TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 pp1–13
DOI: 10.26599/TST.2025.9010078
Volume x, Number x, xxxx xxxx

SDRec: Synergistic Drug Recommendation Network Based on Medical Domain Knowledge

Kang An, Mingyu Lu, Fei Chen, Xiaoxia Liu, and Yijia Zhang*

Abstract: Drug recommendation is a prominent topic in the medical field, aiming to suggest a set of safe drug combinations based on patients' clinical information and medical knowledge. Existing research has primarily focused on utilizing patients' electronic health records (EHR) and avoiding adverse drug-drug interaction (ADDI) for drug recommendation. However, they ignore the influence of drug synergy in the recommendation system. To address this limitation, we propose the Synergistic drug recommendation network based on medical domain knowledge (SDRec), which introduces the concept of drug synergy into the field for the first time. Specifically, we design a drug synergy module that extracts drug features from the perspective of molecular structures and models the interaction information between different drug pairs to provide safe and effective drug combinations for patients. Additionally, we incorporate a drug synergy graph derived from medical domain knowledge and model it using a graph convolutional network. Considering the safety of the recommended drug, we introduce a contrastive loss function during training to balance the features of drug synergy and adverse drug reactions, thereby minimizing potential side effects. Our experimental findings indicate that SDRec achieves notable performance enhancements compared to several baseline methods on the MIMIC-III and MIMIC-IV datasets.

Key words: drug recommendation; drug drug interaction; drug synergy

1 Introduction

In clinical practice, how to accurately recommend safe drugs is a difficult problem. When formulating treatment plans, physicians must comprehensively consider diagnoses, medical histories, and drug-drug interactions to ensure optimal therapeutic outcomes while avoiding adverse effects. Deep learning technology is constantly applied in various fields of medical treatment, including drug recommendation, drug relocation, gene synthetic lethality, etc. It is particularly important to improve the accuracy and safety of drug recommendation to help doctors assist in decision-making[1, 2]. The methods for recommending drugs fall into three categories: disease-drug rule-based, patient visit instance-based, and patient historical electronic health record (EHR)-based. The rule-based drug recommendation model relies on predefined rules and clinical treatment protocols set by domain experts, mapping medical knowledge to corresponding drugs[3, 4]. The instance-based models focus on recommending drugs directly based on the

• Kang An, Fei Chen, and Yijia Zhang are with School of Information Science and Technology, Dalian Maritime University, Dalian 116024, China. E-mail: 120221416_ankang@dlmu.edu.cn; chenf@dlmu.edu.cn; zhangyijia@dlmu.edu.cn.
• Mingyu Lu is with School of Artificial Intelligence, Dalian Maritime University, Dalian 116024, China. E-mail: lumingyu@dlmu.edu.cn.
• Xiaoxia Liu is with Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA 94305, USA. E-mail: xxliu@stanford.edu.
* To whom correspondence should be addressed.
Manuscript received: 2024-11-29; revised: 2025-01-19; accepted: 2025-05-07

clinical diagnosis and surgery of the patient in this visit, but these models often overlook the patient's medical history[5, 6]. The historical EHR-based models integrate historical medical records with external knowledge of drug interactions to ensure both safety and efficacy in drug usage[7–12]. However, a notable limitation in existing studies is the insufficient consideration of drug-drug synergy (DDS), despite its critical role in enhancing therapeutic outcomes.

In clinical practice, synergistic drug combinations are widely applied. The core concept of drug synergy is that the combined use of different drugs can mutually enhance therapeutic efficacy, resulting in better outcomes than using any single drug alone[13, 14]. As shown in Fig. 1, for hypertensive patients, using antihypertensive drugs alone may have limited efficacy. However, when combined with diuretics, blood pressure can be effectively controlled. This mechanism arises because antihypertensive drugs alleviate cardiac load, while diuretics reduce fluid and salt levels through urinary excretion. Together, they achieve better blood pressure management and reduce cardiac burden. Physicians often leverage such synergistic effects when formulating treatment plans to optimize outcomes, reduce dosages, and mitigate adverse drug reactions. Motivated by this concept, we explore how to fully incorporate drug synergy into drug recommendation to better guide clinical decision-making and improve therapeutic outcomes.

This research presents the Synergistic drug recommendation network based on medical domain knowledge (SDRec), an innovative synergistic drug recommendation network that incorporates medical domain knowledge and pioneers the use of drug synergy. A specialized drug synergy module is developed to analyze and capture drug characteristics at the molecular structure level, enabling the modeling of interactions between drugs to identify the most effective combinations for patients. Furthermore, we employ graph convolutional networks (GCNs)[15] to model domain knowledge, including the DDS graph, the EHR graph, and the drug-drug interaction (ADDI) graph. Synergistic drug features benefit patients, while adverse interaction features pose risks. To dynamically balance these features, we adopt a contrastive loss function to maximize the retention of synergistic features while minimizing potential harm from adverse interactions. Comprehensive experiments on the MIMIC-III and MIMIC-IV datasets demonstrate that our proposed model SDRec achieves substantial improvements over previous baselines.

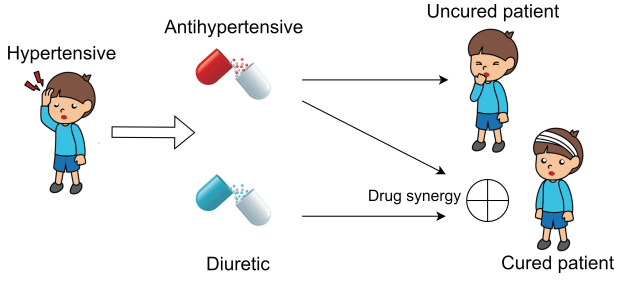

Fig. 1 An example of drug synergy.

The main contributions of this study are as follows:

(1) We introduce SDRec, a synergistic drug recommendation network that is the first to integrate the concept of drug synergy. Utilizing GCNs to model the drug synergy graph, SDRec enhances the learning of synergistic features while reducing adverse effects, improving the accuracy and safety of drug recommendations.

(2) We develop a drug synergy module that captures drug features at the molecular structure level, modeling interactions between drug pairs. This allows for the identification of effective synergistic drug combinations to optimize treatment outcomes.

(3) Experiments conducted on the MIMIC-III and MIMIC-IV datasets validate the effectiveness of our model, with results showing remarkable performance gains compared to previous baselines.

2 Related Work

This section covers related work, focusing on joint drug recommendation and the representation of drug molecular structures.

2.1 Joint drug recommendation

The methods for recommending drugs fall into three categories: disease-drug rule-based, patient visit instance-based, and patient historical EHR-based.

2.1.1 Rule-based drug recommendation

Rule-based methods recommend drugs based on predefined rules, typically requiring domain experts to establish specific guidelines and clinical treatment protocols. The system maps medical knowledge to corresponding drugs. For example, Himabindu et al.[3] used decision lists to map diagnostic test results to treatment plans. Mehnaz et al.[4] proposed an information decision system that combines natural language processing with ontology-based semantic

annotation to provide drug recommendations. However, rule-based methods are traditional, heavily reliant on manual input and design, and often fail to accurately recommend drugs or account for drug safety. As deep learning emerged, rule-based approaches gradually became obsolete.

2.1.2 Instance-based drug recommendation

The instance-based approaches recommend drugs directly based on the clinical diagnosis and surgery of the patient in this visit. For example, LEAP[6] takes the diagnose information of a patient's current visit as input and uses a recurrent decoder to serialize recommended drugs based on diseases with different weights. SMR[5] constructs a heterogeneous graph from electronic prescription records and medical knowledge graphs, embedding the graph to reformulate drug recommendation as a link prediction task while also considering adverse drug reactions. However, these methods only model a patient's current clinical information without comprehensively modeling the patient's representation, limiting their effectiveness in drug recommendation.

2.1.3 Historical EHR-based drug recommendation

Historical EHR-based methods incorporate patients' temporally dependent historical medical records and medical knowledge of drug-drug interactions to provide comprehensive patient modeling. For instance, GAMENet[8] introduces the ADDI knowledge graph into drug recommendation for the first time, using the ADDI graph as a memory module. SafeDrug[11] uses both global and local modules to comprehensively model the structure of drug molecules. MoleRec[10] models relationships between drug molecules and substructures to identify substructures truly effective for treating diseases. MICRON[16] proposes a novel recurrent residual network to predict changes in drug use. COGNet[12] leverages historical diagnostic and drug information with a copy-and-predict mechanism for drug recommendation. DGCL[9] uses distance-aware loss to model the differences between current cases and historical records, employing graph contrastive learning to train adverse drug-drug interaction graphs and electronic health records. DKINet[17] captures domain knowledge and integrates it with patients' clinical presentations. Carmen[7] incorporates patients' historical visits into molecular graph representation learning, and ACDNet[18] uses an attention mechanism to model patients' historical visits on both global and local levels. While these methods effectively model patient representations, incorporating historical visits, current clinical presentations, and ADDI graphs to reduce adverse drug reactions, they overlook the critical importance of drug synergy in treatment planning. To address this, we propose SDRec, which not only models a drug synergy knowledge graph but also introduces a drug synergy module to maximize the benefits of synergistic drug use.

2.2 Representation of drug molecular structures

For drug molecular structure representation, early work focused on directly learning adaptive "descriptors" that are highly sensitive to molecular topology from molecular graphs rather than relying on predefined descriptors[19]. The rise of deep learning has given birth to two new representation approaches[20]: **(1) SMILES-based representation:** This approach treats the Simplified Molecular Input Line Entry System (SMILES) representation[21], which retains raw molecular information, as sequential data. It uses recurrent neural networks (RNNs)[22] for modeling or variational autoencoders[23] for encoding. **(2) Graph-based representation:** This approach models drug molecular graphs as graph-structured data, where atoms represent nodes and chemical bonds represent edges. It uses graph neural networks (GNNs)[24] to learn drug molecules with different structures, representing structured data as serialized embeddings. In our model SDRec, we use the latter representation approach to model the drug molecular structures.

3 Problem Formulation

3.1 EHR

In the MIMIC dataset, each patient has at least one hospital visit record. The visit records of patient x can be represented as: $V_x = \{v_x^1, v_x^2, \ldots, v_x^N\}$, where N represents the number of hospital visits of patient x. The EHR of patient x can be represented as: $E_x = \{d_x^t, p_x^t, m_x^t\}$, where t denotes the t-th visit of patient x, d_x^t represents the diagnostic records of the patient, p_x^t represents the procedural records of the patient, and m_x^t represents the medication records of the patient.

3.2 DDS, ADDI, and EHR graph

We represent the DDS graph, ADDI graph, and EHR graph using adjacency matrices $G_{\text{dds}} = \{A, I_{\text{dds}}\}$,

$G_{\text{ADDI}} = \{A, I_{\text{ADDI}}\}$, and $G_{\text{ehr}} = \{A, I_{\text{ehr}}\}$ respectively. Here, A represents the set of drugs, consisting of 131 different medications. I denotes the edges between drugs. Specifically, $I_{\text{dds}}[p, q] = 1$ indicates a synergy between drug p and drug q, $I_{\text{ADDI}}[p, q] = 1$ indicates an interaction between drug p and drug q, namely antagonistic interaction, and $I_{\text{ehr}}[p, q] = 1$ indicates that drugs p and q have been co-recommended simultaneously.

4 Method

As shown in Fig. 2, the SDRec model is divided into five modules: **(1) Patient representation module:** This module focuses on modeling patients' diagnostic, procedural, and prescription records. By integrating data from these aspects, the module provides a holistic understanding of a patient's overall health status. **(2) Drug Synergy Module:** The module models the molecular structure of drugs. By learning the interactions between drug molecules, it dynamically adjusts the drug molecular synergy matrix to achieve more accurate representations of drug molecules. This facilitates uncovering synergistic relationships among drugs and improves the precision of drug recommendations. **(3) Domain knowledge representation module:** In this module, the DDS graph, EHR graph, and ADDI graph are modeled. In dynamic learning of these graph features, we introduce the contrastive loss function. This comprehensive modeling approach enhances the understanding of the interconnections among different domain knowledge sources, thereby providing more robust information for drug recommendations. **(4) Drug prediction module:** This module combines patient health data with drug features to predict and recommend medications suited to the patient's clinical requirements. **(5) Training and inference module:** The training process incorporates multiple loss functions to ensure a balance between accuracy and safety in drug recommendations. This ensures that the model considers both the efficacy and potential adverse effects of drugs during the recommendation process. The design of this module enhances the model's robustness and practical utility, providing more reliable support for clinical decision-making.

4.1 Patient representation module

In this module, we model the longitudinal electronic health records of patients, which consist of diagnostic records, procedural records, and medication records. Diagnostic and procedural records, i.e., disease information, are encoded with the International Classification of Diseases (ICD) codes, while medication records utilize the Anatomical Therapeutic Chemical (ATC)※ codes for encoding.

4.1.1 Representation of diagnosis, procedure, and prescription

※https://www.who.int/tools/atc-ddd-toolkit/atc-classification

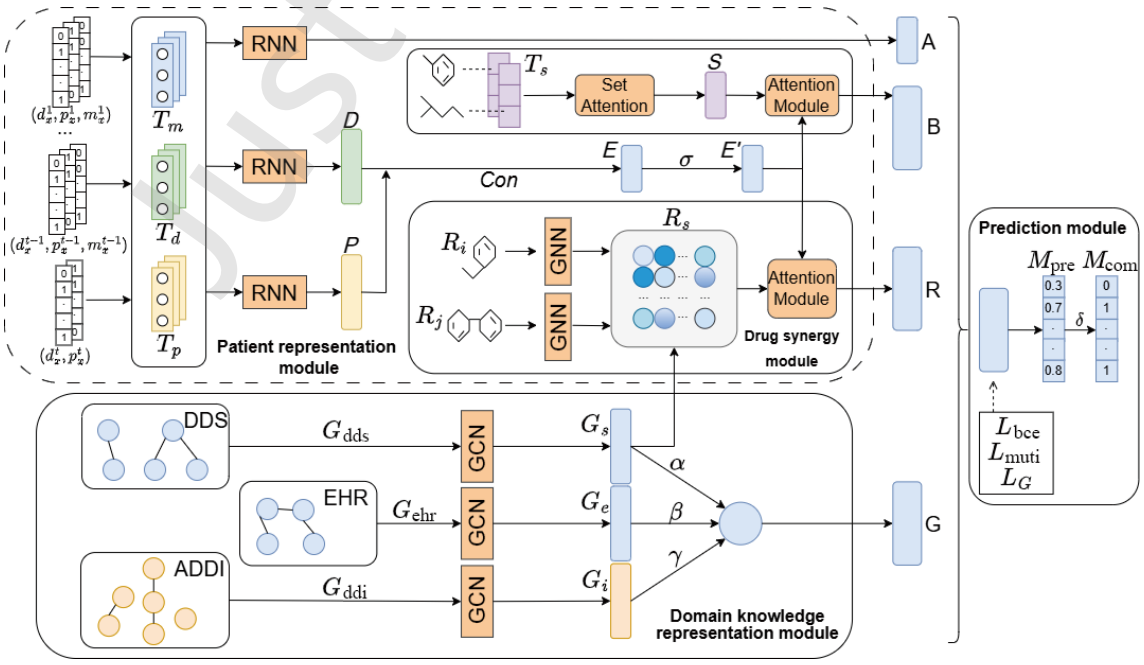

Fig. 2 Synergistic drug recommendation network based on medical domain knowledge.

For the t-th visit of patient x, we model the diagnostic records and procedural records of the t-th visit, and the drug records from the $(t-1)$-th visit to capture historical drug information. Specifically, for the patient's first visit, the historical drug information is set to null. These records are represented as $\{d_x^t, p_x^t, m_x^{t-1}\}$. We design three embedding tables, $T_d \in R^{|D|\times d}$, $T_p \in R^{|P|\times d}$, and $T_m \in R^{|M|\times d}$, where d represents the embedding dimension, and $|D|$, $|P|$, and $|M|$ represent the number of diagnoses, procedures, and drugs, respectively. The visit records are mapped to the embedding space through vector dot products to obtain the corresponding embedding vectors,

$$v_d^t = d_x^t \cdot T_d \tag{1}$$

$$v_p^t = p_x^t \cdot T_p \tag{2}$$

$$v_m^t = m_x^{t-1} \cdot T_m \tag{3}$$

To comprehensively model the longitudinal EHRs, we employ a Recurrent Neural Network (RNN)[22], which effectively captures sequential dependencies and contextual information. Using a gated recurrent unit (GRU)[25], the longitudinal EHRs are encoded as follows:

$$o_d^t, h_d^t = \mathrm{GRU}_d(v_d^t, h_d^{t-1}) \tag{4}$$

$$o_p^t, h_p^t = \mathrm{GRU}_p(v_p^t, h_p^{t-1}) \tag{5}$$

$$o_m^t, h_m^t = \mathrm{GRU}_m(v_m^t, h_m^{t-1}) \tag{6}$$

where o^t and h^t denote the output and hidden states of the RNN, respectively. We integrate the patient's diagnostic and procedural information as a patient representation,

$$E = \mathrm{Con}([h_d, h_p], [o_d, o_p]) \tag{7}$$

where Con denotes the concatenation function. Similarly, we represent historical drug information as

$$A = \mathrm{Con}(o_m, h_m) \tag{8}$$

4.1.2 Drug substructure representation

In the drug substructure representation, we explore the relationship between drug efficacy and specific drug substructures. Since there may be synergistic interactions among substructures, accurately modeling these relationships is crucial. First, we decompose drug molecules into distinct substructures using the breaking retrosynthetically interesting chemical substructures (BRICS)[26] method. Subsequently, we employ a set attention block (SAB)[27], which leverages the permutation-invariant property of set data to process the embedding table and capture the synergy among substructures. This process yields

$$S = \mathrm{SAB}(T_s) = \mathrm{LA}(U + \mathrm{FF}(U)) \tag{9}$$

$$U = \mathrm{LA}(T_s + \mathrm{SATT}(T_s)) \tag{10}$$

where T_s represents the drug substructure embedding table, LA denotes layer normalization, FF represents a feedforward neural network, and U represents the intermediate variable. SATT represents the self-attention mechanism:

$$\mathrm{SATT}(X) = \mathrm{Softmax}\left(\frac{Q_X K_X^{\mathrm{T}}}{\sqrt{d_k}}\right) V_X \tag{11}$$

where Q_X, K_X, and V_X denote the query, key, and value matrices, respectively, and d_k is the dimension. Using the attention mechanism, we identify substructures that are most effective for treating the disease, resulting in

$$B = \mathrm{ATT}(E', S) \tag{12}$$

$$E' = \sigma(\mathrm{FF}(E)) \tag{13}$$

$$\mathrm{ATT}(a, b) = \mathrm{Softmax}\left(\frac{Q_a K_b^{\mathrm{T}}}{\sqrt{d_k}}\right) \tag{14}$$

where σ is an activation function and $\mathrm{ATT}(a, b)$ is the attention function.

4.2 Drug synergy module

The structural properties of drug molecules are critical in drug design as they directly influence efficacy and potency. We adopt an advanced approach to represent drug molecules while considering their interactions. Unlike traditional neural networks, GNNs capture complex relationships between nodes and edges, enabling global representation of the graph structure. By modeling node features, GNNs aggregate information from both the nodes themselves and their neighbors, propagating and learning information across the graph to iteratively update the global representation of each node. This aggregation process can be expressed as

$$f_n^i(p) = AGG(f_n^{i-1}(p), f_n^{i-1}(q), f_e^{i-1}(p, q) \mid C[p, q] = 1) \tag{15}$$

where $C[p, q] = 1$ denotes an edge (i.e., a molecular bond) between nodes p and q, $f_e^{i-1}(p, q)$ represents the

feature of the edge connecting p and q, and $f_n^i(p)$ denotes the i-th aggregated feature of node p. We convert the SMILES molecular formula[21] of each drug into a molecular graph structure, where chemical elements are represented as nodes and chemical bonds as edges. The molecular graph is denoted as $R = \{N, C\}$, where N represents the nodes and C represents the edges. We utilize a 5-layer GNN to model the molecular graphs of drugs R_i and R_j:

$$O_i = \mathrm{GNN}(f_n^{(5)}(u)|u \in N_i) \tag{16}$$

$$O_j = \mathrm{GNN}(f_n^{(5)}(u)|u \in N_j) \tag{17}$$

where $u \in N$ denotes all nodes in the graph undergoing aggregation and update operations. To capture drug synergy, we design a learnable drug synergy matrix R_s, initialized based on G_{dds}. When there is synergy between drugs R_i and R_j, there is an edge in the G_{dds} graph, where $R_s[i, j] = 1$ and conversely $R_s[i, j] = 0$.

$$R_s = \mathrm{FF}((w_1 O_i)^T \cdot w_2 O_j) \tag{18}$$

where FF denotes a feedforward neural network, and w_1, w_2 are learnable parameters. Using an attention mechanism, we derive the synergistic representation between drug molecules:

$$R = \mathrm{ATT}(E', R_s) \tag{19}$$

$$E' = \sigma(FF(E)) \tag{20}$$

4.3 Domain knowledge representation module

In drug research, drugs may exhibit synergistic effects but can also induce severe adverse reactions. Effectively modeling the DDS graph, the EHR graph, and the ADDI graph to learn their critical features is vital for patient safety and treatment efficacy. For the drug synergy graph, we leverage DDInter[28] as a resource to organize drug synergy relationships into adjacency lists, converting them into graph structures. GCN propagates node features and connection relationships through convolution operations, generating new node representations that integrate the features of both the node itself and its neighbors, enriching semantic information. GCN can be generally defined as

$$\mathrm{GCN}(X, W) = \sigma\left(\widetilde{P}^{-\frac{1}{2}}\widetilde{W}\widetilde{P}^{-\frac{1}{2}}XQ\right) \tag{21}$$

where $X \in R^{N \times d}$, W denotes the adjacency matrix, $\widetilde{W} = W + I$, and I is the identity matrix. $\widetilde{P}$ is the degree matrix of $\widetilde{W}$, and Q is a learnable parameter matrix. The DDS graph, EHR graph, and ADDI graph are modeled by GCN, resulting in

$$G_s = \mathrm{GCN}(\mathrm{RELU}(\mathrm{GCN}(X_s, W_s)), W_s) \tag{22}$$

$$G_e = \mathrm{GCN}(\mathrm{RELU}(\mathrm{GCN}(X_e, W_e)), W_e) \tag{23}$$

$$G_i = \mathrm{GCN}(\mathrm{RELU}(\mathrm{GCN}(X_i, W_i)), W_i) \tag{24}$$

where $\mathrm{RELU}(x) = \max(0, x)$ denotes the rectified linear unit, which introduces nonlinearity to capture complex feature interactions. G_s represents the synergistic effect of drug molecules, G_e represents the coexistence of drug molecules in the patient's electronic health record, and G_i represents the harmful interactions between drug molecules. We use the contrastive loss function[29] to extract positive features from the EHR and DDS graphs while reducing negative features from the ADDI graph. During training, the relative distances among the three graphs are iteratively adjusted,

$$D = \|(G_s + G_e) - G_i\|_2 \tag{25}$$

$$\mathcal{L}_G = (1 - Z)D^2 + Z \cdot \frac{1}{2}(\max\{0, g - D\})^2 \tag{26}$$

where $\|\cdot\|_2$ denotes the L2 norm (euclidean distance) of a vector, D represents the Euclidean distance between vectors, g denotes the graph margin, and $Z = 1$ indicates drug synergy without adverse interactions; otherwise, $Z = 0$. We get the final representation of knowledge in the medical field,

$$G = \alpha G_s + \beta G_e - \gamma G_i \tag{27}$$

where α, β, and γ are learnable dynamic parameters.

4.4 Drug prediction module

In the drug prediction module, we incorporate multiple features, including the attention representation of drug substructures for diseases B, the synergistic representation of drug molecules R, patient historical medication information A, and the drug knowledge representation G. The dimensions of these four features obtained in the previous modules are consistent. We connect these features and get the corresponding drug prediction score,

$$M_{\mathrm{pre}} = \sigma(\mathrm{FF}(\mathrm{Con}(B + R + A + G))) \tag{28}$$

where FF is a feedforward neural network and Con denotes concatenation. We set a threshold $\delta = 0.5$, and the final recommended drug set M_{com} is the drug whose predicted score is greater than the threshold.

4.5 Training and inference

In the training phase, we use three loss functions. The multi-label margin loss and the binary cross-entropy loss are defined as follows:

$$\mathcal{L}_{\text{mul}} = \frac{1}{|M|} \sum_{j \in y_n} \sum_{i \notin y_n} \max(0,\ 1 - x_n[j] + x_n[i]) \tag{29}$$

$$\mathcal{L}_{\text{bce}} = -\sum_{i=1}^{|M|} y_i \times \log(\sigma(x_i)) + (1 - y_i) \times \log(1 - \sigma(x_i)) \tag{30}$$

where $|M|$ is the total number of drugs, x denotes the prediction results, derived from M_{com}, and y represents the ground truth labels. Additionally, we employ the contrastive loss $\mathcal{L}_G$ in the drug knowledge representation module to extract positive features. The loss function is defined as

$$\mathcal{L} = \eta(\theta \mathcal{L}_{\text{mul}} + (1 - \theta)\mathcal{L}_{\text{bce}}) + (1 - \eta)\mathcal{L}_G \tag{31}$$

where η and θ are hyperparameters controlling the contributions of each loss component. Given the inherent presence of adverse drug-drug interactions in the dataset, the parameter η is used to specifically adjust $\mathcal{L}_G$ to achieve more accurate drug recommendations.

5 Experiment

We carried out comparative experiments to benchmark the SDRec model against state-of-the-art methods on two datasets. To evaluate the contribution of each module, we conducted ablation studies and case analyses. Additionally, we analyzed the limitations of the proposed model.

5.1 Experimental setup

In this study, we set the learning rate of the Adam optimizer to 0.0005, and the dropout ratio to 0.7. We set the hyperparameters η and θ as 0.95 and 0.05, respectively.

5.1.1 Datasets

For our experiments, we utilized the publicly available MIMIC-III[30] and MIMIC-IV[31] datasets. The MIMIC datasets contain information about patients in the hospital's ICU. These datasets include rich medical data, such as patient diagnostic records, surgical details, prescription information, hospitalization durations, and impact reports. To ensure data accuracy and consistency, we preprocessed the MIMIC-III dataset following the methodology of SafeDrug[11], resulting in a dataset comprising 6350 patients, 1958 diagnostic categories, 1430 surgical categories, and 131 medications. Similarly, the MIMIC-IV dataset was preprocessed in accordance with the methodology used in the Carmen[7] study, yielding a dataset containing 9036 patients, 1892 diagnostic categories, 4939 surgical categories, and 131 medications.

5.1.2 Evaluation metrics

We selected the following commonly used evaluation metrics to assess the model's performance:

(1) Jaccard similarity score (Jaccard): Evaluates the overlap between the predicted and true drug sets.

(2) Precision-recall area under curve (PRAUC): Summarizes the precision-recall tradeoff for the predicted drugs.

(3) F1 score (F1): Assesses the balance between precision and recall across all classes.

5.2 Baseline methods

In this paper, we compare our model SDRec† with several methods, including:

(1) LEAP[6]: LEAP takes the diagnosis information of a patient's current visit as input and uses a recurrent decoder to serialize recommended drugs based on diseases with different weights.

(2) GAMENET[8]: GAMENET focuses on modeling patients' longitudinal EHR and introduces an ADDI knowledge graph for the first time. It incorporates ADDI knowledge into the model's memory module via graph convolutional networks.

(3) SafeDrug[11]: SafeDrug uses both global and local modules to comprehensively model the structure of drug molecules. This approach is the first in the field to model drug molecular structures while introducing an ADDI-controllable loss function.

(4) MoleRec[10]: MoleRec focuses on drug molecular structure and drug substructure by modeling the interactions between drug substructures and the connections between substructures and specific drugs to find drug substructures that are effective for patients.

(5) Carmen[7]: In order to distinguish between drugs with different functions but similar molecular structures, Carmen fuses the historical visit representation into the drug molecular diagram representation and learns to distinguish between different drug molecules.

(6) COGNet[12]: COGNet uses the multi-head

†https://github.com/AK-321/SDRec

attention mechanism in the patient representation stage and introduces the replication and prediction mechanism in the prediction stage. By retrieving the patient's historical information, including the historical medical information and the historical medication information, COGNet decides whether to recommend the past drug or the new drug directly.

(7) DGCL[9]: DGCL uses a distance-based detection loss function to model discrepancies between current diagnoses and historical records. It employs graph contrastive learning to jointly train ADDI and EHR graphs, effectively controlling for ADDI.

5.3 Result analysis

As shown in Table 1, we conducted a large number of experiments on two MIMIC datasets and analyzed the experimental results. We mainly selected three evaluation indicators: Jaccard, F1, and PRAUC. Our proposed model, SDRec, was superior to the mainstream drug recommendation models in terms of indicators, showing remarkable performance improvement and verifying the effectiveness of the proposed drug synergy mechanism.

Specifically, LEAP, a typical model in the instance-based drug recommendation approach, fails to fully leverage patients' historical information, recommending drugs solely based on their current visit, which limits its recommendation accuracy. In contrast, GAMENET models longitudinal patient visit records and simultaneously uses GCN to model EHR and ADDI graphs, which was the first time that ADDI knowledge was introduced in the field, effectively reducing the occurrence of drug side effects. SafeDrug, MoleRec, and Carmen focus on modeling drug molecules, with SafeDrug and MoleRec providing detailed representations of drug substructures, while Carmen incorporates historical visit information into molecular representation learning. These innovative approaches offer new perspectives on studying the chemical properties of drug molecules and achieve commendable performance. COGNet emphasizes that the modeling of patient historical medication information is particularly helpful for patients with long-term medication, while the use of attention mechanisms to comprehensively model patient representation further improves its performance. DGCL uses graph contrastive learning to ensure the safety of recommended drugs and improve the accuracy of drug recommendation. Our proposed model incorporates medical domain knowledge about drug synergy and thoroughly considers synergistic relationships between drug molecules and their substructures. It demonstrates superior performance on both the MIMIC-III and MIMIC-IV datasets, further establishing its effectiveness and leading position in the domain of drug recommendation.

Table 1 Performance comparison on MIMIC-III and MIMIC-IV. The best score results are highlighted in bold.

Model	MIMIC-IV			MIMIC-III		
	Jaccard (%)	Prauc (%)	F1 (%)	Jaccard (%)	Prauc (%)	F1 (%)
LEAP	43.50	63.21	59.63	45.44	65.71	61.63
GAMENET	47.84	73.95	63.63	51.59	76.84	67.15
SafeDrug	48.57	74.00	64.35	50.35	75.76	66.14
MoleRec	48.85	73.97	64.52	53.05	77.36	68.43
Carmen	50.06	74.62	65.69	52.67	76.52	68.12
COGNET	49.67	74.07	65.13	53.36	77.39	68.69
DGCL	50.54	76.03	66.02	53.42	78.24	68.74
SDRec	**51.14**	**76.31**	**66.56**	**54.25**	**78.58**	**69.54**

5.4 Ablation study

To better understand the contribution of each component in our model, we conducted an ablation study by designing several ablation models to evaluate the impact of each module on overall performance. The ablation models and their effects are as follows:

(1) SDRec w/o *B*: The model excludes drug substructure representations and ignores associations between drug substructures and specific diseases.

(2) SDRec w/o *A*: This model excludes the representation of the patient's history of medication, and the patient representation consists only of diagnostic and surgical information.

(3) SDRec w/o *G*: This model excludes the drug synergy graph from the external knowledge of medicine and retains only the traditional EHR graph and ADDI graph.

(4) SDRec w/o *R*: This model excludes the drug synergy module and ignores the synergies between drugs.

Table 2 presents the performance comparison of these ablation models. Notably, the w/o *R* model shows the most remarkable performance drop, which indicates that the drug synergy module plays a key role in the whole model and also indicates that the synergistic drug combinations are closely related to the patient representation. The results of the w/o *G* model show a decline in the accuracy of recommendations using only

Table 2 Performance comparison of ablation studies. The best score results are highlighted in bold.

Model	MIMIC-IV			MIMIC-III		
	Jaccard (%)	Prauc (%)	F1 (%)	Jaccard (%)	Prauc (%)	F1 (%)
SDRec w/o *B*	50.86	75.91	66.33	53.98	78.30	69.32
SDRec w/o *A*	50.82	75.89	66.29	53.92	78.27	69.23
SDRec w/o *G*	50.71	75.79	66.16	53.88	78.21	69.17
SDRec w/o *R*	50.58	75.67	66.08	53.72	78.07	69.03
SDRec	**51.14**	**76.31**	**66.56**	**54.25**	**78.58**	**69.54**

the EHR and ADDI graphs in traditional models when modeling medical domain knowledge, further emphasizing the actual validity of drug synergy graphs. The w/o *A* model results demonstrate that neglecting patients' historical medication information and relying solely on diagnostic information yields incomplete patient representations, leading to decreased performance. This underscores the importance of historical medication data, especially for chronic disease patients. Finally, the w/o *B* model results show that removing the substructure module and disregarding the interactions between substructures leads to a slight performance decline, validating the impact of specific substructures on disease treatment. In summary, the ablation study confirms the effectiveness of drug synergy, molecular synergy, and historical medication records in our model, providing a clearer understanding of the contributions of each component.

5.5 Case study

To verify the application of the proposed model in real-world patient cases, we conducted a case study. Using actual patient case data, we compared two approaches: one that includes the drug molecular synergy module and another that excludes it (SDRec w/o *R*). Table 3 shows the recommended prescriptions for the patient's visits and the correct prescriptions, described using Anatomical Therapeutic Chemical 4 codes.

We selected a patient case where the patient had two clinical visits. During the first visit, the absence of historical drug records led to lower recommendation accuracy in both approaches. However, in the second visit, the availability of the patient's historical drug information resulted in improved recommendation accuracy for both methods. Notably, in our model, drugs C08D (selective calcium channel blockers acting directly on the heart) and C03C (high-ceiling diuretics) were simultaneously recommended, whereas they were not recommended by the model without the drug molecular synergy module. These two drugs exhibit synergistic effects: C08D dilates blood vessels and reduces myocardial oxygen demand by blocking calcium ion channels, while C03C lowers blood pressure by promoting sodium excretion. Consequently, their combination is commonly used in treating hypertension. These real-world case results demonstrate that our model can learn associations between drugs and recommend safe and effective synergistic drug combinations.

5.6 Visual analysis

Based on the case study, we conducted a visual analysis of the model. During the patient's second visit, the SDRec model recommended a combination of drugs along with the patient's actual prescription, totaling 24 drugs. We visualized the synergy scores of these 24 drugs using a heatmap, where the synergy scores were obtained from the drug synergy matrix R_s during the model training process.

The drug synergy heatmap for the 24 drugs is shown in Fig. 3, where darker colors indicate higher drug synergy scores. We can intuitively observe that most drugs do not exhibit significant drug synergy, with only a few pairs showing synergistic effects. In the case study, we focused on the drugs C08D (selective calcium channel blockers acting directly on the heart) and C03C (high-ceiling diuretics), which had relatively

Table 3 Case study of a patient's medication visit and recommended results. Correct drug predictions are highlighted in bold.

Method	SDRec w/o *R*	SDRec	Actual drugs
Visit1	**C08C**, J01A, **B05C**, B03B, N06B, **C01D**, **V04C**, **V03A**, **C01E**, D04A, **A12C**, **S01K**, **C10A**, J02A, **N05C**, **B02B**	**C08C**, **C03C**, **B05C**, B03B, R03A, **C01D**, A07D, **V03A**, **C01E**, **C08D**, **A12C**, D04A, **C10A**, **A12A**, **N05C**, J02A	C08C, C03C, B05C, G04B, D06A, C01D, V04C, V03A, C01E, C08D, A12C, S01K, C10A, A12A, N05C, B02B
Visit2	**C03C**, **C07A**, P01B, **D06A**, **C01D**, **C08C**, **V03A**, **C01E**, N06B, **S01K**, **V04C**, **J06B**, D04A, **C03A**, N02C, **C10A**, N04B, **N05C**, **J01X**, **B02B**	**C03C**, **C07A**, P01B, **D06A**, **C01D**, **C08C**, **V03A**, **C01E**, B03B, D04A, **V04C**, **J06B**, **C08D**, **C03A**, **A12C**, **C10A**, N04B, **N05C**, **J01X**, **B02B**	C03C, C07A, B05C, D06A, C01D, C08C, V03A, C01E, G04B, S01K, V04C, J06B, C08D, C03A, A12C, C10A, A12A, N05C, J01X, B02B

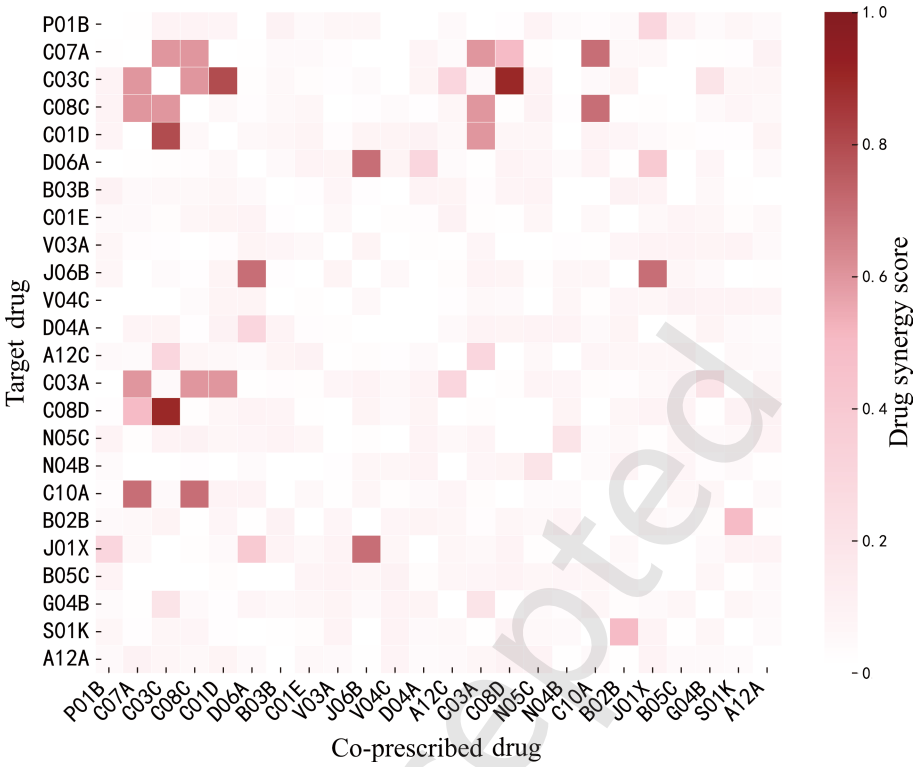

Fig. 3 Drug synergy score heatmap.

high synergy scores, as expected. We found that the drugs C08C (selective calcium channel blockers with a primary effect on blood vessels), C07A (beta-blocking agents), C03C (high-ceiling diuretics), C03A (low-ceiling diuretics, thiazides), and C10A (lipid-modifying agents, plain) were recommended simultaneously. In the row corresponding to drug C08C in the heatmap, these drugs all showed relatively high synergy scores with C08C, further demonstrating the significance of our model's drug synergy module. Additionally, we observed that the drug P01B (antimalarials) was incorrectly recommended by our model. The heatmap analysis revealed that P01B (antimalarials) and J01X (other antibacterials) had low synergy scores. Overall, the visual analysis further validated our model's ability to integrate and learn drug synergy knowledge, thereby improving the accuracy of drug recommendations.

5.7 Error analysis

To ensure the safety of drug recommendations, we analyzed the recommendation results by examining the drug labels in the MIMIC-III dataset. As shown in Fig. 4, statistical analysis of the 131 drugs revealed that 15 drugs were used more than 4000 times, while 25 drugs were used fewer than 200 times. This imbalance in label frequency likely causes drugs frequently

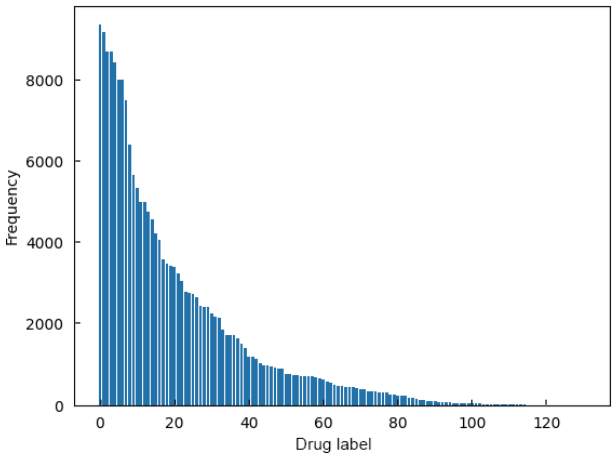

Fig. 4 Distribution of drug labels.

appearing in the dataset to be recommended more often, whereas infrequently used drugs may be overlooked due to lower recommendation scores—a common long-tail distribution issue.

In our case studies, the drug N04B (dopaminergic agents), which is widely used as a first-line emergency drug, was frequently recommended incorrectly, with a high recommendation score in the model. This erroneous recommendation validates our dataset analysis and highlights a critical challenge in drug recommendation systems: the insufficient

representation of rare drugs in the dataset may lead to under-recommendation or misrecommendation of these drugs.

In a further study, we plan to construct a memory module for storing patient consultation data with various characteristics. Since the diversity and variability of the data can affect the effectiveness of recommendations, we use a global greedy farthest point sampling scheme to select 300 benchmark patients with the largest differences from the patient dataset. Based on this, we obtain a benchmark patient dataset containing rare diseases and rare drugs, and calculate the similarity between patients and benchmark patients to recommend known treatment plans, thus alleviating the problem of under-representation of rare drugs in the dataset. This approach is expected to improve the recommendation accuracy and safety of all drugs, better serving clinical practice and patient treatment.

6 Conclusion

This study proposes a novel drug recommendation framework, SDRec, which leverages a medical domain knowledge-based synergistic drug recommendation network. By constructing a drug synergy module, we can uncover the interactions between drug molecules, thereby identifying effective drug combinations tailored to patients. In addition, we introduce additional medical domain knowledge to construct the synergistic relationships between drugs as a drug synergy graph and use GCN modeling to obtain a representation of the relationship between drugs. This is also the first time that the concept of drug synergy has been introduced to this field, providing a new perspective and approach. Comparative experiments and ablation experiments are carried out on the public MIMIC dataset, and the experimental results show the superior performance of the SDRec framework compared to existing drug recommendation methods. In future research, we will focus on drug-drug interactions, including side effects, synergism, and antagonism, in order to clarify the mechanism of drug-drug interactions and improve the safety of drug recommendations.

Acknowledgment

This work was supported by the Natural Science Foundation of China (No. 62072070).

References

[1] S. Liu, X. Wang, Y. Xiang, H. Xu, H. Wang, and B. Tang, Multi-channel fusion LSTM for medical event prediction using EHRs, *J. Biomed. Inform.*, vol. 127, p. 104011, 2022.

[2] P. Symeonidis, S. Chairistanidis, and M. Zanker, Recommending what drug to prescribe next for accurate and explainable medical decisions, in *Proc. 34th Int. Symp. Computer-Based Medical Systems* (*CBMS*), Aveiro, Portugal, 2021, pp. 213–218.

[3] H. Lakkaraju and C. Rudin, Learning cost-effective and interpretable treatment regimes, in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics*, Fort Lauderdale, Florida, USA, 2017, pp. 166–175.

[4] M. Adnan, J. Warren, and M. Orr, Ontology based semantic recommendations for discharge summary medication information for patients, in *Proc. 23rd Int. Symp. Computer-Based Medical Systems* (*CBMS*), Bentley, WA, Australia, 2010, pp. 456–461.

[5] F. Gong, M. Wang, H. F. Wang, S. Wang, and M. Y. Liu, SMR: Medical knowledge graph embedding for safe medicine recommendation, *Big Data Res*., vol. 23, p. 100174, 2021.

[6] Y. Zhang, R. Chen, J. Tang, W. F. Stewart, and J. Sun, LEAP: Learning to prescribe effective and safe treatment combinations for multimorbidity, in *Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Halifax, NS, Canada, 2017, pp. 1315–1324.

[7] Q. Chen, X. Li, K. Geng, and M. Wang, Context-aware safe medication recommendations with molecular graph and DDI graph embedding, in *Proc. 37th AAAI Conf. Artificial Intelligence*, Washington, DC, USA, 2023, pp. 7053-7060.

[8] J. Shang, C. Xiao, T. Ma, H. Li, and J. Sun, GAMENet: Graph augmented MEmory networks for recommending medication combination, in *Proc. 33rd AAAI Conf. Artificial Intelligence*, Honolulu, Hawaii, USA, 2019, pp. 1126–1133.

[9] X. Li, Y. Zhang, X. Li, H. Wei, and M. Lu, DGCL: Distance-wise and graph contrastive learning for medication recommendation, *J. Biomed. Inform*., vol. 139, p. 104301, 2023.

[10] N. Yang, K. Zeng, Q. Wu, and J. Yan, MoleRec: Combinatorial drug recommendation with substructure-aware molecular representation learning, in *Proc. ACM Web Conf. 2023*, Austin, TX, USA, 2023, pp. 4075–4085.

[11] C. Yang, C. Xiao, F. Ma, L. Glass, and J. Sun, SafeDrug: Dual molecular graph encoders for recommending effective and safe drug combinations, arXiv preprint arXiv: 2105.02711, 2021.

[12] R. Wu, Z. Qiu, J. Jiang, G. Qi, and X. Wu, Conditional generation net for medication recommendation, in *Proc. ACM Web Conf. 2022*, Virtual Event, 2022, pp. 935–945.

[13] T. H. Li, C. C. Wang, L. Zhang, and X. Chen, SNRMPACDC: Computational model focused on Siamese network and random matrix projection for anticancer synergistic drug combination prediction, *Brief. Bioinform*.,

vol. 24, no. 1, p. bbac503, 2023.
[14] T. Wang, R. Wang, and L. Wei, AttenSyn: An attention-based deep graph neural network for anticancer synergistic drug combination prediction, *J. Chem. Inf. Model.*, vol. 64, no. 7, pp. 2854–2862, 2024.
[15] T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, in *Proc. 5th Int. Conf. Learning Representations*, 2017. https://openreview.net/forum?id=SJU4ayYgl.
[16] C. Yang, C. Xiao, L. Glass, and J. Sun, Change matters: Medication change prediction with recurrent residual networks, arXiv preprint arXiv: 2105.01876, 2021.
[17] S. Liu, X. Wang, X. Zhao, and H. Chen, DKINet: Medication recommendation via domain knowledge informed deep learning, in *Proc. 2024 IEEE Int. Conf. Bioinformatics and Biomedicine* (*BIBM*), 2024, pp. 3521–3526.
[18] J. Mi, Y. Zu, Z. Wang, and J. He, ACDNet: Attention-guided collaborative decision network for effective medication recommendation, *J. Biomed. Inform*., vol. 149, p. 104570, 2024.
[19] C. Merkwirth and T. Lengauer, Automatic generation of complementary descriptors with molecular graph networks, *J. Chem. Inf. Model.*, vol. 45, no. 5, pp. 1159–1168, 2005.
[20] X. Chen, X. E. Liu, and J. Wu, Research progress on drug representation learning, *J. Tsinghua Univ.* (*Sci. Technol.*), vol. 60, no. 2, pp. 171–180, 2020.
[21] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Comput. Sci.*, vol. 28, no. 1, pp. 31–36, 1988.
[22] W. Zaremba, I. Sutskever, and O. Vinyals, Recurrent neural network regularization, arXiv preprint arXiv: 1409.2329, 2014.
[23] D. P. Kingma and M. Welling, Auto-encoding variational bayes, in *Proc. 2nd Int. Conf. Learning Representations*, 2014. https://dblp.org/db/conf/iclr/iclr2014.html#Kingma-W13.
[24] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, The graph neural network model, *IEEE Trans. Neural Network.*, vol. 20, no. 1, pp. 61–80, 2009.
[25] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv: 1412.3555, 2014.
[26] J. Degen, C. Wegscheid-Gerlach, A. Zaliani, and M. Rarey, On the art of compiling and using 'drug-like' chemical fragment spaces, *ChemMedChem*, vol. 3, no. 10, pp. 1503–1507, 2008.
[27] J. Lee, Y. Lee, J. Kim, A. Kosiorek, S. Choi, and Y. W. Teh, Set transformer: A framework for attention-based permutation-invariant neural networks, in *Proc. 36th Int. Conf. Machine Learning*, Long Beach, California, USA, 2019, pp. 3744–3753.
[28] G. L. Xiong, Z. J. Yang, J. C. Yi, N. N. Wang, L. Wang, H. M. Zhu, C. K. Wu, A. P. Lu, X. Chen, S. Liu, et al., DDInter: An online drug–drug interaction database towards improving clinical decision-making and patient safety, *Nucleic Acids Res*., vol. 50, no. D1, pp. D1200-D1207, 2022.
[29] R. Hadsell, S. Chopra, and Y. LeCun, Dimensionality reduction by learning an invariant mapping, in *Proc. 2006 IEEE Computer Society Conf. Computer Vision and Pattern Recognition* (*CVPR'06*), New York, NY, USA, 2006, pp. 1735–1742.
[30] A. E. W. Johnson, T. J. Pollard, L. Shen, L. W. H. Lehman, M. L. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, MIMIC-III, a freely accessible critical care database, *Sci. Data*, vol. 3, no. 1, p. 160035, 2016.
[31] A. E. W. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. C. Hao, B. Moody, B. Gow, et al., MIMIC-IV, a freely accessible electronic health record dataset, *Sci. Data*, vol. 10, no. 1, p. 1, 2023.

Kang An received the BS degree from Qingdao University, China, in 2022. He is currently a graduate student at School of Information Science and Technology, Dalian Maritime University, China. His research interests focus on drug recommendations.

Mingyu Lu received the BS degree from Heilongjiang University, China, in 1985 and the MS and PhD degrees from Tsinghua University, China, in 1988 and 2002, respectively. He is currently a professor at School of Artificial Intelligence, Dalian Maritime University, China. He has published nearly 200 academic papers in the fields of artificial intelligence, machine learning, and data mining.

Fei Chen received the BS, MS, and PhD degrees from Dalian Maritime University, China, in 2002, 2005, and 2011, respectively. He is a lecturer at School of Information Science and Technology, Dalian Maritime University, China. He has published several research papers on topics in text mining

Xiaoxia Liu received BSc degree from Dalian Jiaotong University, China, in 2011, and Ph.D. degree from Dalian University of Technology, China, in 2018. She is currently a postdoctoral researcher at Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA, USA. Her main research interests include bioinformatics and text mining.

Yijia Zhang received the MS degree in 2009 and the PhD degree in 2014 from Dalian University of Technology, China. He is a professor at School of Information Science and Technology, Dalian Maritime University, China. His main research interests include natural language processing and bioinformatics.