0.6260	0.7195	0.0771	12.7840
Set prediction $\longrightarrow$ Sequential prediction	0.5377	0.6906	0.7794	0.0738	21.4862	0.4638	0.6105	0.6299	0.6958	13.6584
SSPNet	**0.5517**	**0.7057**	**0.7947**	0.0773	21.6921	**0.4948**	**0.6423**	**0.7386**	0.0713	12.4286

Table 3: Ablation study on MIMIC-III and MIMIC-IV.

Medication Representation Module

We evaluated the validity of the EHR graph information. The results of SSPNet *w/o* G_c indicate that the EHR graph information contributes to the final outcome. The EHR graph encompasses co-occurrence information of medications and drug-drug interaction (DDI) information, which can enhance the representation of medications.

Patient Representation Module

We first assessed the validity of the patients' diagnoses and procedures information. We separately removed the information regarding diagnoses and procedures. As indicated by SSPNet *w/o* D and SSPNet *w/o* P, removes one of them lead to a diminished model performance.

To verify the advantages of our set encoders, we compared it with methods for modeling patient health conditions using aggregation. As show in Table 3, the SAB$\longrightarrow$Dual-RNN [Le *et al.*, 2018] indicates that after replacing the SAB module with the Dual-RNN (the details can referred to [Yang *et al.*, 2021; Yang *et al.*, 2023]), the performance of SSPNet decreased. Applying the set encoder to obtain the representation of patients' health status can better model the information regarding patients' health conditions.

Personalized Drug Representation Module

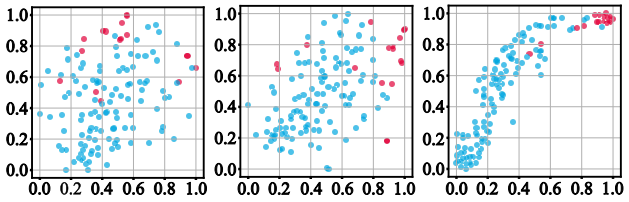
We have validated the superior performance of our proposed PDRM. As indicated by SSPNet *w/o* PDRM, the performance of SSPNet decreases when it is not equipped with the PDRM. The proposed PDRM effectively enhances the model's ability to capture historical medication information.

Drug Set Prediction Module

To verify the superiority of our ensemble prediction module, we replaced our medication set prediction module with the decoder module of a traditional Transformer, similar to that used in COGNet. The results are presented in Set prediction $\longrightarrow$ Sequential prediction. Recommending medications in the form of sets can avoid the sequential recommendation process which introduces specific medication orders, thus enhancing the accuracy of the model.

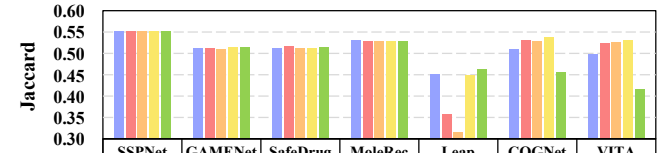
5.6 Case Study

We present a sample patient with historical visits to illustrate our model's drug recommendation. Figure 2 shows the visualization of drug representations through t-SNE [Van der Maaten and Hinton, 2008] under the initial, personalized, and predictive phases, corresponding to Equations (7), (20) and (22), respectively. In the initial phase, the drug representations have identical distributions for each patient. Many existing methods recommend drugs based on drug representations in the initial phase, which leads to suboptimal results. In the personalized phase, our PDMR module personalizes the drug representations according to the drugs the patient has previously used. Compared with the initial phase, the target drug representations are more separated from other drugs. Drug recommendations based on personalized drug representations can effectively improve the model accuracy. In the predictive phase, our permutation-invariant framework predicts the representations of recommended drugs based on the patient's health status set and the personalized drug representations.



(a) Initial (b) Personalized (c) Predictive

Figure 2: Visualization of drug representations at different phases. Red dots indicate the target drug; blue dots indicate other drugs.



	SSPNet	GAMENet	SafeDrug	MoleRec	Leap	COGNet	VITA
No process	0.5512	0.5110	0.5120	0.5297	0.4522	0.5094	0.4982
Frequent	0.5517	0.5126	0.5157	0.5285	0.3581	0.5304	0.5238
Rare first	0.5516	0.5092	0.5124	0.5291	0.3143	0.5290	0.5263
Vocabulary	0.5527	0.5141	0.5111	0.5283	0.4484	0.5385	0.5318
Random	0.5510	0.5131	0.5139	0.5276	0.4634	0.4566	0.4167

Figure 3: Performance of various methods on MIMIC-III across different label order.

Model	Jaccard	F1	PRAUC
MoleRec + PDRM	0.5372	0.6902	0.7740
SafeDrug + PDRM	0.5203	0.6762	0.7689
GAMENet + PDRM	0.5143	0.6694	0.7672

Table 4: The performance of different models equipped with our PDRM on MIMIC-III.

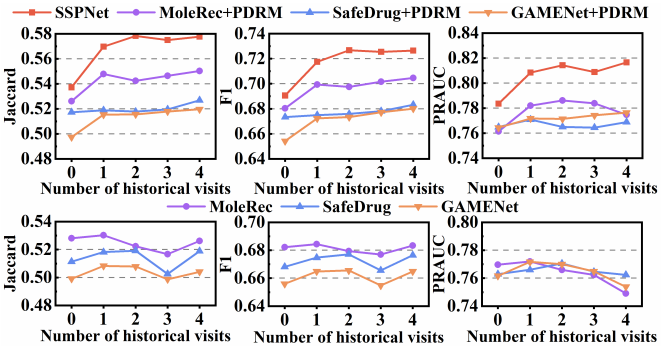


Figure 4: Performance of Various Models Across Varying Numbers of Historical Visits

5.7 The Stability of Set-to-Set Framework

We further assessed the performance of SSPNet, GAMENet, SafeDrug, MoleRec, Leap, COGNet and VITA under following five different input drug sequences:

- *No process* preserves the inherent sequence of drugs, without performing any manipulation.
- *Frequent* is sorted the order of the drugs in descending order based on their frequency.
- *Rare first* is arranged in ascending order by frequency.
- *Vocabulary* is sorted of drugs in alphabetical order.
- *Random* is randomly shuffled the order of the drugs.

The results are shown in Figure 3. Our model appropriately treats patient health information and medications as sets, and employs permutation-invariant encoders and decoders to predict drug prescriptions. Thus, our model exhibits stable performance and possesses superior robustness. MoleRec, SafeDrug, and GAMENet models aggregate patient diagnostic and surgical information to obtain patient health status, rather than treating patient health information as a set. This approach leads to a slight deficiency in the robustness of these models. VITA, COGNet and Leap which regard drug recommendation as a sequential decision-making process are sensitive to the medical implications of drug order.

5.8 Compatibility of PDRM

We validated the capability of our proposed PDRM module to capture historical medication information and its orthogonality. We applied the personalized drug representation model to MoleRec, SafeDrug and GAMENet. The performance is shown in Table 4. The performance of these models is enhanced after the integration of our PDRM module. We further evaluated the performance with varying numbers of historical visits. The results shown in Figure 4 as shown. Since most of the patients in the MIMIC-III data set have less than five visits [Wu *et al.*, 2022], we conducted experiments in patients with fewer than five visits. MoleRec and SafeDrug have only taken into account the current and historical health conditions of patients, neglecting the drugs that the patient has previously used. Upon the incorporation of the personal medication module, they are able to account for historical medication information. GAMENet has individually considered the relationship between patients' health status and medication as well as their historical drug use, leading to suboptimal outcomes. When GAMENet is equipped with our PDRM module, it can jointly consider the relationship between patients' health status, medication, and historical medication information. Our proposed PDRM not only shows exhibits flexibility in its application in various models, but also excellent performance across varying numbers of historical visits.

6 Conclusion

In this paper, we introduce SSPNet, a novel end-to-end framework designed to tackle the intricate task of automated medication recommendation. Unlike existing methods, which often overlook the essential set-based requirements, SSPNet employs a set-based approach to encode comprehensive patient health condition and predicts appropriate prescription through a permutation-consistent decoder. Furthermore, we have designed a personalized drug representation module with excellent compatibility to capture the medications previously used by patients. Our extensive evaluations on the MIMIC-III and MIMIC-IV datasets demonstrate that SSPNet outperforms existing state-of-the-art models.

Ethical Statement

The research presented in this paper utilizes the MIMIC dataset, which is a deidentified data set that is compliant with the Health Insurance Portability and Accountability Act (HIPAA). Under HIPAA regulations, the use of deidentified data for research purposes does not constitute human subjects research. The authors of this study have taken all necessary precautions to ensure that the research is conducted in an ethical manner, respecting the privacy and rights of the individuals represented in the dataset.

Acknowledgments

This work is partially supported by the Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No.2023B1212010007). Liang-Jie Zhang and Lin Ma are the corresponding authors.

References

[Ba *et al.*, 2016] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[Bhoi *et al.*, 2021] Suman Bhoi, Mong Li Lee, Wynne Hsu, Hao Sen Andrew Fang, and Ngiap Chuan Tan. Personalizing medication recommendation with a graph-based approach. *ACM Transactions on Information Systems (TOIS)*, 40(3):1–23, 2021.

[Carion *et al.*, 2020] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.

[Chen *et al.*, 2023] Qianyu Chen, Xin Li, Kunnan Geng, and Mingzhong Wang. Context-aware safe medication recommendations with molecular graph and ddi graph embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 7053–7060, 2023.

[Cowie *et al.*, 2017] Martin R Cowie, Juuso I Blomster, Lesley H Curtis, Sylvie Duclaux, Ian Ford, Fleur Fritz, Samantha Goldman, Salim Janmohamed, Jörg Kreuzer, Mark Leenay, et al. Electronic health records to facilitate clinical research. *Clinical Research in Cardiology*, 106:1–9, 2017.

[Evans, 2016] R Scott Evans. Electronic health records: then, now, and in the future. *Yearbook of medical informatics*, 25(S 01):S48–S61, 2016.

[Johnson *et al.*, 2016] Alistair EW 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. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[Johnson *et al.*, 2023] Alistair EW Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, et al. Mimic-iv, a freely accessible electronic health record dataset. *Scientific data*, 10(1):1, 2023.

[Kim *et al.*, 2024] Taeri Kim, Jiho Heo, Hongil Kim, Kijung Shin, and Sang-Wook Kim. Vita:'carefully chosen and weighted less' is better in medication recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8600–8607, 2024.

[Kingma and Ba, 2014] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Kosiorek *et al.*, 2020] Adam R Kosiorek, Hyunjik Kim, and Danilo J Rezende. Conditional set generation with transformers. *arXiv preprint arXiv:2006.16841*, 2020.

[Le *et al.*, 2018] Hung Le, Truyen Tran, and Svetha Venkatesh. Dual memory neural computer for asynchronous two-view sequential learning. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1637–1645, 2018.

[Lee *et al.*, 2019] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International conference on machine learning*, pages 3744–3753. PMLR, 2019.

[Li *et al.*, 2023] Chengxi Li, Yejing Wang, Qidong Liu, Xiangyu Zhao, Wanyu Wang, Yiqi Wang, Lixin Zou, Wenqi Fan, and Qing Li. Strec: Sparse transformer for sequential recommendations. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 101–111, 2023.

[Locatello *et al.*, 2020] Francesco Locatello, Dirk Weissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, and Thomas Kipf. Object-centric learning with slot attention. *Advances in neural information processing systems*, 33:11525–11538, 2020.

[Mi *et al.*, 2024] Jiacong Mi, Yi Zu, Zhuoyuan Wang, and Jieyue He. Acdnet: Attention-guided collaborative decision network for effective medication recommendation. *Journal of Biomedical Informatics*, 149:104570, 2024.

[Read *et al.*, 2011] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. *Machine learning*, 85:333–359, 2011.

[Shang *et al.*, 2019a] Junyuan Shang, Tengfei Ma, Cao Xiao, and Jimeng Sun. Pre-training of graph augmented transformers for medication recommendation. *arXiv preprint arXiv:1906.00346*, 2019.

[Shang *et al.*, 2019b] Junyuan Shang, Cao Xiao, Tengfei Ma, Hongyan Li, and Jimeng Sun. Gamenet: Graph augmented memory networks for recommending medication combination. In *proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1126–1133, 2019.

[Skou *et al.*, 2022] Søren T Skou, Frances S Mair, Martin Fortin, Bruce Guthrie, Bruno P Nunes, J Jaime Miranda, Cynthia M Boyd, Sanghamitra Pati, Sally Mtenga, and Susan M Smith. Multimorbidity. *Nature Reviews Disease Primers*, 8(1):48, 2022.

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

[Tan *et al.*, 2022] Yanchao Tan, Chengjun Kong, Leisheng Yu, Pan Li, Chaochao Chen, Xiaolin Zheng, Vicki S Hertzberg, and Carl Yang. 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*, pages 3970–3980, 2022.

[Tatonetti *et al.*, 2012] Nicholas P Tatonetti, Patrick P Ye, Roxana Daneshjou, and Russ B Altman. Data-driven prediction of drug effects and interactions. *Science translational medicine*, 4(125):125ra31–125ra31, 2012.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[Wang *et al.*, 2021] Yanda Wang, Weitong Chen, Dechang Pi, Lin Yue, Sen Wang, and Miao Xu. Self-supervised adversarial distribution regularization for medication recommendation. In *IJCAI*, pages 3134–3140, 2021.

[Wang *et al.*, 2024] Yanda Wang, Lin Yue, and Ying Li. Topological knowledge enhanced personalized ranking model for sequential medication recommendation. In *International Conference on Advanced Data Mining and Applications*, pages 109–124. Springer, 2024.

[Wu *et al.*, 2022] Rui Wu, Zhaopeng Qiu, Jiacheng Jiang, Guilin Qi, and Xian Wu. Conditional generation net for medication recommendation. In *Proceedings of the ACM Web Conference 2022*, pages 935–945, 2022.

[Yang *et al.*, 2021] Chaoqi Yang, Cao Xiao, Fenglong Ma, Lucas Glass, and Jimeng Sun. Safedrug: Dual molecular graph encoders for recommending effective and safe drug combinations. *arXiv preprint arXiv:2105.02711*, 2021.

[Yang *et al.*, 2023] Nianzu Yang, Kaipeng Zeng, Qitian Wu, and Junchi Yan. Molerec: Combinatorial drug recommendation with substructure-aware molecular representation learning. In *Proceedings of the ACM Web Conference 2023*, pages 4075–4085, 2023.

[Zaheer *et al.*, 2017] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

[Zhang *et al.*, 2017] Yutao Zhang, Robert Chen, Jie Tang, Walter F Stewart, and Jimeng Sun. 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*, pages 1315–1324, 2017.

[Zhang *et al.*, 2019] Yan Zhang, Jonathon Hare, and Adam Prugel-Bennett. Deep set prediction networks. *Advances in Neural Information Processing Systems*, 32, 2019.

[Zhang *et al.*, 2023] Yingying Zhang, Xian Wu, Quan Fang, Shengsheng Qian, and Changsheng Xu. Knowledge-enhanced attributed multi-task learning for medicine recommendation. *ACM Transactions on Information Systems*, 41(1):1–24, 2023.

[Zheng *et al.*, 2023] Zhi Zheng, Chao Wang, Tong Xu, Dazhong Shen, Penggang Qin, Xiangyu Zhao, Baoxing Huai, Xian Wu, and Enhong Chen. Interaction-aware drug package recommendation via policy gradient. *ACM Transactions on Information Systems*, 41(1):1–32, 2023.