Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 16, No. 2, 2026

A Medication Recommendation Method Based on Multi-View Medication Representation and a Time-Aware Copy Mechanism

Kefan Zhang

School of Computer Science and Technology, Taiyuan Normal University, Taiyuan, Shanxi, 030000, China

Abstract: Electronic health records provide an important data foundation for intelligent medication assistance. Although existing medication recommendation methods have achieved notable progress in recommendation accuracy, they still suffer from insufficient medication representation learning and coarse-grained utilization of historical prescriptions. To address these limitations, this paper proposes a medication recommendation method based on multi-view medication representation and a time-aware copy mechanism. At the medication representation level, information from the EHR co-occurrence graph, drug--drug interaction graph, and drug molecular structure graph is integrated to construct a multi-view medication representation, thereby enhancing the structural and semantic expressiveness of medication embeddings. At the level of historical information utilization, a time-aware copy mechanism is introduced, which jointly considers temporal decay and diagnosis similarity during historical prescription reuse, so as to improve the specificity and reliability of historical prescription selection. Experiments conducted on the public MIMIC-III dataset demonstrate that the proposed method outperforms multiple baseline models in terms of Jaccard, F1, and PR-AUC, while also achieving better safety performance on DDI Rate. Furthermore, ablation studies verify the effectiveness of both the multi-view medication representation module and the time-aware copy mechanism. The results indicate that the proposed method can improve recommendation accuracy while simultaneously accounting for medication safety, thus providing an effective solution for intelligent medication assistance.

Keywords: Medication Recommendation; Electronic Health Records; Multi-view Medication Representation; Time-aware Copy Mechanism; Drug--drug Interaction; Generative Model.

1. Introduction

With the continuous advancement of medical informatization, a large amount of clinical data, including diagnoses, procedures, and prescriptions, has been accumulated in electronic health records (EHRs), providing an important foundation for research on intelligent medication assistance. The goal of the medication recommendation task is to generate reasonable medication combination suggestions for physicians based on a patient's current visit information and historical clinical records, thereby supporting clinical prescription decision-making. Compared with traditional medication practices that mainly rely on physicians' experience, deep learning-based medication recommendation methods are able to mine potential associations between patient conditions and medication combinations from large-scale real-world clinical data. Such methods are therefore of great significance for improving prescription decision efficiency, reducing the risk of inappropriate medication use, and promoting the development of smart healthcare.

In recent years, various methods have been proposed for medication recommendation in the EHR setting. Early approaches generally formulated this task as an independent multi-label classification problem. Although simple to implement, such methods have difficulty capturing the combinational dependencies among medications. With the development of medical sequential modeling, an increasing number of studies have begun to model patients' longitudinal visit sequences and improve recommendation performance through recurrent neural networks, attention mechanisms, memory networks, and generative frameworks. Meanwhile, medication co-occurrence relationships, drug--drug interaction (DDI) [1] relationships, and drug molecular structure information have gradually been incorporated into medication recommendation models to enhance medication representation and improve recommendation safety. In particular, generative recommendation methods formulate medication combination prediction as a conditional generation process, making them more suitable for handling the complex label dependency problem in polypharmacy scenarios.

To address the above issues, this paper proposes a medication recommendation method based on multi-view medication representation and a time-aware copy mechanism under a generative medication recommendation framework. The proposed method is built upon the Copy-or-Predict framework of COGNet[2]. While preserving the overall generation process, it improves the model from two aspects: medication representation learning and historical prescription utilization. On the medication representation side, this paper integrates three types of information, namely the EHR co-occurrence graph, DDI graph, and drug molecular structure graph, to construct a multi-view medication representation, thereby enhancing the structural and semantic expressiveness of medication embeddings. On the historical information utilization side, a time-aware copy mechanism is introduced, which jointly considers temporal decay and diagnosis similarity when copying historical prescriptions. In this way, the model can improve the specificity of historical prescription selection and reduce the interference of outdated or weakly relevant historical records in current recommendation.

The main contributions of this paper are summarized as follows:

- A multi-view medication representation module is proposed to jointly model medication information from the perspectives of clinical co-occurrence relationships, interaction relationships, and molecular structure, thereby enhancing the model's ability to represent complex medication semantics.
- A time-aware copy mechanism is proposed to introduce both temporal constraints and disease-similarity constraints into the process of historical prescription reuse, thereby enabling more fine-grained utilization of historical information.

2. Problem Definition

The goal of medication recommendation is to predict an appropriate set of medications for a patient at the current visit based on longitudinal clinical records. Since a patient's condition evolves over time, prescription decisions are typically influenced not only by the information observed at the current encounter, but also by patterns reflected in previous visits. Therefore, the task can be formulated as a sequential modeling problem over electronic health records .

Let P denote the set of patients. For any patient $p \in P$, the complete clinical history is represented as a time-ordered visit sequence $V_p = \{v_1, v_2, \ldots, v_T\}$, where T is the total number of visits and v_t denotes the record of the t-th visit. Each visit usually consists of three major components, including diagnoses, procedures, and prescribed medications. Accordingly, a visit can be represented as $v_t = (D_t, O_t, M_t)$, where D_t denotes the diagnosis set, O_t denotes the procedure set, and M_t denotes the ground-truth medication set at visit t.

Based on the above formulation, the medication recommendation task aims to learn a predictive function that maps the historical visit sequence together with the clinical observations at the current visit to a candidate medication set for the current encounter. Since multiple medications are often prescribed simultaneously, this task can be naturally viewed as a multi-label prediction problem. Let M denote the medication vocabulary. The model output can then be represented as a probability vector of length $|M|$, where each dimension indicates the probability that the corresponding medication should be recommended for the current visit. Unlike general multi-label recommendation tasks, medication recommendation must consider not only predictive accuracy but also medication safety. In particular, the model should minimize the risk of recommending potentially harmful drug-drug interactions while maintaining strong predictive performance. Motivated by the limitations of existing methods in drug representation learning and historical prescription utilization, this study develops a medication recommendation framework based on multi-view drug representation and a time-aware copy mechanism.

3. Medication Recommendation Method Based on Multi-View Medication Representation and a Time-Aware Copy Mechanism

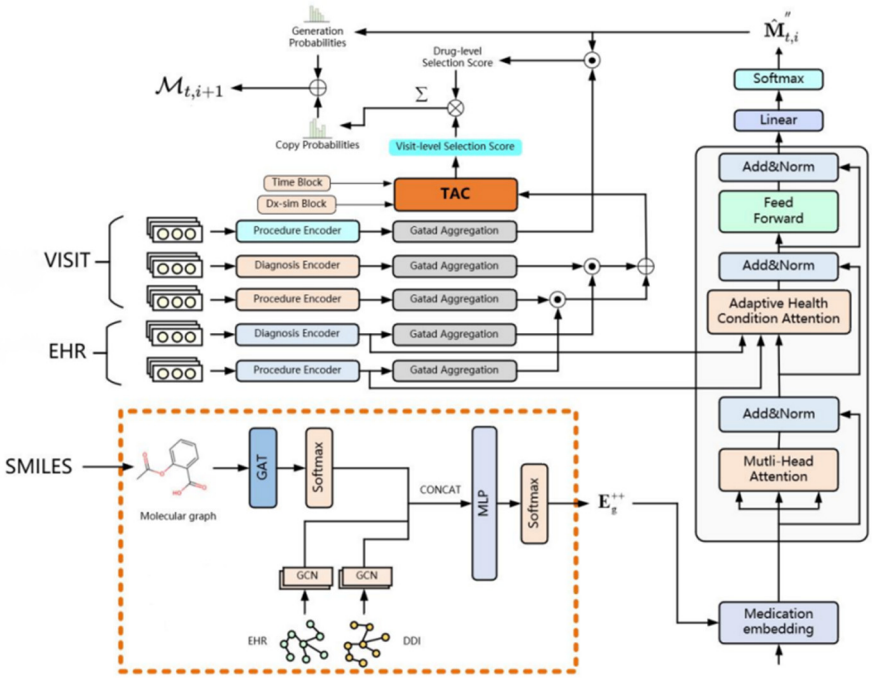

Fig 1. Overall Model Framework

3.1. Overall Model Framework

The MVTA_COGNet model (as illustrated in Figure 1) is built upon the COGNet architecture and preserves its fundamental Encoder--Decoder and Copy-or-Predict framework. Without altering the original generation process, the proposed model introduces improvements from two aspects, namely medication representation learning and historical prescription utilization, so as to enhance the accuracy and rationality of medication recommendation results. For a given patient, the model takes the current visit information and historical visit records as input, and progressively generates the medication sequence corresponding to the current visit in an autoregressive manner.

Overall, the model consists of five main components: an encoder, a multi-view medication representation module, a decoder, a time-aware copy module, and the loss function and

training strategy. The encoder is used to model the patient's current clinical state and historical context. The multi-view medication representation module integrates information from the EHR co-occurrence graph, DDI graph, and molecular structure graph to enhance the semantic representation of medications. The decoder is responsible for progressively generating medication representations. The time-aware copy module dynamically selects candidate medications from historical prescriptions by jointly considering temporal intervals and diagnosis similarity. Finally, by combining the generation distribution and the copy distribution, the model produces the output probability at the current decoding step.

3.2. Multi-View Drug Representation Module

To strengthen drug representation in the generation branch, we introduce a multi-view drug representation module that jointly models drug-side information from multiple complementary sources. Specifically, the module integrates EHR co-occurrence relations, drug-drug interaction relations, and molecular structure information to construct a semantically enriched drug representation matrix, which serves as the drug dictionary for the generation distribution in the decoder. For each drug, representations are learned from three perspectives: the EHR co-occurrence view captures collaborative prescription patterns in real clinical practice, the DDI view encodes potential risk relationships between drugs, and the molecular structure view provides intrinsic chemical property information. By combining these views within a unified representation space, the model is able to characterize clinical usage patterns, safety constraints, and molecular priors simultaneously, thereby improving the discriminative power of candidate drug representations.

First, under the EHR co-occurrence view, two drugs are considered related if they appear in the same prescription. Based on this assumption, we construct a drug co-occurrence graph, where nodes correspond to drugs and edges indicate co-prescription relations. To capture collaborative medication patterns from historical prescriptions, a graph convolutional network (GCN)[3] is applied to encode the co-occurrence graph. Let $H^{(0)}$ denote the initial drug embedding matrix. The propagation rule at the l-th GCN layer is defined as:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \quad (1)$$

where $\tilde{A}$ is the adjacency matrix with self-loops, $\tilde{D}$ is the corresponding degree matrix, $W^{(l)}$ is the learnable parameter matrix at layer l, and $\sigma(\cdot)$ denotes a nonlinear activation function. After multiple propagation layers, the model obtains the drug representation matrix E^{ehr} under the EHR co-occurrence view. This representation encodes prescription-level co-usage patterns and provides clinically grounded prior knowledge for subsequent drug generation.

Second, to explicitly incorporate medication safety information, we further construct a DDI-based representation view using known drug interaction knowledge. Let G^{ddi} denote the drug interaction graph, where nodes still represent drugs and edges indicate known interacting drug pairs. Similar to the co-occurrence view, a GCN is employed to learn structural representations from the DDI graph. The propagation process can be formulated as:

$$Z^{(l+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}Z^{(l)}U^{(l)}\right) \quad (2)$$

where $\hat{A}$ is the adjacency matrix of the DDI graph, $\hat{D}$ is the associated degree matrix, and $U^{(l)}$ denotes the learnable parameter matrix. The resulting representation matrix under the interaction view is denoted by E^{ddi}. Unlike the EHR co-occurrence view, which emphasizes collaborative usage patterns, the DDI view focuses on modeling potential risk structures among drugs, allowing the model to encode safety-aware priors during the generation process.

In addition to clinical relational information, the molecular structure of a drug is also an important source of representation. For each drug, a molecular graph is constructed from its SMILES sequence, where atoms are treated as nodes and chemical bonds as edges. Based on this graph, a graph attention network (GAT) is adopted to encode molecular structure features. For a node and its neighboring nodes in the molecular graph, the attention coefficient is computed as follows:

$$\alpha_{ij} = \mathrm{softmax}_j\left(\mathrm{LeakyReLU}\left(a^{\top}\left[Wh_i \parallel Wh_j\right]\right)\right) \quad (3)$$

where h_i denotes the node feature, W and a are learnable parameters, and $\parallel$ denotes vector concatenation. Neighboring information is then aggregated according to the learned attention weights, and a graph-level readout operation is further applied to obtain the molecular representation of each drug. By repeating this process for all drugs, the model derives the molecular-view representation matrix E^{mol}. Compared with representations based solely on clinical relations, this view provides intrinsic chemical information and is therefore beneficial for modeling low-frequency drugs and rare medication combinations.

After obtaining representations from the three views, we employ a fusion mechanism to integrate them into a unified drug representation. For each drug, the three view-specific embeddings e^{ehr}, e^{ddi}, and e^{mol} are first concatenated to form a joint feature vector:

$$e^{cat} = \left[e^{ehr} \parallel e^{ddi} \parallel e^{mol}\right] \quad (4)$$

The concatenated feature is then passed through a multilayer perceptron to produce the final multi-view drug representation:

$$e^{mv} = \mathrm{MLP}(e^{cat}) \quad (5)$$

Applying the same operation to all drugs yields the final multi-view drug representation matrix E^{mv}. This representation jointly encodes co-prescription information, interaction-based risk information, and molecular structural information, and is used as the drug dictionary for the generation distribution. As a result, the model gains a richer and more discriminative representation space for candidate medications. To maintain consistency between the input embedding space and the output prediction space, the decoder input embeddings and the output-side multi-view drug representations may share parameters fully or partially, or be aligned through a linear transformation.

Overall, the proposed multi-view drug representation module unifies three complementary perspectives of a drug—clinical co-occurrence, interaction risk, and molecular structure—within a single representation framework. By moving beyond a single-view representation scheme, the module provides a more complete characterization of drug semantics and lays a stronger foundation for accurate and safe medication generation.

3.3. Time-Aware Copy Mechanism

To improve the specificity and reliability of historical prescription copying, this study extends the original Copy-or-Predict framework by introducing a time-aware copy mechanism, which explicitly enhances the modeling of

reference weights assigned to historical visits in the copy pathway. While preserving semantic matching information, the proposed mechanism further incorporates both the temporal interval between historical and current visits and the similarity between their diagnosis sets to dynamically regulate the contribution of historical records. In this way, it reduces the interference caused by outdated prescriptions or weakly relevant historical visits in the current recommendation process. The central intuition is that, in sequential visit scenarios, historical prescriptions do not provide equal clinical value: records that are temporally closer to the current visit and more similar to the current condition should generally be prioritized.

The proposed improvement mainly operates in the first stage of the copy mechanism, namely, the redefinition of visit-level weights over historical visits. Let h_k denote the decoder hidden state at the k-th generation step, and let v_i represent the embedding of the i-th historical visit. Then, the semantic matching score in the original copy pathway can be formulated as

$$s_{k,i}^{sem} = f(h_k, v_i) \tag{6}$$

where $f(\cdot)$ denotes the semantic matching function between the hidden state at the current decoding position and the representation of the historical visit. This score reflects the relevance between the current generation state and the content of the historical visit, and thus serves as the fundamental basis for historical prescription copying. However, semantic matching alone is insufficient, because even when two historical records are semantically similar, their clinical reference value may still vary substantially due to temporal distance and changes in disease status.

To address this issue, we further introduce a temporal decay factor to characterize how the reference value of a historical visit changes over time. Let t_{cur} denote the time of the current visit, and t_i the time of the i-th historical visit. The temporal interval is defined as $\Delta t_i = t_{cur} - t_i$. The temporal modulation term is then given by

$$g_i^{time} = \exp(-\lambda \Delta t_i) \tag{7}$$

where λ is a temporal decay coefficient controlling the rate at which historical information attenuates over time. A larger temporal interval indicates that the historical visit is further away from the current time point and therefore likely to have lower reference value for the current prescription decision. By incorporating this term, the model can naturally suppress the influence of distant historical records in the copy pathway, thereby assigning greater importance to more recent prescriptions during decision-making.

In addition to temporal information, we also consider the consistency between the current visit and each historical visit in terms of diagnostic structure. Let D_{cur} denote the diagnosis set of the current visit and D_i the diagnosis set of the i-th historical visit. Their diagnostic-set similarity is measured using the Jaccard coefficient:

$$g_i^{diag} = \frac{|D_{cur} \cap D_i|}{|D_{cur} \cup D_i|} \tag{8}$$

This similarity quantifies the degree of overlap between the current and historical disease conditions at the diagnosis level. A higher similarity indicates that the corresponding historical prescription is more likely to be aligned with the current therapeutic objective, and thus should be assigned greater reference value. Compared with modulation based solely on temporal information, the incorporation of diagnostic similarity further highlights historical records that are structurally more consistent with the current clinical condition, thereby improving the clinical specificity of the copying decision.

By jointly considering semantic matching, temporal decay, and diagnosis-set similarity, the final score for each historical visit at the current decoding position is defined as

$$s_{k,i} = s_{k,i}^{sem} \cdot g_i^{time} \cdot g_i^{diag} \tag{9}$$

where $s_{k,i}^{sem}$ provides the basic semantic matching information between the current decoding state and the historical visit, g_i^{time} suppresses historical records with weaker temporal relevance, and g_i^{diag} enhances those that are more consistent with the current disease structure. Based on the joint score, the time-aware visit-level copy weight is obtained via normalization:

$$\alpha_{k,i} = \frac{\exp(s_{k,i})}{\sum_j \exp(s_{k,j})} \tag{10}$$

where $\alpha_{k,i}$ denotes the reference intensity assigned by the model to the i-th historical visit at the k-th generation step. This weight is subsequently used to aggregate historical prescription information and to participate in the computation of the copy distribution for the current medication.

Through this design, the time-aware copy mechanism enables the model to refine the utilization of historical prescriptions by jointly considering disease-course temporal information and diagnostic-structure information on top of the original semantic matching basis. Compared with copy strategies that rely solely on semantic matching, the proposed mechanism can more effectively reduce the interference of outdated or weakly relevant historical records, thereby improving the stability and specificity of the copy pathway. In scenarios where the patient's condition remains relatively stable, the mechanism helps the model preserve previously appropriate prescriptions; when the condition changes, it suppresses historical medication information that is no longer applicable, thus enhancing the rationality and robustness of the recommendation results.

3.4. Model Training

Since medication combination generation can be formulated as a conditional sequence generation problem, this study adopts the cross-entropy loss based on maximum likelihood estimation (MLE) as the primary training objective. During training, the model takes the patient's historical visit sequence and current visit information as input, and progressively generates the medication sequence corresponding to the current prescription during decoding. Let the ground-truth medication sequence for the t-th visit of the n-th patient be denoted by $Y^{(n,t)} = \{y_1, y_2, \ldots, y_L\}$, where L is the length of the medication sequence for that visit. We employ the teacher forcing strategy during training; specifically, at the k-th generation step, the decoder predicts the probability distribution of the current medication conditioned on the first $k-1$ ground-truth medications. This strategy improves training stability and accelerates convergence.

Under this formulation, the generation loss for a single visit is defined as

$$\mathcal{L}_{gen}^{(n,t)} = -\sum_{k=1}^{L} \log P\left(y_k \mid y_{<k}, X^{(n,t)}\right) \tag{11}$$

where $X^{(n,t)}$ denotes the input information associated with the current visit, including the patient's historical visit records, current diagnosis information, and procedure information, and $y_{<k}$ represents the ground-truth medication prefix before the current position. This loss function directly constrains the consistency between the model output distribution and the ground-truth prescription sequence, allowing the generation pathway and the copy pathway to be optimized jointly under a unified objective.

By accumulating the generation loss over all patients and their valid visits, the overall cross-entropy loss is obtained as

$$\mathcal{L}_{ce} = \sum_{n} \sum_{t} \mathcal{L}_{gen}^{(n,t)} \tag{12}$$

This term serves as the main supervision signal during training, guiding the model to learn the mapping among the current clinical state, historical prescriptions, and medication relation information, thereby improving the agreement between the recommended results and the ground-truth prescriptions.

Considering that the proposed model adopts a Copy-or-Predict mechanism, if the gating distribution becomes overly biased toward either the generation branch or the copy branch at an early training stage, the other branch may receive insufficient effective gradients, which can adversely affect overall optimization. To alleviate this issue, we introduce a lightweight gating entropy regularization term in addition to the main loss, in order to encourage moderate diversity in the gating distribution during early training. Let the gating distribution at the k-th generation step be denoted by $g_k = [g_k^{gen}, g_k^{copy}]$. Its entropy is defined as

$$\mathcal{H}(g_k) = -\sum_{r\in\{gen,copy\}} g_k^r \log g_k^r. \tag{13}$$

Based on this, the gating entropy regularization term is formulated as

$$\mathcal{L}_{ent} = -\sum_{k} \mathcal{H}(g_k) \tag{14}$$

The purpose of this term is to prevent the gating mechanism from collapsing too rapidly to a single pathway during early training, so that both the generation branch and the copy branch can be sufficiently trained, thereby improving the overall stability and robustness of the model.

By combining the primary objective with the auxiliary regularization term, the final optimization objective of the model can be written as

$$\mathcal{L} = \mathcal{L}_{ce} + \beta \mathcal{L}_{ent} \tag{15}$$

where β is a balancing coefficient that controls the contribution of the gating entropy regularization term to the total loss. Since the final output probability is differentiable with respect to the parameters of the encoder, the multi-view medication representation module, the time-aware copy module, and the decoder, the entire model can be trained end-to-end under a unified objective.

4. Experiments and Results

4.1. Experimental Setup

To evaluate the effectiveness of the proposed method in the medication recommendation task, experiments were conducted on the publicly available MIMIC-III dataset [4]. This dataset is derived from real-world electronic health records collected in intensive care units and contains information on diagnoses, procedures, and prescribed medications across multiple hospital admissions, thereby providing reliable support for longitudinal medication recommendation research. In accordance with the requirements of this task, we mainly used structured data from the diagnosis, procedure, and prescription tables, and constructed patient visit sequences for medication recommendation through unified code mapping. To ensure the effectiveness of sequential modeling, only patients with no fewer than two visits were retained. In addition, the medication vocabulary was restricted to the top 300 most frequent drugs, so as to preserve the major clinical prescribing patterns while controlling the size of the output space.

In terms of evaluation, the proposed method was assessed from three perspectives: recommendation accuracy, ranking capability, and medication safety. Specifically, Jaccard and F1-score were adopted to measure the consistency between the recommended medication set and the ground-truth prescription, PR-AUC was used to evaluate the model's ranking and discrimination ability over candidate drugs, and DDI Rate was employed to quantify the potential risk of drug--drug interactions in the recommended results. In addition, the average number of recommended medications (Avg#Meds) was reported for auxiliary analysis of the output scale. Among these metrics, Jaccard and F1-score reflect the overall performance of the model at the set-prediction level, PR-AUC measures the quality of probability-based ranking, and DDI Rate evaluates the safety of the recommended medication combinations.

To comprehensively assess the performance of the proposed method, we compared it with a variety of representative baseline models for medication recommendation, including the traditional multi-label learning methods LR and ECC, the history-aware sequential models RETAIN and DMNC, the generative recommendation model LEAP, and models that incorporate structured medication relation modeling or safety constraints, including GAMENet, MICRON, SafeDrug, and COGNet. Among them, COGNet is the most closely related to our approach and can be used to directly verify the effectiveness of the proposed multi-view medication representation and time-aware copy mechanism. All models were trained and evaluated under the same data split, evaluation metrics, and experimental conditions to ensure the fairness and comparability of the results.

For hyperparameter settings, the embedding dimensions of diagnoses, procedures, and medications were all set to 64. Both the encoder and decoder adopted a two-layer Transformer architecture with two attention heads. The EHR co-occurrence graph and the DDI graph were both encoded using two-layer graph convolutional networks, with output dimensions aligned with the hidden dimension of the main model. The molecular structure view representations were obtained through offline precomputation and then used as static features during training of the main model. During training, the Adam optimizer was employed with an initial learning rate of 0.001, a batch size of 16, and a maximum of 50 epochs. During inference, beam search was used for sequence decoding, with the beam size set to 5. These unified settings help ensure training stability, fairness, and reproducibility under limited computational resources.

4.2. Comparative Experimental Results

To verify the effectiveness of the proposed method, we conducted a systematic comparison between MVTA_COGNet and a series of representative baseline models. All models were trained and evaluated under the same data split, evaluation metrics, and experimental settings. The comparison results are reported in Table 1.

Table 1. Comparison with Baseline Methods

Models	Jaccard	F1	PRAUC	DDI	Avg#Meds
LR	0.4865±0.0021	0.6234±0.0019	0.7509±0.0018	0.0829±0.0009	16.1773±0.0942
ECC	0.4996±0.0049	0.6369±0.0044	0.6844±0.0038	0.0846±0.0018	18.0722±0.1914
RETAIN	0.4887±0.0028	0.6481±0.0027	0.7556±0.0033	0.0835±0.0020	20.4051±0.2832
LEAP	0.4521±0.0024	0.6138±0.0026	0.6549±0.0033	0.0731±0.0008	18.7138±0.0666
DMNC	0.4864±0.0025	0.6529±0.0030	0.7580±0.0039	0.0842±0.0011	20.0180±0.0150
GAMENet	0.5067±0.0025	0.6626±0.0025	0.7631±0.0030	0.0864±0.0006	27.2145±0.1141
MICRON	0.5100±0.0033	0.6654±0.0031	0.7687±0.0026	0.0641±0.0007	17.9267±0.2172
SafeDrug	0.5213±0.0030	0.6768±0.0027	0.7647±0.0025	0.0589±0.0005	19.9178±0.1604
COGNet	0.5336±0.0011	0.6869±0.0010	0.7739±0.0009	0.0852±0.0005	28.0903±0.0950
MVTA-COGNet	0.5450±0.0012	0.6975±0.0011	0.7805±0.0010	0.0798±0.0006	27.6546±0.1043

Overall, the proposed model achieves the best performance on Jaccard, F1-score, and PR-AUC, demonstrating clear advantages in both recommendation accuracy and ranking capability. At the same time, the DDI Rate of MVTA_COGNet is lower than that of the direct baseline COGNet, while the average number of recommended medications does not increase substantially. This indicates that the performance gain is not achieved by simply enlarging the recommendation set; rather, the proposed model simultaneously improves recommendation effectiveness and medication safety while maintaining a relatively stable output scale.

From the comparison across different categories of methods, traditional multi-label approaches such as LR[5] and ECC show relatively weaker overall performance. These methods typically formulate medication recommendation as an independent label prediction problem and lack explicit modeling of the dependencies within medication combinations, which limits their effectiveness in polypharmacy scenarios. In contrast, history-aware sequential models such as RETAIN and DMNC achieve improvements on some metrics, suggesting that longitudinal clinical information provides important guidance for current prescription decisions. However, because their modeling of medication relations and historical prescription reuse remains insufficient, their overall performance is still constrained.

Further examination of the generative and structure-enhanced models shows that GAMENet[6], MICRON, SafeDrug[7], and COGNet consistently outperform traditional methods, indicating that medication co-occurrence relations, drug--drug interaction constraints, molecular structure information, and historical prescription copying mechanisms all contribute positively to medication recommendation performance. Among them, SafeDrug and MICRON achieve relatively favorable results in terms of DDI Rate, suggesting that explicit safety modeling can effectively reduce the risk of potential drug--drug interactions. In contrast, COGNet performs better on accuracy-related metrics, indicating that a generative framework combined with a historical copying mechanism is beneficial for improving the quality of medication combination prediction.

4.3. Ablation Study Analysis

To further verify the independent contribution of each key component to the overall model performance, we conducted removal-based ablation studies on the full MVTA_COGNet model. Since the main improvements proposed in this work consist of the multi-view medication representation module in the generation pathway and the time-aware mechanism in the copy pathway, we constructed several model variants from three perspectives: overall module removal, sub-view information removal, and elimination of key modulation factors in the copy pathway. The experimental results are presented in Table 2.

Table 2. Comparison with Baseline Methods

Models	**Jaccard**	**F1-score**	**PR-AUC**	**DDI Rate**	**Avg#Meds**
w/oMulti-View	0.5361±0.0013	0.6898±0.0012	0.7754±0.0011	0.0841±0.0006	27.95±0.11
w/oMolecularView	0.5402±0.0012	0.6926±0.0011	0.7776±0.0010	0.0829±0.0006	27.78±0.10
w/oTime-Aware	0.5415±0.0013	0.6937±0.0012	0.7784±0.0011	0.0837±0.0006	27.83±0.09
w/oDiagnosis-Sim	0.5437±0.0012	0.6962±0.0011	0.7759±0.0010	0.0811±0.0006	27.65±0.09
MVTA_COGNet	**0.5450±0.0012**	**0.6975±0.0011**	**0.7805±0.0010**	**0.0798±0.0006**	27.60±0.10

From the ablation results related to the generation pathway, removing the multi-view medication representation module leads to a clear decline in Jaccard, F1-score, and PR-AUC, while the DDI Rate also increases. This indicates that the multi-view medication representation module not only improves the accuracy of medication combination generation, but also enhances the safety of the recommendation results to some extent, confirming that it is a core enhancement component in the generation pathway. Compared with single-view medication representation, this module integrates clinical co-occurrence relations, drug--drug interaction relations, and molecular structure information, thereby providing the model with a more comprehensive semantic representation of medications.

Furthermore, when only the molecular structure view is removed, the model still performs better than the variant without the entire multi-view module, but it consistently underperforms the full model. This suggests that molecular structure information is not a redundant design; rather, it provides additional structural complementarity beyond EHR co-occurrence relations and DDI risk relations. Especially in scenarios involving complex medication combinations and low-frequency drugs, molecular structure information helps the model more accurately distinguish candidate medications with similar functions but different clinical applicability, thereby contributing positively to the overall recommendation performance.

From the ablation results related to the copy pathway, removing the time-aware mechanism causes performance degradation in both accuracy-related and safety-related metrics to varying degrees. This demonstrates that the time-aware mechanism can effectively suppress the interference of outdated historical prescriptions in the current recommendation process, thereby improving the specificity and stability of historical information utilization. Compared with a copy strategy that relies solely on semantic matching, the temporal decay factor enables the model to focus more on recent records that are more relevant to the current visit, thus enhancing the practical effectiveness of the copy pathway.

When the diagnosis-set similarity modulation is further removed while retaining the temporal component, the model performance also declines, although the degradation is smaller than that observed when the entire time-aware mechanism is removed. This indicates that diagnosis-set similarity plays a supportive enhancement role in historical prescription selection, whereas temporal decay remains the primary contributing factor in the copy pathway. In other words, temporal decay and diagnostic-structure similarity are complementary in historical prescription filtering: the former mainly reflects the timeliness of historical information, while the latter characterizes the consistency between the historical condition and the current condition. Their joint effect further improves the refinement of the copying decision.

Overall, the results in Table 2 show that both the proposed multi-view medication representation module and the time-aware copy mechanism contribute positively to the performance improvement of the model. Specifically, the former mainly strengthens the model's ability to capture medication semantics, while the latter improves the specificity and reliability of historical prescription utilization. Their combined effect enables the model to achieve a better balance between recommendation accuracy and medication safety, further validating the rationality and effectiveness of the proposed design.

5. Summary

To address the limitations of existing medication recommendation methods in insufficient medication representation learning and coarse-grained utilization of historical prescriptions, this study proposes a medication recommendation framework based on multi-view medication representation and a time-aware copy mechanism. Under a generative medication recommendation paradigm, the proposed method improves the model from two complementary aspects: enhanced medication representation and refined historical information utilization. On the representation side, we integrate information from the EHR co-occurrence graph, DDI graph, and molecular structure graph to construct a multi-view medication representation, thereby strengthening the structural and semantic expressiveness of medication embeddings. On the historical utilization side, we introduce a time-aware copy mechanism that jointly considers temporal decay and diagnosis similarity when copying historical prescriptions, thus improving the specificity and reliability of historical record selection.

Experimental results on the public MIMIC-III dataset demonstrate that the proposed method outperforms multiple baseline models in terms of Jaccard, F1-score, and PR-AUC, while also achieving better safety performance on DDI Rate. In addition, the ablation studies further verify the positive contribution of both the multi-view medication representation module and the time-aware copy mechanism to the overall model performance. These findings indicate that the proposed method is capable of improving recommendation accuracy while simultaneously accounting for medication safety, thereby achieving strong overall performance in the medication combination generation task.

Overall, this work suggests that improving generative medication recommendation models from the perspectives of multi-view medication information modeling and fine-grained historical prescription utilization is an effective strategy for enhancing medication recommendation performance. In future work, additional information sources such as laboratory test results, clinical text, and external medical knowledge could be incorporated to improve the model's adaptability to more complex clinical scenarios, while further enhancing the interpretability and practical applicability of the recommendation results.

References

[1] Zhong J, Zhao H, Zhao Q, Wang J. A Knowledge Graph-Based Method for Drug-Drug Interaction Prediction With Contrastive Learning. IEEE/ACM Trans Comput Biol Bioinform. 2024 Nov-Dec;21(6):2485-2495. doi: 10.1109/TCBB.2024. 3477 410. Epub 2024 Dec 10. PMID: 39383072.

[2] Wu R, Qiu Z, Jiang J, et al. Conditional generation net for medication recommendation[C]//Proceedings of the ACM web conference 2022. 2022: 935-945.

[3] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint arXiv: 1609. 02907, 2016.

[4] Johnson, A., Pollard, T., Shen, L. et al. MIMIC-III, a freely accessible critical care database. Sci Data 3, 160035 (2016).

[5] Tsoumakas G, Katakis I. Multi-label classification: An overview[J]. International Journal of Data Warehousing and Mining (IJDWM), 2007, 3(3): 1-13.

[6] Shickel B, Tighe P J, Bihorac A, et al. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis[J]. IEEE journal of biomedical and health informatics, 2017, 22(5): 1589-1604.

[7] Wishart D S, Feunang Y D, Guo A C, et al. DrugBank 5.0: a major update to the DrugBank database for 2018[J]. Nucleic acids research, 2018, 46(D1): D1074-D1082.