Drug Combination Recommendation Based on Multi-View Drug Feature Learning

Yanjie Zhao[1], Xiaomei Yu[1], Jianlong Zhao[2], Shucheng Liu[3], Wei Zhang[3]
{tracer_jj@163.com, yxm0708@126.com, zhaojianlong@hisense.com
liushucheng@shandong.cn, zhangwei@shandong.cn }
[1] School of Information Science and Engineering, Shandong Normal University, 250358, Jinan, China
[2] Research and Development Center, Qingdao Hisense TransTech Co., Ltd., Qingdao, China
[3] Shandong Provincial Motor Vehicle Exhaust Pollution Monitoring Center, 250102, Jinan, China

Abstract. With the booming development of artificial intelligence technology and the medical industry, deep learning-based drug recommendation models have been playing an increasingly crucial role in the healthcare community. However, the existing approaches suffer from inadequate drug representations and the lack of multi-view drug information. To address these challenges, this paper proposes a novel model named Graph-based Multi-View Learning for Drug Recommendation (MVRM). Specifically, the semantic information of drugs are learned from electronic health records (EHRs) and drug molecular structures are extracted. Moreover, functionally differentiated drug molecular representations are leveraged to distinguish consistent 2D structures from distinct 3D geometric conformations. By incorporating both semantic information and structural relationships, comprehensive drug representations are generated and adverse drug-drug interactions are avoided. Furthermore, a cross-view contrastive learning mechanism is deployed to fully explore the EHR data from multiple views, thereby achieving safe and effective drug recommendations. Finally, the experimental results on EHR datasets demonstrate the outstanding performance of MVRM, compared with the state-of-the-art baseline models in several metrics.

Keywords: Medication recommendation, electronic health records, multi-view learning, contrastive learning

1 Introduction

With the continuous growth of EHRs data and the rapid advancement of artificial intelligence technologies, deep learning-based drug recommendation[1] has attracted widespread attention in the medical field[2].

Drug recommendation aims to provide patients with personalized drug combinations based on their medical visit history[3] and pharmaceutical knowledge[4], and the existing methods include instance-based approaches and longitudinal ones[5, 6, 7]. The former focus on recommending drugs

IIKI 2025, December 18-21, Chengdu, People's Republic of China

DOI 10.4108/eai.18-12-2025.2365281

according to the patient's current health status, while the latter[8] takes into account the progression of patients' diseases, offering personalized medication prescription suggestion[9] based on the temporal dependencies in longitudinal EHR data. Moreover, existing studies have shown that the biochemical activity of a drug is usually associated with a small number of conserved molecular substructures[10]. Therefore, an increasing number of researches persist to incorporate drug molecular structure and improve the accuracy of drug recommendation[11, 12, 13].

However, the relationships between drugs are intricate[14], and different drugs have distinct chemical properties with complex and variable molecular structures. Therefore, learning effective drug molecular representations and integrating multi-structural drug information to obtain comprehensive drug representations remains a challenging task[15].

To address the above challenges, this paper proposes a drug recommendation method based on multi-view learning, which integrates the intrinsic chemical features of drugs, drug interaction relationships in EHRs, and external drug knowledge to obtain more accurate and effective drug representations. This enables better drug recommendations for downstream tasks, offering more personalized, accurate, and interpretable drug treatment combinations for patients. Specifically, our contributions are as follows:

- We extract the dependency relationships between drugs based on their co-occurrence frequency in different EHR visit records and associate these dependencies with drug molecular representations. This helps differentiate between drug molecules with the same molecular structure but different functions.

- We introduced drug substructure information into molecular representation learning, employing a graph isomorphism network (GIN), we extract drug molecular structure and substructure features, and design a gated attention mechanism to dynamically combine molecular structure and substructure information. This allows adaptive learning of drug molecular representations with substructure features.

- We propose a cross-view contrastive learning method, which systematically integrates the intrinsic chemical features of drugs, drug interaction relationships in EHR, and external drug DDI knowledge through an contrastive learning mechanism. This enhances the comprehensiveness and reliability of drug representations, supporting more accurate drug recommendations.

We conduct extensive experiments to evaluate the performance of proposed MVRM on real-world datasets. The results demonstrate that our model exhibits superior performance across widely-adopted evaluation metrics compared to baseline models.

2 Problem formulation

2.1 Electronic Health Records

EHRs also known as electronic medical records, document the patient's medical history (e.g., medications, diagnoses, procedures). We represent the electronic medical records of patient v as

$$\mathbf{X_v} = \left\{x^1, x^2, \cdots, x^t\right\} \tag{1}$$

where t represents the total number of visits for patient v. The representation of the patient's t-th visit is defined as:

$$x^i = \left[d^i, p^i, m^i\right] \tag{2}$$

where $d^i \in \{0,1\}^D, p^i \in \{0,1\}^P, m^i \in \{0,1\}^M$ represent different visit information, the sets M, D, and P are the corresponding medical code sets.

2.2 Problem description

In drug recommendation systems, predicting the current recommended drug combination M^t is a crucial step in achieving personalized healthcare. However, DDIs can lead to serious health risks, necessitating the effective reduction of these adverse reactions. This study aims to train three key models: D^t, P^t, and M^{t-1}, to enhance the prediction accuracy of drug combinations while simultaneously reducing DDI rate.

3 Methods

3.1 Overview of MVRM

The framework of our proposed MVRM model is shown in Fig.1, which is divided into four modules: patient representation, drug-drug interaction learning, drug molecular structure learning and cross-view contrastive learning.

3.2 Patient representation

The health of patients is represented by diagnostic and procedure information. Given the diagnosis and procedure codes in the visit records in the initial EHR data, we coded them. We define two embedding tables, $E_d \in \mathbb{R}^{|D| \times l}$, $E_p \in \mathbb{R}^{|P| \times l}$, where each row of E_d and E_p species stores an embedding vector for diagnosis and procedure respectively, and dim represents the embedding dimension. Given the multiple hot vector triplet d^t, p^t, m^t, The embedding matrices E_d, E_p are multiplied by the original multi heat vectors d^t, p^t respectively to encode the current health status of the patient, as follows:

$$d_e^t = d^t E_d, \quad p_e^t = p^t E_p \tag{3}$$

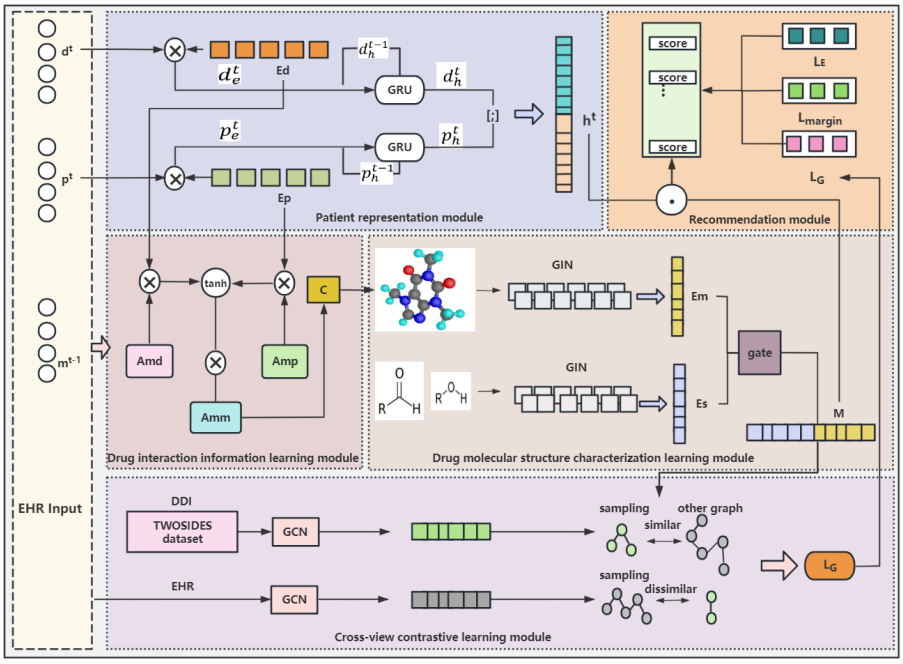

Fig. 1. The framework of MVRM

However, a single visit is not enough to determine a personalized treatment plan. For example, when a patient with a long-term chronic disease and an emergency patient change for the same pneumonia disease, it is inappropriate for us to prescribe the same drug to the patient. Therefore, we need to refer to the patient's death for diagnosis. We use two gating circulation units to deal with the problem of long-term history dependence, so as to capture the longitudinal information of patients:

$$d_h^t = BiGRU_d(d_e^t, d_h^{t-1}) \tag{4}$$

$$p_h^t = BiGRU_p(p_e^t, p_h^{t-1}) \tag{5}$$

where $d_h^t, p_h^t \in \mathbb{R}^l$. The representation of the patient is as follows:

$$h^t = W_h[d_h^t; p_h^t] \tag{6}$$

Finally, we get the final representation of the patient $h^t \in \mathbb{R}^l$, which contains all the diagnostic and surgical information, where $[;]$ is the concatenation operator and W_h is a learnable weight matrix in $\mathbb{R}^{l \times 2l}$.

3.3 Drug-drug interaction learning

The correlations between medications in EHRs facilitate a more comprehensive learning of drug representations. While molecular-based drug representation methods rely on graph representa-

tions of molecules, 2D molecular structures only capture atomic connectivity and overlook 3D spatial configurations. Such structural ambiguity diminishes the accuracy of drug recommendations. Therefore, leveraging the differential co-occurrence patterns of distinct drugs in EHRs to resolve these discrepancies becomes critically important.

Specifically, for the EHR data, we construct the drug-diagnosis co-occurrence matrix $A_{md} \in \mathbb{R}^{|\mathscr{M}|\times|\mathscr{D}|}$,, drug-procedure co-occurrence matrix $A_{mp} \in \mathbb{R}^{|\mathscr{M}|\times|P|}$, and drug-drug co-occurrence matrix $A_{mm} \in \mathbb{R}^{|\mathscr{M}|\times|\mathscr{M}|}$ by statistically analyzing the co-occurrence frequency between different medical codes. We normalize each row of these matrices using the L1 norm.

We define the initial drug context representations in the two views as $C_d = A_{md}E_d$, and $C_p = A_{mp}E_p$, where $C_d, C_p \in \mathbb{R}^{|\mathscr{M}|\times l}$. Then we combine the diagnosis and program views into co-occurrence information C_{dp} by $C_{dp} = [C_d; C_p]W_c$, where $W_c \in \mathbb{R}^{2l\times l}$ is a learnable parameter matrix.Next, we obtain the combined information by $C_{mm} = A_{mm}C_{dp}$, $C = C_{dp} + tanh(C_{dp}W_{s1}) \odot C_{mm}$ uses the feature attention layer to integrate the combination information and the co-occurrence information C_{dp}, in which the feature attention layer uses the activation function $\tanh(\cdot)$ to take $W_{s1} \in \mathbb{R}^{l\times l}$ and realize, so as to adaptively select the valuable features in C_{mm}, and filter out the trivial features according to C_{dp} to represent the element product. Finally, we get the drug interaction information C.

3.4 Drug molecular structure learning

The efficacy and activity of a drug mainly depend on its molecular structure and properties, which makes molecular graph-based learning an important research topic.To obtain a comprehensive and accurate molecular representation, we represent the molecular graph of a drug as $\mathscr{G}_m = (\mathscr{V}_m, \mathscr{E}_m)$, where $\mathscr{V}_m$ And $\mathscr{E}_m$ represents the set of atom sets and edges, respectively. Chemical bonds are modeled as edges. Given a molecule G, we employ GIN[16] to encode both the structural information and substructure information of the drug molecule. Subsequently, a readout function is applied to aggregate and combine all node features.The node update formula for each atom $v \in \mathscr{V}_m$ is as follows:

$$\mathbf{a}^{(k)}(u) = \mathrm{AGG}^{(k)}\left(\left\{\mathbf{h}_n^{(k-1)}(v), \mathbf{h}_e(u,v) | v \in \mathscr{N}(u)\right\}\right) \tag{7}$$

$$\mathbf{h}_n^{(k)}(u) = \mathrm{MLP}^{(k)}\left(\mathbf{a}^{(k)}(u) + \left(1 + \varepsilon^{(k)}\right) \cdot \mathbf{h}_n^{(k-1)}(u)\right) \tag{8}$$

where k represents the number of layers of GNN, $\mathbf{a}^{(k)}(u)$ is the representation of node u at the k-th layer. $\mathscr{N}(u)$ represents the set of nodes directly connected to node u, $\mathbf{h}_n^{(k)}(u)$ is the final representation of node u. The $\varepsilon^{(k)}$ is a learnable hyperparameter related to the number of layers, which is used to control the influence degree of the upper layer node feature $\mathbf{h}_n^{(k-1)}(u)$ in the node update process.

To incorporate drug interaction information C, we encode the neighborhood information along with the C as follows:

$$\hat{\mathbf{h}}_n^{(k)}(u) = \tanh\left(W_{s2}C^m\right) \odot \mathbf{h}_n^{(k)}(u) \tag{9}$$

where W_{s2} denotes a learnable linear transformation matrix, $\tanh(\cdot)$ is an activation function, $\odot$ represents the element-wise multiplication, and $\hat{h}_n^{(k)}(u)$ stands for the drug molecular feature representation of the K-th layer after fusing the interaction information C.

Subsequently, the readout function is used to aggregate the node features from the L-th layer, the formula is as follows:

$$e_m = READOUT\left(\left\{\hat{\mathbf{h}}_n^{(k)}(u)|u \in \mathscr{V}\right\}\right). \tag{10}$$

where e_m is the final representation of the m_{th} drug. Each drug is encoded by the GIN and stored in the drug embedding matrix E_m, where $E_m \in \mathbb{R}^{|\mathscr{M}|\times l}$.

Similarly, the substructural information of the drug molecule is also obtained through GIN, resulting in $E_s \in \mathbb{R}^{|\mathscr{M}|\times l}$.

Finally, we design an attention-based gating mechanism to dynamically weight the contributions of the drug molecular structure representation E_m and the drug substructure representation E_s, to obtain the drug molecular structure representation with substructural information M. The gated attention mechanism is defined as follows:

$$\mathbf{gate} = softmax(\mathbf{W}_2\ ReLU(\mathbf{W}_1\ [E_m; E_s\ ] + b_1) + b_2 \tag{11}$$

$$\mathbf{M} = gate_1 \cdot \mathbf{E}_m + gate_2 \cdot \mathbf{E}_s \tag{12}$$

where $\mathbf{W}_1 \in \mathbb{R}^{d\times 2d}$, $\mathbf{b}_1 \in \mathbb{R}^{d\times 2d}$ are the parameters of the first linear layer, and $\mathbf{W}_2 \in \mathbb{R}^{d\times 2d}$, $\mathbf{b}_2 \in \mathbb{R}^{d\times 2d}$ are the parameters of the second linear layer.

3.5 Cross-view contrastive learning

To enforce the similarity of representations of the same drug across different views and alleviate the data sparsity problem in single-view data, we design a cross-view contrastive learning module. First, we convert the EHR adjacency matrix and the external knowledge DDI adjacency matrix from the TWOSIDES dataset into graph structures. Graph convolutional networks (GCNs) [17] are then used to process graph structures and extract aggregated features from nodes and edges:

$$GCN(X,A) = \sigma\left(D^{-\frac{1}{2}}\widetilde{A}D^{-\frac{1}{2}}\right) \tag{13}$$

where $X \in \mathbb{R}^{M\times d}$ represents the number of nodes and d represents the characteristic dimension of the nodes. A represents the adjacency matrix of the edges, $\widetilde{A} = A + I$, where I is the identity matrix, D is the degree matrix of A and W is the learnable parameter. We used GCN to model the EHR and DDI graphs, the formula is as follows:

$$G_e = GCN\left(RELU\left(GCN\left(X_e, A_e\right)W_e\right), A_e\right) \tag{14}$$

$$G_d = GCN\left(RELU\left(GCN(X_e, A_e)W_d\right), A_d\right) \tag{15}$$

Contrastive learning aims to bring similar sample pairs closer while pushing dissimilar samples as far apart as possible. The cross-view contrastive learning module learns the multi-view

differences among EHR knowledge, external DDI knowledge, and drug molecular structure knowledge. By bringing positive samples closer and pushing negative samples apart, it fully integrates multi-view knowledge to enhance the safety of recommended drug combinations. We adopt the Euclidean distance (denoted as D) to measure the distance between multi-view drug knowledge, thereby capturing the differences among them. The distance measurement formula is given by Eq. (16):

$$D = \|G_m + G_e - G_d\|_2 \tag{16}$$

where $\|\cdot\|_2$ represents the L2 norm, also called the Euclidean norm. The L2 norm is used to compute the magnitude or length of a vector.

Subsequently, the distance contrastive loss function is adopted to dynamically learn multi-view drug features to the maximum extent. For positive and negative samples, we expect a margin (denoted as m) between their embeddings to regulate the distance of negative samples. In the experiments, we set m to 2.0, where this parameter represents the minimum distance required between negative samples. When the distance of negative samples is less than $m(D < m)$, the model forces the distance of negative sample pairs to be at least m; otherwise, the gradient of the negative sample term is zero (and optimization ceases). The distance contrastive loss function is given by Eq. (17):

$$L_G = Y\frac{1}{2}(D)^2 + (1-Y)\frac{1}{2}max(0, m-D)^2 \tag{17}$$

If the recommended drug combination has a DDI, then Y is set to 0, and the model adjusts the distance between negative samples via the graph margin (m) in the loss function. Conversely, Y is set to 1, and the positive samples are pulled closer using the first part of the loss function.

3.6 Training and inference

3.6.1 Multi-Label Prediction Loss

The matching score of each drug is the weighted sum of the similarity between h^t and M expressed by the patient. The calculation formula is as follows:

$$\hat{y} = 1/2[\sigma\left(LN\left(\widetilde{E}_m\widetilde{h}^t\right)\right) + \sigma\left(LN\left(\widetilde{M}\widetilde{h}^t\right)\right)] \tag{18}$$

where $\hat{y} \in \mathbb{R}^{|\mathscr{M}|}$, σ represents the sigmoid activation function, and LN represents the layer normalization operation. $\widetilde{E}_m$ and $\widetilde{M}$ are row normalized E_m and M, and $\widetilde{h}^t$ is normalized h^t.

3.6.2 Objective function

In this paper, the drug recommendation problem is defined as a multi class and multi label classification task. First, we take the binary cross entropy (BCE) loss as a part of the goal, and empirically use the multi label hinge loss L_{margin}, aiming to maintain a significant gap between the real label score and other label scores, and at the same time, we add the contrast loss L_G to control

the antagonism between drugs. Therefore, the objective function is the weighted sum of L_E and L_{margin} and L_G, the calculation formula are as follows::

$$L_E = -\sum_{i=1}^{|\mathscr{M}|} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i) \tag{19}$$

$$L_{margin} = -\sum_{i:y(i)=1}\sum_{j:y(j)=0} \frac{max(0, 1-(\widehat{y_i}-\widehat{y_j}))}{|\mathscr{M}|} \tag{20}$$

$$Loss = (1-\alpha)L_E + \alpha L_{margin} + \beta L_G \tag{21}$$

The training algorithm is detailed in Algorithm 1.

Algorithm 1 Training algorithm for MVRM

Require: Training set Q, the iterations T_{max} , trainable parameters, embedded margin m in Eq.(16), learning rate η;
Input: Preprocessed EHR data, EHR adjacency matrix and the external DDI knowledge;
1: Initialize parameters:
2: Calculate adjacency matrix;
3: **for** $T = 1 \rightarrow T_{max}$ **do**
4: Sample a patient $\mathbf{X_v} = \{x^1, x^2, \cdots, x^t\}$ from EHR;
5: Obtain external knowledge and records from Eq. (13), (14),
5: (15);
6: **for** a visit i=1 to X_v **do**
7: Fetch the i^{th} visit of patient X_v;
8: Obtain the current health status of the patient d_e^t, p_e^t from
9: Eq. (3);
10: Obtain the final representation of the patient h^t from Eq.
11: (4), (5), (6);
12: Based on Eq. (7), (8), (9), (10), (11), (12) to obtain the drug
13: molecular structure representation with substructural in-
14: formation M;
15: Calculate contrastive loss L_G based on Eq. (16), (17);
16: Accumulate L_E, and L_{margin} in Eq. (18), (19), (20);
17: Optimizing the loss in Eq. (21);
18: **return** medicine recommendations with the trained model;
19: **end**

4 Experiments and analysis

4.1 Datasets

The EHR data used in the experiments are from the publicly available MIMIC-III[18] and MIMIC-IV[19]. After data processing, the final dataset for MIMIC-III includes 6,350 patients, and the MIMIC-IV dataset includes 9,036 patients.

4.2 Experimental settings

The model is built in Pytorch, with torch version 1.8.0+cu111 and torch-geometric version 2.0.1. The learning rate is set to 0.00015, with two graph convolution layers. The contrastive learning soft matching graph margin is set to 2.0 (m=2.0), the number of epochs is set to 51, and training is conducted on an NVIDIA GeForce RTX 3090 GPU.

4.3 BaseLines

We compared the performance of various methods on the MIMIC-III and MIMIC-IV datasets. The experimental results are presented in Table 1,Table 2 and Fig. 2. The experiments show that MVRM outperforms other baseline methods across multiple performance metrics. The baseline methods are shown below:

- **LEAP**[20] combines deep learning and reinforcement learning techniques to optimize the process of drug discovery. It can quickly identify potential drug candidates among a large number of compounds by dynamically learning the patterns of drug target interactions.
- **GAMENet**[21] processes the graph structure information of drug molecules through graph neural networks, effectively capturing the complex interrelationships between molecules. This has improved the accuracy and specificity of predictions.
- **SafeDrug**[12] focuses on drug safety by building risk assessment models to identify potential side effects in advance and reduce risks in drug development.
- **COGNet**[22] enhances drug discovery through the use of knowledge graphs, integrating biomedical knowledge and data to improve the interpretability of drug target associations.
- **DAINet**[13] improves drug recommendation accuracy by modeling drug relationships through integrating molecular structural features and disease association information.
- **MoleRec**[11] delves into the importance of specific molecular substructures in medications. This approach enhances the precision of medication recommendations by leveraging finer molecular representations.

Table 1, Table 2 and Fig. 2 present the drug recommendation results, showing that both MVRM and its variants outperform the baseline models in terms of performance.

Table 1: Performance comparison on MIMIC-III dataset

Methods	DDI_RATE	F1	Jaccard	PRAUC
LEAP	0.0731±0.0008	0.6138±0.0026	0.4521±0.0024	0.6549±0.0033
GAMENet	0.0864±0.0006	0.6626±0.0025	0.5067±0.0025	0.7631±0.0030
SafeDrug	0.0589±0.0005	0.6768±0.0027	0.5213±0.0030	0.7647±0.0025
COGNET	0.0852±0.0005	0.6869±0.0010	0.5336±0.0011	0.7739±0.0009
DAINet	0.0782±0.0009	0.6804±0.0027	0.5252±0.0024	0.7738±0.0023
MoleRec	0.0724±0.0008	0.6813±0.0029	0.5273±0.0033	0.7736±0.0027
MVRM	**0.0533±0.0015**	**0.6832±0.0002**	**0.5290±0.0030**	**0.7800±0.0011**
MVRM W/OA	0.0535±0.0010	0.6818±0.0012	0.5278±0.0030	0.7762±0.0031
MVRM W/OB	0.0538±0.0020	0.6830±0.0006	0.5283±0.0030	0.7774±0.0021
MVRM W/OC	0.0534±0.0015	0.6827±0.0004	0.5288±0.0030	0.7787±0.0011
MVRM W/OC+D	0.0541±0.0009	0.6829±0.0002	0.5284±0.0030	0.7765±0.0023

Table 2: Performance comparison on MIMIC-IV dataset

Methods	DDI_RATE	F1	Jaccard	PRAUC
LEAP	0.0655±0.0006	0.6012±0.0013	0.4405±0.0024	0.4405±0.0024
GAMENet	0.0734±0.0006	0.6363±0.0014	0.4784±0.0025	0.7395±0.0030
SafeDrug	0.0571±0.0020	0.6435±0.0027	0.4857±0.0030	0.7400±0.0035
COGNET	0.0799±0.0010	0.6541±0.0010	0.5060±0.0011	0.7135±0.0025
DAINet	0.0779±0.0008	0.6558±0.0012	0.5002±0.0016	0.7457±0.0029
MoleRec	0.0713±0.0005	0.6560±0.0029	0.5037±0.0033	0.7500±0.0015
MVRM	**0.0535±0.0015**	**0.6595±0.0012**	**0.5053±0.0010**	**0.7535±0.0011**

First, instance-based methods such as LEAP achieve poor performance because they only focus on data from the current visit and neglect historical information. Longitudinal methods like GAMENet and COGNet perform well by leveraging the similarity of historical visits for medication recommendation. DAINet, SafeDrug, and MoleRec incorporate drug molecular structure information, yielding excellent performance. However, SafeDrug merely appends substructural features as supplementary information to 2D molecular structures through simple concatenation, failing to construct a substructure-aware dual-branch collaborative modeling mechanism and overlooking the impact of subtle differences in molecular substructures on drug efficacy. MoleRec fails to effectively integrate drug-drug interaction information from EHRs into substructure-aware molecular representations, making it difficult to accurately avoid potential risks associated with isomorphic drugs and resulting in recommended medications lacking clinical adaptability. The architectural advantage of MVRM lies in infusing drug-drug interaction information from EHRs into molecular structure encodings while incorporating drug substructural information. This ensures the model fully utilizes EHR-derived drug interaction insights and accounts for the influence of functional groups on drug functions. Additionally, MVRM adopts contrastive learning to perform cross-view fusion of multi-view medication knowledge, enabling it to not only learn discriminative drug representations but also maintain a low DDI rate for recommended medications. Overall, MVRM demonstrates outstanding performance in both accuracy and safety across comparative experiments.

Table 3 evaluates the model complexity of MVRM in comparison with other deep learning-

Table 3: Parameters and training time on MIMIC-III and MIMIC-IV

Methods	Param. in MIMIC-III/IV	Training(s) in MIMIC-III/IV
LEAP	177395/175622	503.20/774.41
GAMENet	455002/669444	147.25/169.36
SafeDrug	366122/609150/	141.35/195.72
COGNET	1357624/1578172	129.31/149.49
DAINet	372842/618467	196.81/361.18
MoleRec	507060/727412	1077.30/1387.10
MVRM	**805455/1025807**	**186.27/267.45**

based baseline methods. Although our model has a larger number of parameters compared to LEAP, DAINet, and MoleRec, it exhibits relatively lower time complexity, demonstrating advantages in efficiency and flexibility.

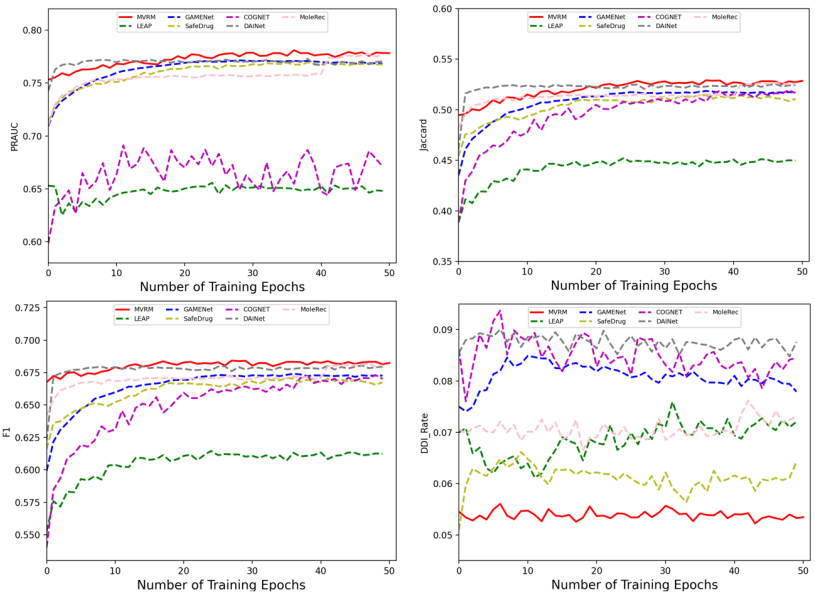

Fig. 2. Performance Evaluation on MIMIC-III Dataset

4.4 Ablation experiments

In this section, we develop four variant versions of the model to investigate the effectiveness of each component. We conduct experimental comparisons between these variants and our proposed model. The final experimental results of MVRM and its variants are presented in Table 1 and Fig. 3. The detailed descriptions of these variants are as follows:

- **MVRM W/OA** the MVRM W/OA remove the drug substructure information extraction of the drug molecular structure characterization learning modul. The rest of the architecture is retained.
- **MVRM W/OB** the MVRM W/OA eliminate the drug interaction information extraction module. The rest of the architecture is retained.
- **MVRM W/OC** the MVRM W/OC eliminate the gate attention fusion mechanism. The rest of the architecture is retained.
- **MVRM W/OC+D** the MVRM W/OC+D remove the cross-graph contrastive learning and gate attention fusion mechanism. The rest of the architecture is retained.

To ensure a fair performance evaluation, we employed identical parameters across all variant models as well as the original MVRM. The training process is briefly illustrated in Fig. 3, which demonstrate consistent performance improvement in MVRM and its variants as the number of training epochs increases.

Firstly, a comparison between MVRM W/OA and MVRM shows that incorporating drug substructure information leads to richer drug representations. Second, the variant MVRM W/OB verifies that extracting relational information between drugs in medical records helps distinguish the geometric structures of drug molecules. The variant MVRM W/OC demonstrates that the gated attention mechanism adaptively selects chemically salient information from both molecular structures and substructures that are more critical to drug efficacy, resulting in clearer drug representations from this perspective. Finally, experimental results of the variant MVRM W/OC+D indicate that leveraging cross-view contrastive learning effectively integrates multi-view drug information, simultaneously bringing safer drug combinations closer while pushing high-risk combinations apart, thereby achieving the goal of controlling the DDI rate of recommended medications.

4.5 Parameter sensitivity analysis

In this section, we investigate the Effect of number of GCN convolutional layers and the weight distribution of gated attention fusion on model performance.

4.5.1 Effect of number of GCN convolutional layers

This study investigates the selection of the number of layers in GCN during node propagation. Fig. 4 presents the variation in evaluation metrics as the number of convolution layers increases within the range of $\{1,2,3,4,5\}$. The experimental results indicate that, for the current Medication Recommendation task, two convolution layers offer the best overall performance, as reflected in the metrics across various evaluation criteria.

4.5.2 Weight distribution of gate attention fusion

The gating mechanism adaptively models feature importance by assigning weighting coefficients to drug molecular and substructure vectors. Analysis of the gated weight heatmap (Fig. 5)

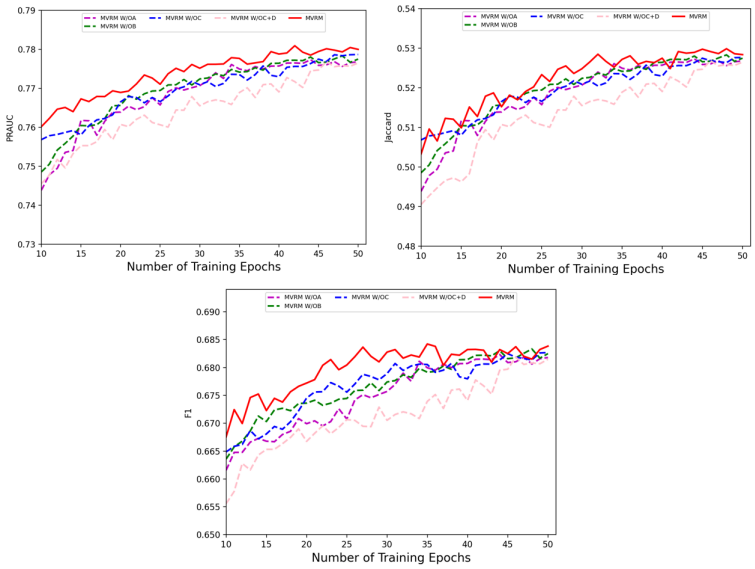

Fig. 3. Performance Evaluation on MIMIC-III Dataset

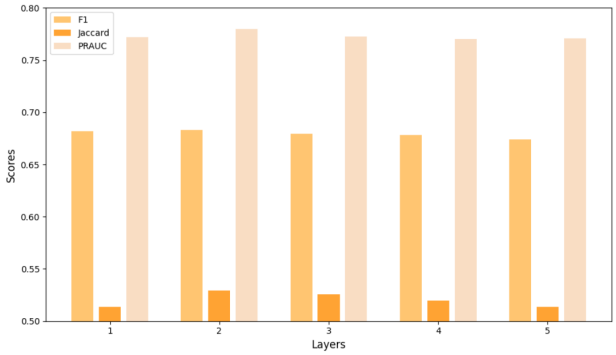

Fig. 4. GCN Layers

shows that weight pattern changes intuitively reflect the model's learning dynamics and weight distribution across different training samples and phases. Experiments confirm the model adaptively adjusts weights based on feature importance, enabling efficient weighted fusion.

4.6 Case study

To validate the effectiveness, safety, and innovation of the proposed MVRM model in clinical medication recommendation, as shown in Fig. 6, we conducted a detailed case study on a patient with two hospital visits from the EHR dataset, where diagnosis codes and medication codes were

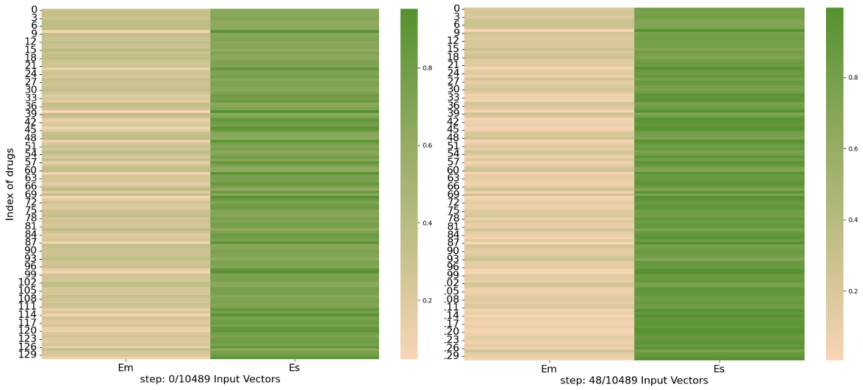

Fig. 5. Weight Distribution of Gate Attention Fusion

represented by ICD-9 and ATC-3 codes, respectively. The patient's core diagnoses included intracerebral hemorrhage (ICD-9: 431), hypertension (ICD-9: 4019), pneumonia (ICD-9: 486), and cardiac arrest (ICD-9: 4275), with a craniotomy (ICD-9: 159) performed during the first visit. In the medication list, we can observe that the model successfully recommended most of the core medications across both hospital visits, such as sodium nitroprusside (C02DB05), mannitol (B05CA01). Additionally, the recommended pantoprazole (A02BC02) exhibits no DDI when used in combination with emergency medications, demonstrating that the model integrates multi-view medication knowledge for reasoning. Furthermore, cefazolin and cephalothin share similar structures and identical antibacterial core functions. However, the MVRM model effectively captures the subtle differences in their functional groups. By incorporating the clinical context of the patient, it accurately recommends cefazolin—characterized by low nephrotoxicity and high applicability—while excluding high-risk cephalothin. This validates the model's effectiveness in discriminative learning for molecular structure representation of medications. Relatively speaking, the model exhibits a certain degree of incorrect and missing recommendations. Nevertheless, overall, it can generate medication regimens that are highly consistent with the patient's clinical needs.

5 Conclusion

In this paper, we propose a multi-view learning-based drug recommendation method that employs GIN to effectively extract the molecular and substructural features of drugs. To address the critical issue that different drugs often share identical molecular representations, which limits the accuracy of traditional recommendation methods, our approach integrates multiple relational features from EHRs—including drug-procedure, drug-diagnosis, and drug-drug relationships—into the drug molecular structure feature space. Furthermore, we enhance the quality of drug representation learning by incorporating detailed substructure information and designing a gated attention mechanism, which adaptively learns drug representations with substructure-aware characteristics and highlights the key substructural components that affect drug efficacy. Finally, we leverage cross-view con-

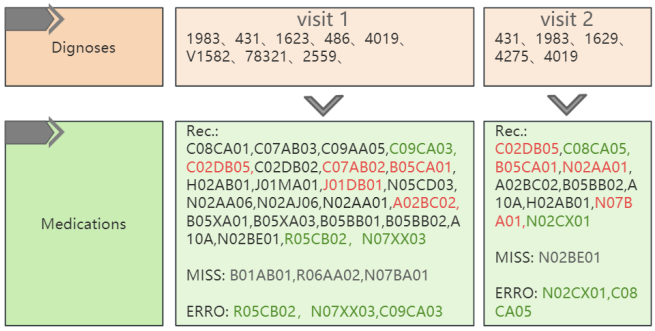

Fig. 6. Patient visit case

trastive learning to align and fuse features from different views, combined with an attention-based fusion strategy, to jointly integrate multi-perspective drug knowledge for more accurate, reliable, and clinically applicable drug recommendations. This integrated framework effectively overcomes the limitations of single-view feature learning and provides a more comprehensive basis for personalized drug recommendation. In future work, we will explore the interpretability of the recommendation process to make the model more suitable for practical clinical decision support scenarios.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] An Y, Zhang L, You M, Tian X, Jin B, Wei X. MeSIN: Multilevel selective and interactive network for medication recommendation. Knowledge-Based Systems. 2021;233:107534.

[2] Mu S, Li C, Li X, Liang S. Medication recommendation via dual molecular modalities and multi-step enhancement. Expert Systems with Applications. 2025;276:127163.

[3] Chen Q, Li X, Geng K, Wang M. Context-aware safe medication recommendations with molecular graph and DDI graph embedding. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37; 2023. p. 7053-60.

[4] Li X, Zhang Y, Islam Fu, Dong D, Wei H, Lu M. JLAN: medical code prediction via joint learning attention networks and denoising mechanism. BMC bioinformatics. 2021;22(1):590.

[5] Liu G, Yu X, Liu Z, Liu S, Li X, Fan X, et al. DNMDR: Dynamic networks and multi-view drug representations for safe medication recommendation. Knowledge-Based Systems. 2025;329:114327.

[6] Yu X, Li X, Zhao F, Yan X, Zheng X, Li T. AKA-SafeMed: A safe medication recommendation based on attention mechanism and knowledge augmentation. Information Sciences. 2024;670:120577.

[7] Ge F, Yu X, Li X, Fan X, Zhao Y. Personalized and safe medication recommendation based on convolutional neural network and transformer architecture. Engineering Applications of Artificial Intelligence. 2025;161:112267.

[8] Wang Y, Min Y, Chen X, Wu J. Multi-view graph contrastive representation learning for drug-drug interaction prediction. In: Proceedings of the web conference 2021; 2021. p. 2921-33.

[9] Li X, Zhang Y, Li X, Wei H, Lu M. DGCL: distance-wise and graph contrastive learning for medication recommendation. Journal of Biomedical Informatics. 2023;139:104301.

[10] Fan X, Yu X, Li X, Ge F, Zhao Y. LMGA: Lightweight multi-graph augmentation networks for safe medication recommendation. Journal of King Saud University - Computer and Information Sciences. 2024;36(10):102245.

[11] Yang N, Zeng K, Wu Q, Yan J. Molerec: Combinatorial drug recommendation with substructure-aware molecular representation learning. In: Proceedings of the ACM web conference 2023; 2023. p. 4075-85.

[12] Yang C, Xiao C, Ma F, Glass L, Sun J. Safedrug: Dual molecular graph encoders for recommending effective and safe drug combinations. arXiv preprint arXiv:210502711. 2021.

[13] Zou X, He X, Zheng X, Zhang W, Chen J, Tang C. DAI-Net: Dual Adaptive Interaction Network for Coordinated Medication Recommendation. Journal on Biomedical and Health Informatics (J-BHI). 2024;28(10):11.

[14] Zhang R, Wang X, Wang P, Meng Z, Cui W, Zhou Y. HTCL-DDI: a hierarchical triple-view contrastive learning framework for drug–drug interaction prediction. Briefings in Bioinformatics. 2023;24(6).

[15] Li X, Yu X, Liu G, Fan X, Ge F, Zhao Y, et al. Transformer-based medication recommendation with a multiple graph augmentation strategy. Expert Systems with Applications. 2024;257:125091.

[16] Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? arXiv preprint arXiv:181000826. 2018.

[17] Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. arXiv preprint arXiv:171010903. 2017.

[18] Johnson AE, Pollard TJ, Shen L, Lehman LwH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific data. 2016;3(1):1-9.

[19] Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. Mimic-iv (version 1.0). PhysioNet. 2021;101:e215-20.

[20] Zhang Y, Chen R, Tang J, Stewart WF, Sun J. LEAP: learning to prescribe effective and safe treatment combinations for multimorbidity. In: proceedings of the 23rd ACM SIGKDD international conference on knowledge Discovery and data Mining; 2017. p. 1315-24.

[21] Shang J, Xiao C, Ma T, Li H, Sun J. Gamenet: Graph augmented memory networks for recommending medication combination. In: proceedings of the AAAI Conference on Artificial Intelligence. vol. 33; 2019. p. 1126-33.

[22] Wu R, Qiu Z, Jiang J, Qi G, Wu X. Conditional generation net for medication recommendation. In: Proceedings of the ACM web conference 2022; 2022. p. 935-45.