MGRec: Structure-Grounded Medication Recommendation via Condition-Aware Molecular Representation Learning

Jinke Feng[1,2] , **Wenjie Du**[1,2,*]

[1]University of Science and Technology of China (USTC), Hefei, China
[2]Suzhou Institute for Advanced Research, USTC, Suzhou, China
{fengjinke, duwenjie}@mail.ustc.edu.cn

Abstract

Medication recommendation plays a critical role in clinical decision-making by supporting personalized and safe treatment planning. Existing methods rely heavily on historical co-occurrence patterns and primarily optimize discrete prescription prediction objectives, limiting generalization in rare or emerging disease settings. We propose MGRec, a framework that shifts learning from discrete prescription prediction to condition-aware modeling in a continuous molecular representation space. MGRec treats molecular structure as the primary modeling target and employs a Therapeutic–Safety Factorized Conditional Variational Autoencoder to disentangle therapeutic and safety-related factors in a condition-aware molecular latent space. The model infers treatment-relevant molecular representations conditioned on current patient-specific clinical context, which are mapped to clinically approved medications for final recommendation. To improve clinical safety, we further introduce a DDI (drug-drug interaction)-guided latent regularization to integrate drug interaction knowledge at the representation level. Experiments on two real-world benchmarks demonstrate that MGRec achieves state-of-the-art accuracy and reduced interaction risk, particularly in data-sparse scenarios.

1 Introduction

The rapid digitalization of healthcare has led to an unprecedented accumulation of patient data, significantly transforming clinical decision-making [Senbekov *et al.*, 2020; Agrawal and Prabakaran, 2020]. Among these advancements, medication recommendation systems have emerged as essential tools for optimizing treatment strategies by leveraging vast repositories of Electronic Health Records (EHRs) [Dagliati *et al.*, 2021; Garriga *et al.*, 2022]. Recent advances in artificial intelligence (AI) have enabled the development of predictive models that integrate EHR data to provide more informed and data-driven medication recommendations [Zhang *et al.*, 2017; Shang *et al.*, 2019; Yang *et al.*, 2021b]. These models aim to enhance therapeutic efficacy by identifying optimal drug [1] regimens tailored to patient characteristics.

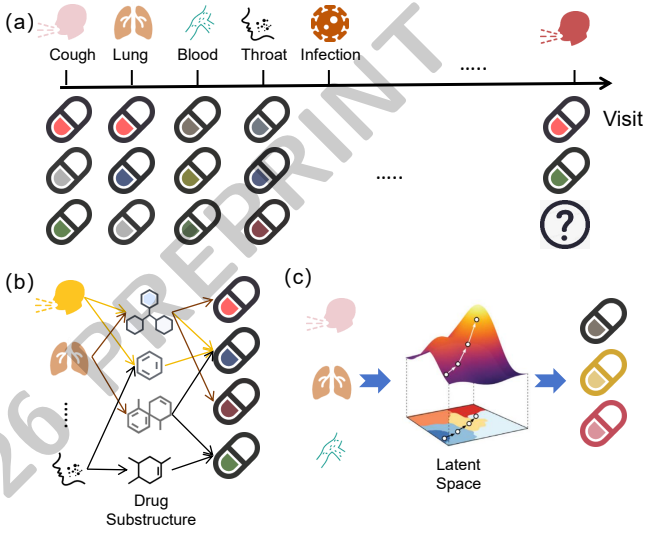

Figure 1: Overview of various medication recommendation approaches: (a) Historical data-driven, (b) Molecular substructure-based, (c) Structure-Grounded approach.

Early medication recommendation models are primarily historical data-driven methods, learning from co-occurrence patterns between patient conditions and discrete medication sets [Choi *et al.*, 2016; Zhang *et al.*, 2017; Shang *et al.*, 2019; Liu *et al.*, 2023; Yang *et al.*, 2021a; Wu *et al.*, 2022], as illustrated in Figure 1(a). While effective in routine clinical scenarios, these methods exhibit limited generalization in sparse settings applied for rare or emerging diseases, where the corresponding medication combinations are rarely observed or entirely absent from historical data. [Jones *et al.*, 2008; Schieppati *et al.*, 2008; Esteva *et al.*, 2019; Faviez *et al.*, 2020]. Recent advancements [Yang *et al.*, 2021b; Yang *et al.*, 2023; Kuang and Xie, 2024] have incorporated molecular substructure to enrich the drug representations, and capture finer-grained associations between patient conditions and pharmacological characteristics, thereby achieving

*: corresponding author

[1]In this paper, "medication" and "drug" are used interchangeably to refer to substances used for the treatment of diseases.

improved predictive performance, as shown in Figure 1(b). However, these methods still fundamentally rely on observed patient–medication data to drive learning, with structural information serving to enhance drug representations rather than providing independent supervision to actively learn clinically plausible treatment associations that are not explicitly present in the data, leading to constrained generalization in data sparse settings.

These limitations suggest that when supervision over patient–medication data is scarce, simply enriching drug representations is insufficient, as learning signals remain coupled to observed prescription labels [Kou *et al.*, 2024; Kou *et al.*, 2025]. A more effective strategy is therefore to shift the learning focus away from direct prediction of discrete medication sets to treating molecular structure itself as the modeling target, where the model can be trained to infer the structural and pharmacological characteristics that plausible medications should possess. In this way, structural information directly contributes learning signals through a continuous representation space, rather than relying solely on whether specific prescriptions are observed. Such a reformulation allows the model to exploit smooth relationships induced by both patient similarity and molecular similarity, providing a more suitable foundation for medication recommendation under sparse supervision. Moreover, this formulation naturally enables safety considerations, such as drug–drug interaction (DDI) risks, to be incorporated directly into the representation learning process, rather than imposed as post-hoc constraints on discrete recommendations.

Motivated by this reformulation, we propose the Medication Structure Grounded Recommendation Framework (**MGRec**), which grounds medication recommendation in a continuous molecular representation space while retaining discrete medications as clinically actionable outputs. As illustrated in Figure 1(c), MGRec integrates patient-specific clinical context with molecular structural knowledge through a Therapeutic–Safety Factorized Conditional Variational Autoencoder (**TSF-CVAE**). TSF-CVAE performs clinical–molecular context fusion to construct a unified condition representation, and learns a factorized latent space that disentangles therapeutic relevance from safety-sensitive molecular factors. The model infer molecular representations that capture core therapeutic characteristics for the current patient conditions. Final recommended medications labels are instantiated by mapping inferred representations to the closest clinically approved drugs, ensuring practical applicability.

Meanwhile, to enhance the clinical safety of the recommended drug combinations, MGRec incorporates **DDI-Guided Latent Regularization**, which injects pharmacological interaction knowledge into the safety subspace and integrates safety considerations directly into the learned condition-aware molecular representations. MGRec not only improves recommendation accuracy and safety for rare and emerging diseases. In conclusion, the main contributions of this paper can be summarized as follows:

- We propose MGRec, a medication recommendation framework that shifts the learning objective from direct discrete prescription prediction to patient condition-aware modeling in a continuous molecular representation space, improving robustness under sparse supervision.
- We introduce a Therapeutic–Safety Factorized Conditional VAE (TSF-CVAE) that disentangles therapeutic and safety-related molecular factors, together with a DDI-guided latent regularization mechanism that internalizes pharmacological safety constraints at the representation level.
- Extensive experiments on two real-world benchmark datasets demonstrate that MGRec achieves state-of-the-art performance in both recommendation accuracy and interaction risk reduction, especially in data-sparse settings.

2 Related Work

2.1 Medication Recommendation

Existing approaches for medication recommendation can be broadly categorized into historical data-driven methods and molecular substructure-informed methods. Historical data-driven methods rely on patient EHR data to recommend medications, and can be further divided into instance-based methods and longitudinal methods. Instance-based methods such as LEAP [Zhang *et al.*, 2017] focus on the patient information from a specific visit. Longitudinal methods incorporate historical records to capture temporal dependencies. An example is GAMENet [Shang *et al.*, 2019], which integrates memory neural networks with graph convolutional networks to encode longitudinal EHR data. Molecular substructure-informed methods integrate drug molecular structures to enhance recommendation accuracy and safety. These methods include SafeDrug [Yang *et al.*, 2021b], which incorporates molecular structure information to improve medication safety, and MoleRec [Yang *et al.*, 2023], which directly associates patient health conditions with molecular substructures. Notably, RAREMed [Zhao *et al.*, 2024] is the most relevant work to ours, as it explicitly focuses on rare disease medication recommendation. It leverages pre-training to improve medication prediction for rare disease cases, while remaining constrained by its reliance on historical patient records.

2.2 VAE for Condition-Aware Molecular Representation Learning

Variational Autoencoders (VAEs) [Kingma, 2013] provide a principled framework for modeling complex data distributions in a continuous latent space and have been widely adopted for learning molecular representations with semantic structure [Wigh *et al.*, 2022; Zeng *et al.*, 2022; Shi and others, 2025; Ochiai *et al.*, 2023]. In the molecular domain, VAEs have been shown to effectively encode molecular graphs into smooth latent spaces that preserve structural similarity, enabling interpolation and uncertainty-aware inference [Gómez-Bombarelli *et al.*, 2018; Samanta *et al.*, 2020; Jin *et al.*, 2018] for downstream task such as molecule generation [Xue *et al.*, 2019; Pang *et al.*, 2023]. In this work, we leverage VAEs not as generators of new molecules, but as conditional distributional models for learning patient-specific molecular representations to provide continuous supervision and safety control.

3 Problem Formulation

For a patient x, the EHR is represented as a sequence $\mathcal{V}^{(x)} = [v^{(1)}, v^{(2)}, \ldots, v^{(N_x)}]$, where N_x is the total number of visits for the patient, and $v^{(i)}$ represents the details of the i-th visit. Each visit $v^{(i)}$ consists of three main components: $v^{(i)} = [\mathbf{v}_d^{(i)}, \mathbf{v}_p^{(i)}, \mathbf{v}_m^{(i)}]$. Here, $\mathbf{v}_d^{(i)} \in \{0,1\}^{|D|}$, $\mathbf{v}_p^{(i)} \in \{0,1\}^{|P|}$ and $\mathbf{v}_m^{(i)} \in \{0,1\}^{|M|}$ are multi-hot vectors representing the diagnosis, procedures and medications prescribed respectively. At time step t, given the longitudinal diagnosis sequence $\mathbf{v}_d^t = \left[\mathbf{v}_d^{(1)}, \mathbf{v}_d^{(2)}, \ldots, \mathbf{v}_d^{(t)}\right]$, procedure sequence: $\mathbf{v}_p^t = \left[\mathbf{v}_p^{(1)}, \mathbf{v}_p^{(2)}, \ldots, \mathbf{v}_p^{(t)}\right]$ and medication sequence: $\mathbf{v}_m^{t-1} = \left[\mathbf{v}_m^{(1)}, \mathbf{v}_m^{(2)}, \ldots, \mathbf{v}_m^{(t-1)}\right]$, our objective is to learn a function $f(\cdot)$ that generates a multi-label output $\hat{\mathbf{m}}^{(t)} \in \{0,1\}^{|\mathcal{M}|}$. Specifically, $\hat{\mathbf{m}}^{(t)} = f(\mathbf{v}_d^t, \mathbf{v}_p^t, \mathbf{v}_m^{t-1})$.

4 Methodology

As shown in Figure 2, our proposed MGRec consists of three components: 1) **Health condition embedding module** processing longitudinal patient medical records to obtain a comprehensive representation of patient health condition, 2) **Conditional representation learning module** learning ideal molecular representations under various contextualized conditions, 3) **Drug inference module** inferring personalized drugs according to patient's current health condition.

4.1 Health Condition Embedding Module

To comprehensively characterize a patient's health condition, we consider two key aspects: the current health condition and the previous health condition. Two embedding layers $\mathbf{E}_d$ and $\mathbf{E}_p$ are employed to transform the $\mathbf{v}_d^t$ and $\mathbf{v}_p^t$ into vectorized representations:

$$\mathbf{V}_{\mathrm{d}}^t = \mathbf{E}_d \mathbf{v}_d^t, \quad \mathbf{V}_{\mathrm{p}}^t = \mathbf{E}_p \mathbf{v}_p^t. \tag{1}$$

Subsequently, to capture temporal dependencies across multiple visits, we apply transformer-based encoders [Vaswani *et al.*, 2017], denoted as $\mathrm{T}(\cdot)$, to yield contextualized encodings of the diagnosis and procedure sequences:

$$\mathbf{h}_{\mathrm{d}}^t = \mathrm{T}(\mathbf{V}_{\mathrm{d}}^t), \quad \mathbf{h}_{\mathrm{p}}^t = \mathrm{T}(\mathbf{V}_{\mathrm{p}}^t). \tag{2}$$

The current health condition is constructed by concatenating the encoded representations at time t: $\mathbf{e}_c = \left[\mathbf{h}_{\mathrm{d}}^{(t)}, \mathbf{h}_{\mathrm{p}}^{(t)}\right] \in \mathbb{R}^d$. Similarly, the previous health condition at time i (where $i \leq t-1$) is obtained by aggregating the corresponding representations from the patient's prior visit (if available): $\mathbf{e}_p^{(i)} = \left[\mathbf{h}_{\mathrm{d}}^{(i)}, \mathbf{h}_{\mathrm{p}}^{(i)}\right] \in \mathbb{R}^d$. For cold-start scenarios, where the patient has only one visit (i.e., $t = 1$), the model exclusively utilizes the current health condition for downstream tasks.

4.2 Conditional Representation Learning Module

Molecular graph encoding via GIN. We firstly employ a Graph Isomorphism Network (GIN) [Xu *et al.*, 2018] to obtain molecular representations of drugs in the given medication sequence $\mathbf{v}_m^t$. Each drug molecule is treated as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where nodes $v_i \in \mathcal{V}$ represent atoms, and edges $e_{ij} \in \mathcal{E}$ correspond to chemical bonds. The node feature vector $\mathbf{z}_i \in \mathbf{R}^d$ encodes atomic properties, while edge features $\mathbf{e}_{ij} \in \mathbf{R}^d$ capture bond types, detailed summary of the selected features can be found in Appendix. The message-passing process at the $(l+1)$-th layer is defined as follows:

$$\mathbf{h}_i^{(l+1)} = \mathrm{MLP}((1+\alpha^{(l)})\mathbf{h}_i^{(l)} + \sum_{j \in \mathcal{N}(i)} \mathbf{h}_j^{(l)} + \sum_{(i,j) \in \mathcal{E}} \mathbf{e}_{ij}), \tag{3}$$

where $\mathbf{h}_i^{(l)}$ represents the hidden state of node i at layer l, $\mathcal{N}(i)$ denotes the neighboring nodes, and $\alpha^{(l)}$ is a learnable scaling parameter that enhances the expressiveness of the aggregation process. After L layers of message passing, the final node representations are obtained as:

$$\mathbf{h}_i^{(L)} = \mathrm{GIN}_L\left(\mathbf{h}_i^{(0)}, \mathbf{e}_{ij}, \alpha^{(l)}\right). \tag{4}$$

Then we aggregate the learned node features using global sum pooling to derive a holistic molecular representation of the drug: $\mathbf{h}_{\mathrm{m}} = \sum_{i \in \mathcal{V}} \mathbf{h}_i^{(L)} \in \mathbb{R}^d$

Pharma-health contextualized condition construction. After obtaining the structural representation of each drug molecule, we further embed their category labels derived from Anatomical Therapeutic Chemical (ATC) classification system [Schellekens *et al.*, 2011; Garbe *et al.*, 1993]. Specifically, we utilize an embedding layer to transform each drug category label into a dense vector representation $V_m^{(i)} = E_m v_m^{(i)} \in \mathbb{R}^d$. Then we compute an adaptive gating vector:

$$\mathbf{g} = \sigma\left(\mathbf{W}_g\left[V_m^{(i)}, \mathbf{e}_p^{(i)}\right] + \mathbf{b}_g\right) \tag{5}$$

where σ denotes sigmod activation function, $\mathbf{W}_g \in \mathbb{R}^{d \times 2d}$ represents the learnable weight matrix, and $\mathbf{b}_g \in \mathbb{R}^d$ is the bias term. The final gated pharma-health contextualized condition $\mathbf{c}$ which effectively encodes both the categorical pharmacological properties and clinical context of patient's health trajectory is computed as:

$$\mathbf{c} = \mathbf{g} \otimes V_m^{(i)} + (\mathbf{1} - \mathbf{g}) \otimes \mathbf{e}_p^{(i)} \tag{6}$$

where $\otimes$ denotes the Hadamard product and $\mathbf{1}$ represents an all-ones vector.

Therapeutic–safety factorized latent modeling. Given drugs prescribed under certain medical contextualized conditions $\mathbf{c}$, we introduce a latent vector $\mathbf{z} = [\mathbf{z}_{\text{ther}}, \mathbf{z}_{\text{safety}}]$ that factorizes into a therapeutic subspace latent vector $\mathbf{z}_{\text{ther}} \in \mathbb{R}^{d_{\text{ther}}}$ and a safety-related subspace vector $\mathbf{z}_{\text{safety}} \in \mathbb{R}^{d_{\text{safety}}}$. These disentangled latent representations are designed to capture the underlying therapeutic substructures that drive prescriptions and the safety patterns, respectively. We assume that the ideal molecular representations for the given $\mathbf{c}$ follow the conditional distribution

$$p_\theta(\mathbf{h_m} \mid \mathbf{c}) = \int p_\theta(\mathbf{h_m} \mid \mathbf{z}, \mathbf{c})\, p_\theta(\mathbf{z} \mid \mathbf{c})\, d\mathbf{z}, \tag{7}$$

where $\mathbf{h_m}$ is the ideal representation and θ denotes the distribution parameters.

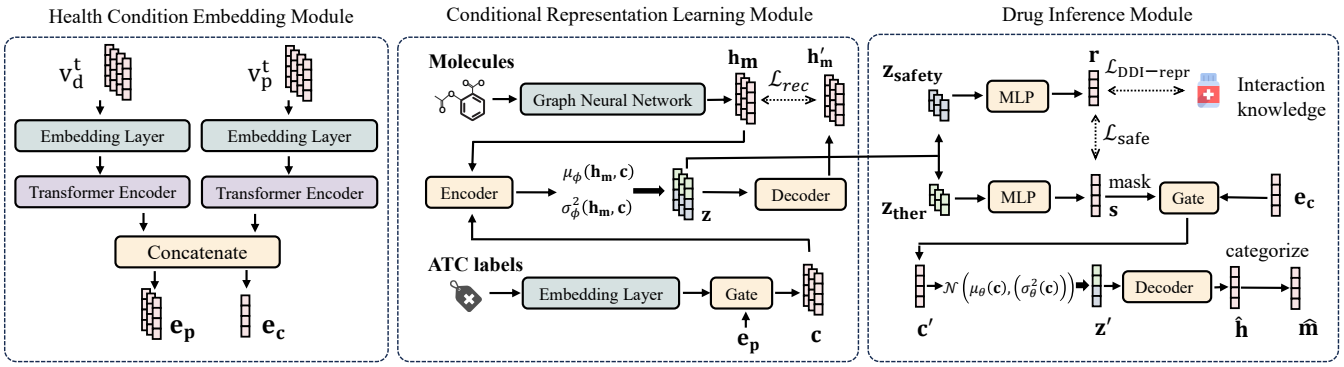

Figure 2: Overview of MGRec. Patient records are encoded into a health representation $\mathbf{e}_c$, while drug molecules are mapped to molecular embeddings $\mathbf{h}_m$. A pharma-health contextualized condition $\mathbf{c}$ conditions a Therapeutic–Safety Factorized CVAE, which infers disentangled latent variables $\mathbf{z}_{\text{ther}}$ and $\mathbf{z}_{\text{safety}}$. Condition-aware molecular representations $\hat{\mathbf{h}}$ are inferred and instantiated as discrete medications via similarity matching, with DDI-guided regularization enforcing safety constraints in the latent space.

As the integral is generally intractable, to optimize $p_\theta(\mathbf{h_m} \mid \mathbf{c})$, we approximate the true posterior $p_\theta(\mathbf{z} \mid \mathbf{h_m}, \mathbf{c})$ with a variational distribution $q_\phi(\mathbf{z} \mid \mathbf{h_m}, \mathbf{c})$. Applying the variational lower bound [Kingma, 2013; Jensen, 1906], we obtain

$$\log p_\theta(\mathbf{h_m} \mid \mathbf{c}) \geq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{h_m},\mathbf{c})}\left[\log p_\theta(\mathbf{h_m} \mid \mathbf{z}, \mathbf{c})\right] - D_{\text{KL}}\left(q_\phi(\mathbf{z} \mid \mathbf{h_m}, \mathbf{c}) \,\|\, p_\theta(\mathbf{z} \mid \mathbf{c})\right). \tag{8}$$

Both the conditional prior and posterior are parameterized as diagonal Gaussians over the concatenated latent vector $\mathbf{z}$:

$$p_\theta(\mathbf{z} \mid \mathbf{c}) = \mathcal{N}\big(\mu_\theta(\mathbf{c}), \text{diag}(\sigma_\theta^2(\mathbf{c}))\big), \tag{9}$$

$$q_\phi(\mathbf{z} \mid \mathbf{h_m}, \mathbf{c}) = \mathcal{N}\big(\mu_\phi(\mathbf{h_m}, \mathbf{c}), \text{diag}(\sigma_\phi^2(\mathbf{h_m}, \mathbf{c}))\big), \tag{10}$$

where $\mu_\theta(\mathbf{c}), \sigma_\theta^2(\mathbf{c}) \in \mathbb{R}^d$ and $\mu_\phi(\mathbf{h_m}, \mathbf{c}), \sigma_\phi^2(\mathbf{h_m}, \mathbf{c}) \in \mathbb{R}^d$ (with $d = d_{\text{ther}} + d_{\text{safety}}$) denote the mean and variance of the prior and posterior, respectively. The conditional prior $p_\theta(\mathbf{z} \mid \mathbf{c})$, parameterized by a multilayer perceptron (MLP) conditioned solely on $\mathbf{c}$, provides a context-dependent latent distribution aligned with the pharma-health context.

The reconstruction term in Eq. (8) encourages the decoded representation to match the prescribed drug, while the KL term aligns the approximate posterior with the condition-aware prior. Let $q_\phi = q_\phi(\mathbf{z} \mid \mathbf{h}_m, \mathbf{c})$ and $p_\theta = p_\theta(\mathbf{z} \mid \mathbf{c})$. The overall objective is:

$$\max_{\theta,\phi} \mathbb{E}_{q_\phi}\left[\log p_\theta(\mathbf{h}_m \mid \mathbf{z}, \mathbf{c})\right] - D_{\text{KL}}(q_\phi \| p_\theta). \tag{11}$$

To parameterize the approximate posterior $q_\phi(\mathbf{z} \mid \mathbf{h_m}, \mathbf{c})$ and the likelihood $p_\theta(\mathbf{h_m} \mid \mathbf{z}, \mathbf{c})$, we employ two MLPs as the conditional encoder and conditional decoder, respectively. The encoder outputs the posterior parameters

$$(\mu_\phi, \sigma_\phi^2) = \text{Encoder}_\phi(\mathbf{h_m}, \mathbf{c}), \tag{12}$$

where $\mu_\phi, \sigma_\phi^2 \in \mathbb{R}^d$. We similarly partition the posterior parameters into therapeutic and safety components, $\mu_\phi = [\mu_\phi^{\text{ther}}; \mu_\phi^{\text{safety}}]$ and $\sigma_\phi = [\sigma_\phi^{\text{ther}}; \sigma_\phi^{\text{safety}}]$, and apply the reparameterization trick [Kingma, 2013]:

$$\mathbf{z}_{\text{ther}} = \mu_\phi^{\text{ther}} + \sigma_\phi^{\text{ther}} \odot \boldsymbol{\epsilon}_{\text{ther}}, \quad \mathbf{z}_{\text{safety}} = \mu_\phi^{\text{safety}} + \sigma_\phi^{\text{safety}} \odot \boldsymbol{\epsilon}_{\text{safety}}, \tag{13}$$

with $\boldsymbol{\epsilon}_{\text{ther}}, \boldsymbol{\epsilon}_{\text{safety}} \sim \mathcal{N}(0, I)$ and $\mathbf{z} = [\mathbf{z}_{\text{ther}}; \mathbf{z}_{\text{safety}}]$.

The conditional decoder reconstructs the pharma-health contextualized molecular representation $\mathbf{h}'_\mathbf{m}$ from the disentangled latent vector $\mathbf{z}$ and condition $\mathbf{c}$:

$$\mathbf{h}'_\mathbf{m} = \text{Decoder}_\theta(\mathbf{z}, \mathbf{c}). \tag{14}$$

The training loss of this module consists of a reconstruction term and a KL regularization term. The reconstruction loss is implemented as the mean squared error (MSE) between $\mathbf{h_m}$ and $\mathbf{h}'_\mathbf{m}$:

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{h_m},\mathbf{c})}\left[\text{MSE}(\mathbf{h_m}, \mathbf{h}'_\mathbf{m})\right]. \tag{15}$$

The KL divergence term aligns the posterior with the condition-aware prior. Let $\mu_{\phi,k}, \sigma_{\phi,k}^2$ and $\mu_{\theta,k}, \sigma_{\theta,k}^2$ denote the k-th elements of the posterior and prior parameters, respectively. Then

$$\begin{aligned}\mathcal{L}_{\text{KL}} &= D_{\text{KL}}\left(q_\phi(\mathbf{z} \mid \mathbf{h_m}, \mathbf{c}) \,\|\, p_\theta(\mathbf{z} \mid \mathbf{c})\right) \\ &= \frac{1}{2}\sum_{k=1}^{d}\left[\log\frac{\sigma_{\theta,k}^2}{\sigma_{\phi,k}^2} + \frac{\sigma_{\phi,k}^2 + \left(\mu_{\phi,k} - \mu_{\theta,k}\right)^2}{\sigma_{\theta,k}^2} - 1\right].\end{aligned} \tag{16}$$

4.3 Drug Inference Module

Therapeutic matching and category filtering. We first determine the necessary drug categories based on the current patient health condition. Specifically, we construct a filtering mask as a constraint on the learned drug latent space.

To achieve this, we integrate all therapeutic latent vectors with the current patient condition by an MLP to obtain a therapeutic matching score for each drug category:

$$\mathbf{s} = \sigma(\text{MLP}([\mathbf{z}_{\text{ther}}, \mathbf{e}_c])), \tag{17}$$

where s represents the matching scores for N drug categories, $\sigma(\cdot)$ is the sigmoid activation function. Subsequently, we apply a threshold $\tau = 0.5$ to construct a filtering mask:

$$\mathbf{mask} = \mathbf{I}(\mathbf{s} > \tau) \tag{18}$$

where $\mathbf{I}(\cdot)$ is the indicator function, returning 1 if $\mathbf{s} > 0.5$ and 0 otherwise, $\mathbf{mask} \in \{0, 1\}^{|M|}$ is the filtering mask applied to the drug categories.

Latent sampling for condition-aware inference. For all drug categories where the mask value equals 1, we utilize their corresponding drug category label embeddings with the patient's current status $\mathbf{e_c}$ to obtain current contextualized conditions $\mathbf{c}'$. Subsequently, we obtain latent vectors from the condition-aware prior and use the decoder to draw latent vectors from

$$\mathbf{z}' = [\mathbf{z}'_{\text{ther}}, \mathbf{z}'_{\text{safety}}] \sim \mathcal{N}\big(\mu_\theta(\mathbf{c}'), \operatorname{diag}(\sigma^2_\theta(\mathbf{c}'))\big). \quad (19)$$

The latent vector is then used to decode a therapeutic molecular representation:

$$\hat{\mathbf{h}} = \text{Decoder}_\theta(\mathbf{z}', \mathbf{c}'), \quad (20)$$

To obtain the final prescribed drug labels, we evaluate the embedding similarity between each generated molecular structure in $\hat{\mathbf{h}}$ and all drugs within its corresponding category. Specifically, we categorize each generated molecule as the most similar drug within its category, thereby obtaining the final recommended drug set $\hat{\mathbf{m}}$:

$$\hat{\mathbf{m}} = \arg \max_{\mathbf{h_m} \in \mathcal{M}_c} \text{Sim}(\hat{\mathbf{h}}, \mathbf{h_m}), \quad (21)$$

where $\mathcal{M}_c$ represents the set of candidate drugs within the category, and $\text{Sim}(\cdot, \cdot)$ denotes the cosine similarity of molecular structural embeddings.

To ensure clinical safety, we introduce a DDI-guided latent regularization mechanism driven by the disentangled safety vector $\mathbf{z}_{\text{safety}}$, which plays two roles: (i) learning a latent representation that reflects the structure of the DDI graph, and (ii) regulating the recommended drug set to mitigate DDIs.

Safety subspace learning via the DDI graph. To ensure that the latent space captures clinically meaningful interaction patterns, we impose a representation-level DDI supervision on $\mathbf{z}_{\text{safety}}$. For each drug pair (i, j), we compute a latent interaction score:

$$r_{ij} = \sigma\Big(f_{\text{DDI}}(\mathbf{z}^{(i)}_{\text{safety}}, \mathbf{z}^{(j)}_{\text{safety}})\Big), \quad (22)$$

where f_{DDI} is a lightweight MLP. Given the DDI adjacency matrix $D \in \mathbb{R}^{|M|\times|M|}$, the safety subspace is trained to reproduce the interaction structure via

$$\mathcal{L}_{\text{DDI-repr}} = -\sum_{i,j} w_{ij} \left[D_{ij} \log r_{ij} + (1 - D_{ij}) \log(1 - r_{ij})\right], \quad (23)$$

where w_{ij} is an optional normalization weight to address class imbalance.

Safety regularization on the recommended set. $\mathcal{L}_{\text{DDI-repr}}$ shapes the safety subspace, but it does not directly constrain the recommended drug set. We therefore introduce a representation-level safety regularizer:

$$\mathcal{L}_{\text{safe}} = \sum_{i<j} s_i s_j \, r_{ij}, \quad (24)$$

where s_i and s_j are the therapeutic matching scores assigned to drugs i and j. This loss discourages the model from simultaneously assigning high confidence to drug pairs that the safety subspace predicts as high-risk, thereby coupling safety awareness with therapeutic relevance recommendation.

Following previous work [Shang *et al.*, 2019; Yang *et al.*, 2021b], we treat the prediction of each drug categorical labels as an independent task and use the BCE loss for optimization:

$$\mathcal{L}_{\text{pred}} = -\sum_{i=1}^{|M|} \left[m_i \log(s_i) + (1 - m_i) \log(1 - s_i)\right], \quad (25)$$

where s_i is the correspond therapeutic matching score as the recommendation probability.

Overall objective. The overall training loss is:

$$\mathcal{L} = \mathcal{L}_{pred} + \alpha \mathcal{L}_{rec} + \beta \mathcal{L}_{KL} + \gamma_1 \mathcal{L}_{\text{DDI-repr}} + \gamma_2 \mathcal{L}_{\text{safe}} \quad (26)$$

where α, β, γ_1, and γ_2 are hyperparameters that respectively control the contributions of reconstruction quality, latent regularization, DDI-aware representation shaping, and safety regularization.

5 Experiments

In this section, we conduct extensive experiments to make a comprehensive evaluation of our proposed method and answer the following four questions:

- **RQ1:** How does the performance of the proposed MGRec compare to that of existing methods?
- **RQ2:** Does MGRec effectively provide reliable recommendations for patients with rare and emerging diseases?
- **RQ3:** How do the different components and hyperparameters of MGRec affect its performance?

5.1 Experiment Setup

Dataset. Electronic health record (EHR) data from two real-world datasets are utilized, specifically MIMIC-III [Johnson *et al.*, 2016] and MIMIC-IV [Johnson *et al.*, 2023]. In line with prior studies [Shang *et al.*, 2019; Yang *et al.*, 2021b], the datasets are processed and randomly split into training, validation, and testing sets with a ratio of 4:1:1. The statistics of the processed datasets are provided in Appendix.

Evaluation Metrics. Four commonly adopted metrics in medication recommendation [Shang *et al.*, 2019; Yang *et al.*, 2021b; Yang *et al.*, 2023] are used as the evaluation metrics: Drug-Drug-Interaction Rate (DDI), Jaccard Similarity Score (Jaccard), F1-score, and Precision-Recall Area Under Curve (PRAUC). Detailed formulation can be found in Appendix.

Baseline. We compare MGRec against ten state-of-the-art medication recommendation models, covering both historical data-driven approaches and molecular substructure-informed approaches. Detailed introduction can be found in Appendix.

Experimental settings. The hyperparameters for all baseline models are selected based on their performance on the validation set. Model optimization is performed using the AdamW optimizer [Loshchilov and Hutter, 2017] with a learning rate of 1e-4 and weight decay of 1e-3. The implementation is based on PyTorch and runs on an NVIDIA Tesla V100 (16GB) GPU. All of the experiment are conducted eight times, and the mean and standard deviation of the results are reported. The hyperparameters could be found in Appendix.

Method	MIMIC-III				MIMIC-IV			
	Jaccard ↑	PRAUC ↑	F1 ↑	DDI ↓	Jaccard ↑	PRAUC ↑	F1 ↑	DDI ↓
LR	$0.4935_{\pm 0.0021}$	$0.7634_{\pm 0.0020}$	$0.6512_{\pm 0.0018}$	$0.0788_{\pm 0.0009}$	$0.4152_{\pm 0.0023}$	$0.6783_{\pm 0.0022}$	$0.5651_{\pm 0.0019}$	$0.0732_{\pm 0.0013}$
LEAP	$0.4521_{\pm 0.0043}$	$0.7081_{\pm 0.0029}$	$0.6152_{\pm 0.0037}$	$0.0720_{\pm 0.0015}$	$0.3909_{\pm 0.0039}$	$0.6542_{\pm 0.0014}$	$0.5439_{\pm 0.0023}$	$0.0550_{\pm 0.0014}$
GAMENet	$0.5210_{\pm 0.0025}$	$0.7780_{\pm 0.0028}$	$0.6762_{\pm 0.0021}$	$0.0781_{\pm 0.0007}$	$0.4401_{\pm 0.0023}$	$0.6833_{\pm 0.0018}$	$0.5933_{\pm 0.0016}$	$0.0718_{\pm 0.0006}$
MICRON	$0.5119_{\pm 0.0029}$	$0.7490_{\pm 0.0023}$	$0.6676_{\pm 0.0027}$	$0.0610_{\pm 0.0009}$	$0.4495_{\pm 0.0031}$	$0.6753_{\pm 0.0034}$	$0.6033_{\pm 0.0023}$	$0.0592_{\pm 0.0007}$
COGNet	$0.5109_{\pm 0.0019}$	$0.7465_{\pm 0.0012}$	$0.6641_{\pm 0.0014}$	$0.0737_{\pm 0.0007}$	$0.4313_{\pm 0.0001}$	$0.6812_{\pm 0.0015}$	$0.5850_{\pm 0.0013}$	$0.0866_{\pm 0.0009}$
SHAPE	$0.5348_{\pm 0.0005}$	$0.7821_{\pm 0.0005}$	$0.6885_{\pm 0.0003}$	$0.0850_{\pm 0.0003}$	$0.4559_{\pm 0.0004}$	$0.6928_{\pm 0.0005}$	$0.6071_{\pm 0.0004}$	$0.0917_{\pm 0.0002}$
RAREMed	$0.5412_{\pm 0.0018}$	$\underline{0.7910}_{\pm 0.0012}$	$0.6938_{\pm 0.0005}$	$\underline{0.0530}_{\pm 0.0006}$	$0.4620_{\pm 0.0016}$	$0.6995_{\pm 0.0011}$	$0.6152_{\pm 0.0014}$	$\underline{0.0510}_{\pm 0.0006}$
SafeDrug	$0.5255_{\pm 0.0030}$	$0.7732_{\pm 0.0028}$	$0.6804_{\pm 0.0027}$	$0.0688_{\pm 0.0005}$	$0.4560_{\pm 0.0033}$	$0.6858_{\pm 0.0030}$	$0.6098_{\pm 0.0029}$	$0.0689_{\pm 0.0007}$
MoleRec	$0.5303_{\pm 0.0032}$	$0.7795_{\pm 0.0030}$	$0.6844_{\pm 0.0026}$	$0.0692_{\pm 0.0008}$	$0.4502_{\pm 0.0035}$	$0.6867_{\pm 0.0031}$	$0.6040_{\pm 0.0033}$	$0.0699_{\pm 0.0007}$
DrugDoctor	$\underline{0.5420}_{\pm 0.0014}$	$0.7903_{\pm 0.0015}$	$\underline{0.6955}_{\pm 0.0011}$	$0.0603_{\pm 0.0003}$	$\underline{0.4703}_{\pm 0.0016}$	$\underline{0.7008}_{\pm 0.0015}$	$\underline{0.6202}_{\pm 0.0015}$	$0.0705_{\pm 0.0007}$
MGRec	$\mathbf{0.5521}_{\pm 0.0010}$	$\mathbf{0.7940}_{\pm 0.0014}$	$\mathbf{0.7004}_{\pm 0.0009}$	$\mathbf{0.0404}_{\pm 0.0002}$	$\mathbf{0.4872}_{\pm 0.0013}$	$\mathbf{0.7145}_{\pm 0.0009}$	$\mathbf{0.6308}_{\pm 0.0012}$	$\mathbf{0.0425}_{\pm 0.0002}$

Table 1: Performance comparison of different methods on MIMIC-III and MIMIC-IV datasets. The best results are highlighted in **bold**, and the second-best results are underlined.

5.2 Overall Performance Comparison (RQ1)

Table 1 summarizes the overall performance of all methods. MGRec achieves the best Jaccard/F1/PRAUC and the lowest DDI on both datasets. Compared with history-driven and structure-informed baselines, MGRec better balances effectiveness and safety by inferring condition-aware molecular representations and incorporating DDI supervision in the latent safety subspace. These gains persist across MIMIC-III and MIMIC-IV, indicating robust generalization under sparse supervision.

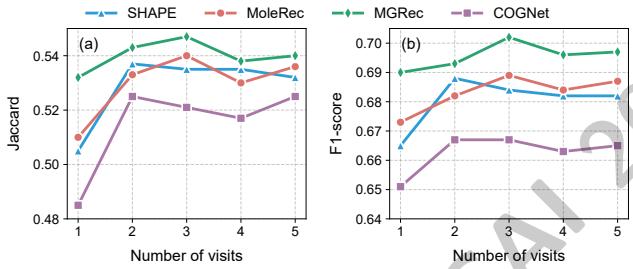

Figure 3: The performances of four models on the patients with specific visits in the MIMIC-III dataset. (a) Jaccard, (b) F1 score.

Figure 3 illustrates the recommendation accuracy of four models across patients with varying numbers of visits in the MIMIC-III dataset. MGRec consistently outperforms the baselines in overall recommendation accuracy. Notably, even for patients with only one visit, where data sparsity poses a significant challenge, MGRec maintains a comparable level of accuracy. This demonstrates MGRec's strong generalization ability to provide reliable recommendations even in data-sparse scenarios.

5.3 Case Study (RQ2)

To intuitively demonstrate the advantage of MGRec in providing reliable medication recommendations for patients with emerging and rare diseases, we sample several example cases and conduct a detailed analysis of their recommendation results. Table 2 presents two representative patients from the MIMIC-III dataset: Patient E with an emerging disease and Patient R with a rare condition.

Category	**Patient E** Jaccard: 0.5432 DDI: 0
Diagnosis	**Postconcussion syndrome**, Cerebral artery occlusion, Subarachnoid hemorrhage, Acute kidney failure, Retention of urine, Hypertensive chronic kidney disease.
Procedure	Insertion of indwelling urinary catheter, Venous catheterization.
Medication	Neomycin, Cefotaxime, Chlorhexidine, Heparin, Glyceryl trinitrate, Sultiame, Amiodarone, Potassium chloride, Furosemide.
Category	**Patient R** Jaccard: 0.5556 DDI: 0
Diagnosis	**Poisoning by aromatic analgesics**, Acute necrosis of liver, Coagulation defects, Acute respiratory failure, Hypernatremia.
Procedure	Venous cutdown, Exchange transfusion, Mechanical ventilation.
Medication	Potassium chloride, Glucose, Magnesium carbonate, Chlorhexidine, Heparin, Omeprazole, Acetylcysteine, Morphine, Cyproheptadine, Bisacodyl, Ketamine.

Table 2: Recommended results for two patients.

Patient E suffers from **postconcussion syndrome**, a condition that appears **only once** in the dataset. To address this unique case, MGRec specifically recommends Sultiame, an anticonvulsant and antiepileptic drug [Ben-Zeev *et al.*, 2004], to manage postconcussion syndrome-related symptoms [Ryan and Warden, 2003; Gorman and Shahwan, 2016; Bresnahan *et al.*, 2019]. Additionally, the model recommends Furosemide to alleviate retention of urine [Brater, 1998], demonstrating its ability to provide clinically relevant treatment strategies even for unseen conditions.

Patient R has been diagnosed with **poisoning by aromatic analgesics**, a condition with an occurrence rate of only **0.004%** in the dataset. To treat this patient, MGRec recommends Acetylcysteine to eplenish glutathione levels and prevents liver damage [Heard, 2008], as well as Heparin, an anticoagulant, to manage coagulation defects [Oduah *et al.*, 2016]. These recommendations align with clinical practices for poisoning cases, showcasing MGRec's capability to make precise and condition-aware medication suggestions even when handling extremely low-frequency diseases.

5.4 Ablation Study (RQ3)

To examine the contribution of each key component in MGRec, we design the following ablation variants. For clarity, we use concise abbreviations to denote each removed component.

- **w/o $\mathbf{D_r}$** removes the DDI-guided latent regularization, including both representation-level DDI supervision and set-level safety regularization, disabling explicit safety modeling in the latent space.
- **w/o $\mathbf{F_z}$** removes the therapeutic–safety factorization by replacing the disentangled latent representation with a single unified latent variable.
- **w/o $\mathbf{T_e}$** replaces the Transformer-based temporal encoder in the health condition embedding module with a simple embedding layer.
- **w/o $\mathbf{G_e}$** replaces the GIN-based molecular graph encoder with a drug embedding layer, removing explicit molecular structural reasoning.

Model	Jaccard	PRAUC	F1	DDI
w/o $\mathbf{D_r}$	0.5461	0.7924	0.6976	0.0548
w/o $\mathbf{F_z}$	0.5450	0.7910	0.6969	0.0517
w/o $\mathbf{T_e}$	0.5449	0.7905	0.6962	0.0423
w/o $\mathbf{G_e}$	0.5414	0.7882	0.6923	0.0469
MGRec	**0.5521**	**0.7940**	**0.7004**	**0.0404**

Table 3: Ablation study on the MIMIC-III dataset.

Table 3 reports the ablation results on the MIMIC-III dataset. Removing any component leads to consistent performance degradation, confirming the necessity of each design choice. In particular, **w/o $\mathbf{D_r}$** results in a substantial increase in DDI rate, demonstrating the importance of incorporating pharmacological interaction knowledge directly into the latent space. The variant **w/o $\mathbf{F_z}$** exhibits both reduced accuracy and increased DDI risk, indicating that explicitly disentangling therapeutic relevance from safety-sensitive factors enables better control over the effectiveness–safety trade-off.

Moreover, removing the Transformer encoder (**w/o $\mathbf{T_e}$**) or the molecular graph encoder (**w/o $\mathbf{G_e}$**) degrades predictive performance, highlighting the importance of modeling longitudinal patient conditions and molecular structural information, respectively. Notably, the performance drop caused by **w/o $\mathbf{G_e}$** is also accompanied by a higher DDI rate, underscoring the role of molecular structure in mitigating adverse drug interactions.

5.5 Hyperparameter Study (RQ3)

To understand the influence of hyperparameters on the efficacy of MGRec, we considered four hyperparameters: the reconstruction weight (α), KL regularization weight β, representation-level DDI regularization weight γ_1 and generation-level safety regularization weight γ_2.

Fig. 4 illustrates the impact of key hyperparameters on MGRec. As shown in Fig. 4(a), the reconstruction weight

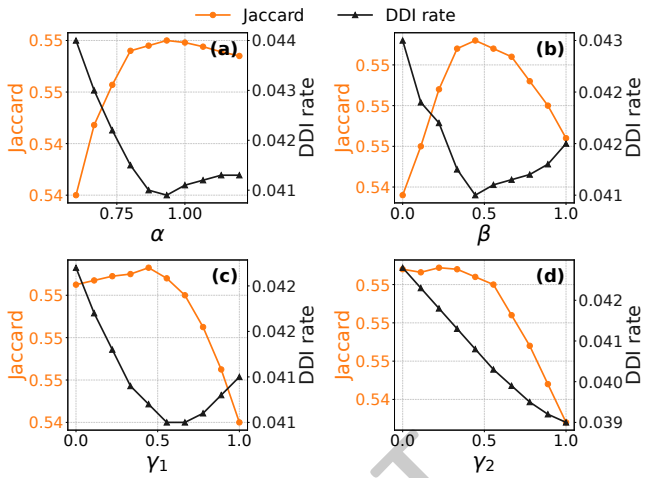

Figure 4: Hyperparameter test on the MIMIC-III dataset. (a) Reconstruction weight, (b) KL regularization weight, (c) Representation-level DDI regularization weight, (d) Generation-level safety regularization weight.

α controls the trade-off between preserving molecular structure fidelity and learning abstract therapeutic representations, where moderate values yield the best accuracy and safety balance. Fig. 4(b) shows that the KL regularization weight β stabilizes the condition-aware latent space and improves generalization under sparse supervision, while overly large values lead to latent collapse and reduced performance.

Fig. 4(c) and Fig. 4(d) examine the two DDI-related weights γ_1 and γ_2. Increasing γ_1 strengthens the alignment between the safety subspace and the DDI graph, effectively reducing interaction risk, whereas excessive emphasis may interfere with therapeutic representation learning. Similarly, γ_2 directly penalizes high-risk co-recommendations at the set level, leading to lower DDI rates but potentially conservative recommendations when overly large. All hyperparameters are selected based on validation performance to balance effectiveness and safety.

6 Conclusion

In this work, we present MGRec, a structure-grounded medication recommendation framework that reframes the learning objective from discrete prescription prediction to condition-aware inference in a continuous molecular representation space. Rather than relying on observed co-occurrence patterns, MGRec models the pharmacological characteristics that plausible medications should possess given a patient's clinical context, enabling more robust generalization under sparse supervision. The proposed Therapeutic–Safety Factorized CVAE disentangles therapeutic relevance from safety-sensitive molecular factors in a unified latent space, while DDI-guided latent regularization embeds pharmacological interaction constraints directly into the representation learning process. Extensive experiments on two real-world EHR benchmarks demonstrate that MGRec consistently improves recommendation accuracy while reducing drug–drug interaction risk, particularly in data-sparse scenarios.

Acknowledgments

This paper is partially supported by the National Natural Science Foundation of China (No.62502491; No.12227901). The AI-driven experiments, simulations and model training were performed on the robotic AI-Scientist platform of Chinese Academy of Sciences.Anhui Science Foundation for Distinguished Young Scholars (No.1908085J24).

References

[Agrawal and Prabakaran, 2020] Raag Agrawal and Sudhakaran Prabakaran. Big data in digital healthcare: lessons learnt and recommendations for general practice. *Heredity*, 124(4):525–534, 2020.

[Ben-Zeev *et al.*, 2004] Bruria Ben-Zeev, Nathan Watemberg, Pinchas Lerman, Itshak Barash, Nathan Brand, and Tally Lerman-Sagie. Sulthiame in childhood epilepsy. *Pediatrics international*, 46(5):521–524, 2004.

[Brater, 1998] D Craig Brater. Diuretic therapy. *New England Journal of Medicine*, 339(6):387–395, 1998.

[Bresnahan *et al.*, 2019] Rebecca Bresnahan, Kirsty J Martin-McGill, Philip Milburn-McNulty, Graham Powell, Graeme J Sills, and Anthony G Marson. Sulthiame add-on therapy for epilepsy. *Cochrane Database of Systematic Reviews*, (8), 2019.

[Choi *et al.*, 2016] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in neural information processing systems*, 29, 2016.

[Dagliati *et al.*, 2021] Arianna Dagliati, Alberto Malovini, Valentina Tibollo, and Riccardo Bellazzi. Health informatics and ehr to support clinical research in the covid-19 pandemic: an overview. *Briefings in bioinformatics*, 22(2):812–822, 2021.

[Esteva *et al.*, 2019] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24–29, 2019.

[Faviez *et al.*, 2020] Carole Faviez, Xiaoyi Chen, Nicolas Garcelon, Antoine Neuraz, Bertrand Knebelmann, Rémi Salomon, Stanislas Lyonnet, Sophie Saunier, and Anita Burgun. Diagnosis support systems for rare diseases: a scoping review. *Orphanet journal of rare diseases*, 15(1):94, 2020.

[Garbe *et al.*, 1993] E Garbe, J Röhmel, and U Gundert-Remy. Clinical and statistical issues in therapeutic equivalence trials. *European journal of clinical pharmacology*, 45:1–7, 1993.

[Garriga *et al.*, 2022] Roger Garriga, Javier Mas, Semhar Abraha, Jon Nolan, Oliver Harrison, George Tadros, and Aleksandar Matic. Machine learning model to predict mental health crises from electronic health records. *Nature medicine*, 28(6):1240–1248, 2022.

[Gómez-Bombarelli *et al.*, 2018] Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.

[Gorman and Shahwan, 2016] Kathleen M Gorman and Amre Shahwan. Sultiame revisited: treatment of refractory absence seizures. *Epileptic Disorders*, 18(3):329–333, 2016.

[Heard, 2008] Kennon J Heard. Acetylcysteine for acetaminophen poisoning. *New England Journal of Medicine*, 359(3):285–292, 2008.

[Jensen, 1906] Johan Ludwig William Valdemar Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta mathematica*, 30(1):175–193, 1906.

[Jin *et al.*, 2018] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation. In *International conference on machine learning*, pages 2323–2332. PMLR, 2018.

[Johnson *et al.*, 2016] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

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

[Jones *et al.*, 2008] Kate E Jones, Nikkita G Patel, Marc A Levy, Adam Storeygard, Deborah Balk, John L Gittleman, and Peter Daszak. Global trends in emerging infectious diseases. *Nature*, 451(7181):990–993, 2008.

[Kingma, 2013] Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[Kou *et al.*, 2024] Zhiqiang Kou, Jing Wang, Jiawei Tang, Yuheng Jia, Boyu Shi, and Xin Geng. Exploiting multi-label correlation in label distribution learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 4326–4334, 2024.

[Kou *et al.*, 2025] Zhiqiang Kou, Si Qin, Hailin Wang, Jing Wang, Mingkun Xie, Shuo Chen, Yuheng Jia, Tongliang Liu, Masashi Sugiyama, and Xin Geng. Label distribution learning with biased annotations assisted by multi-label learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 8 2025.

[Kuang and Xie, 2024] Yabin Kuang and Minzhu Xie. Drugdoctor: enhancing drug recommendation in cold-start scenario via visit-level representation learning and training. *Briefings in Bioinformatics*, 25(6):bbae464, 2024.

[Liu *et al.*, 2023] Sicen Liu, Xiaolong Wang, Jingcheng Du, Yongshuai Hou, Xianbing Zhao, Hui Xu, Hui Wang, Yang Xiang, and Buzhou Tang. Shape: A sample-adaptive hierarchical prediction network for medication recommendation. *IEEE Journal of Biomedical and Health Informatics*, 2023.

[Loshchilov and Hutter, 2017] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[Ochiai *et al.*, 2023] Toshiki Ochiai, Tensei Inukai, Manato Akiyama, Kairi Furui, Masahito Ohue, Nobuaki Matsumori, Shinsuke Inuki, Motonari Uesugi, Toshiaki Sunazuka, Kazuya Kikuchi, et al. Variational autoencoder-based chemical latent space for large molecular structures with 3d complexity. *Communications Chemistry*, 6(1):249, 2023.

[Oduah *et al.*, 2016] Eziafa I Oduah, Robert J Linhardt, and Susan T Sharfstein. Heparin: past, present, and future. *Pharmaceuticals*, 9(3):38, 2016.

[Pang *et al.*, 2023] Chao Pang, Jianbo Qiao, Xiangxiang Zeng, Quan Zou, and Leyi Wei. Deep generative models in de novo drug molecule generation. *Journal of Chemical Information and Modeling*, 64(7):2174–2194, 2023.

[Ryan and Warden, 2003] Laurie M Ryan and Deborah L Warden. Post concussion syndrome. *International review of psychiatry*, 15(4):310–316, 2003.

[Samanta *et al.*, 2020] Soumitra Samanta, Steve O'Hagan, Neil Swainston, Timothy J Roberts, and Douglas B Kell. Vae-sim: a novel molecular similarity measure based on a variational autoencoder. *Molecules*, 25(15):3446, 2020.

[Schellekens *et al.*, 2011] Huub Schellekens, Ety Klinger, Stefan Mühlebach, Jean-Francois Brin, Gert Storm, and Daan JA Crommelin. The therapeutic equivalence of complex drugs. *Regulatory Toxicology and Pharmacology*, 59(1):176–183, 2011.

[Schieppati *et al.*, 2008] Arrigo Schieppati, Jan-Inge Henter, Erica Daina, and Anita Aperia. Why rare diseases are an important medical and social issue. *The Lancet*, 371(9629):2039–2041, 2008.

[Senbekov *et al.*, 2020] Maksut Senbekov, Timur Saliev, Zhanar Bukeyeva, Aigul Almabayeva, Marina Zhanaliyeva, Nazym Aitenova, Yerzhan Toishibekov, and Ildar Fakhradiyev. The recent progress and applications of digital technologies in healthcare: a review. *International journal of telemedicine and applications*, 2020(1):8830200, 2020.

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

[Shi and others, 2025] Jiangming Shi et al. Mipt: Multilevel informed prompt tuning for robust molecular property prediction. In *Forty-second International Conference on Machine Learning*, 2025.

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

[Wigh *et al.*, 2022] Daniel S Wigh, Jonathan M Goodman, and Alexei A Lapkin. A review of molecular representation in the age of machine learning. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(5):e1603, 2022.

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

[Xu *et al.*, 2018] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

[Xue *et al.*, 2019] Dongyu Xue, Yukang Gong, Zhaoyi Yang, Guohui Chuai, Sheng Qu, Aizong Shen, Jing Yu, and Qi Liu. Advances and challenges in deep generative models for de novo molecule generation. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 9(3):e1395, 2019.

[Yang *et al.*, 2021a] Chaoqi Yang, Cao Xiao, Lucas Glass, and Jimeng Sun. Change matters: Medication change prediction with recurrent residual networks. *arXiv preprint arXiv:2105.01876*, 2021.

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

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

[Zeng *et al.*, 2022] Xiangxiang Zeng, Fei Wang, Yuan Luo, Seung-gu Kang, Jian Tang, Felice C Lightstone, Evandro F Fang, Wendy Cornell, Ruth Nussinov, and Feixiong Cheng. Deep generative molecular design reshapes drug discovery. *Cell Reports Medicine*, 3(12), 2022.

[Zhang *et al.*, 2017] Yutao Zhang, Robert Chen, Jie Tang, Walter F Stewart, and Jimeng Sun. Leap: learning to prescribe effective and safe treatment combinations for multimorbidity. In *proceedings of the 23rd ACM SIGKDD international conference on knowledge Discovery and data Mining*, pages 1315–1324, 2017.

[Zhao *et al.*, 2024] Zihao Zhao, Yi Jing, Fuli Feng, Jiancan Wu, Chongming Gao, and Xiangnan He. Leave no patient behind: Enhancing medication recommendation for rare disease patients. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 533–542, 2024.