Highlights

CIDGMed: Causal Inference-Driven Medication Recommendation with Enhanced Dual-Granularity Learning

Shunpan Liang,Xiang Li,Shi Mu,Chen Li,Yu Lei,Yulei Hou,Tengfei Ma

- We use causal inference to uncover the true causal relationships between diseases/procedures and medications, reducing spurious correlations, thereby minimizing recommendation bias and enhancing the interpretability of recommendations.

- We consider both coarse-grained information at the entity level and fine-grained information at the molecular level, leveraging the collaboration between this dual-granularity information to enhance the model's learning capacity.

- We use a bias correction module as a post-processing intervention method during the model recommendation phase to refine the initial results and further optimize the model's predictive performance.

- Extensive experiments demonstrate that CIDGMed significantly outperforms current state-of-the-art models across multiple metrics, achieving a 2.54% increase in accuracy, a 3.65% reduction in side effects, and a 39.42% improvement in time efficiency.

- An intricate and elaborate case study based on a real medical record is designed to elucidate the effectiveness and rationality of the model based on causal inference.

CIDGMed: Causal Inference-Driven Medication Recommendation with Enhanced Dual-Granularity Learning

Shunpan Liang[a,b], Xiang Li[a,c], Shi Mu[a], Chen Li[a,*], Yu Lei[a], Yulei Hou[a] and Tengfei Ma[d]

[a]*Yanshan University, QinHuangDao, 066004, China*

[b]*Xinjiang College of Science & Technology, Korla, 841000, China*

[c]*Peking University, Beijing, 100091, China*

[d]*Hunan University, ChangSha, 410012, China*

ARTICLE INFO

Keywords:
Medication Recommendation
Intelligent Healthcare Management
Causal Inference
Recommender Systems

ABSTRACT

Medication recommendation aims to integrate patients' long-term health records to provide accurate and safe medication combinations for specific health states. Existing methods often fail to deeply explore the true causal relationships between diseases/procedures and medications, resulting in biased recommendations. Additionally, in medication representation learning, the relationships between information at different granularities of medications—coarse-grained (medication itself) and fine-grained (molecular level)—are not effectively integrated, leading to biases in representation learning. To address these limitations, we propose the Causal Inference-driven Dual-Granularity Medication Recommendation method (CIDGMed). Our approach leverages causal inference to uncover the relationships between diseases/procedures and medications, thereby enhancing the rationality and interpretability of recommendations. By integrating coarse-grained medication effects with fine-grained molecular structure information, CIDGMed provides a comprehensive representation of medications. Additionally, we employ a bias correction model during the prediction phase to further refine recommendations, ensuring both accuracy and safety. Through extensive experiments, CIDGMed significantly outperforms current state-of-the-art models across multiple metrics, achieving a 2.54% increase in accuracy, a 3.65% reduction in side effects, and a 39.42% improvement in time efficiency. Additionally, we demonstrate the rationale of CIDGMed through a case study.

1. Introduction

As society advances and the population grows, the healthcare system is facing unprecedented pressures, including the imbalance of medical resources and the complexity of disease diagnosis and treatment, which have become prominent issues in the medical field. In this context, artificial intelligence (AI)–based medication recommendations [51, 3, 21] emerge as a key component, bringing new hope to healthcare. Compared to traditional manual methods, this system, based on extensive patient data, can make rapid responses and decisions, significantly addressing the issue of scarce medical resources. Furthermore, by considering patients' individual characteristics, medical history, and current health status [1], the medication recommendation system can provide accurate and safe medication combinations, playing a key role in alleviating medical pressures and enhancing treatment outcomes.

Compared to recommendation systems (RSs) in other domains, medication recommendation faces unique and significant challenges that prevent the direct application of methodologies from other RSs [26, 45, 32, 4]. These challenges primarily lie in the following two aspects. Firstly,

medication recommendation involves the complex relationships among various medical concepts (medications, diseases, procedures) and the extensive chemical and molecular structure knowledge inherent in medications. Secondly, when evaluating recommendation performance, medication recommendation not only focuses on accuracy but also places a high emphasis on safety concerns. This is because drug-drug [2] interactions (DDIs) [2, 25, 6] can cause side effects in patients.

Early medication RSs [40, 1, 43] primarily focused on analyzing patients' current health status, while overlooking their long-term medical histories. Subsequent researches [20, 16, 59] began to recognize the importance of longitudinal medical records and modelled patients' historical records based on time series analysis. Although these studies improved the accuracy of medication recommendations to some extent, they neglected the safety of medication recommendations, which is crucial in this field. Consequently, some studies [27, 5] have considered drug-drug interactions as the main factor affecting safety and have focused on reducing the impact of these interactions to enhance the safety of medication recommendation systems. Despite the significant achievements in existing research, there remain considerable biases in the recommendation results, primarily stemming from the following aspects:

(1) **Unclear Relationships between Disease/Procedure and Medication.**

---

*Corresponding author.

✉ liangshunpan@ysu.edu.cn (S. Liang); lilixiang222@gmail.com (X. Li); mushi@stumail.ysu.edu.cn (S. Mu); lichen36211@gmail.com (C. Li); leiyu0160@gmail.com (Y. Lei); ylhou@ysu.edu.cn (Y. Hou); tfma@hnu.edu.cn (T. Ma)

[1]Patients' health status encompasses the health status of individuals, including diseases and procedures in healthcare encounters.

[2]The terms "drug" and "medication" are used interchangeably in this paper.
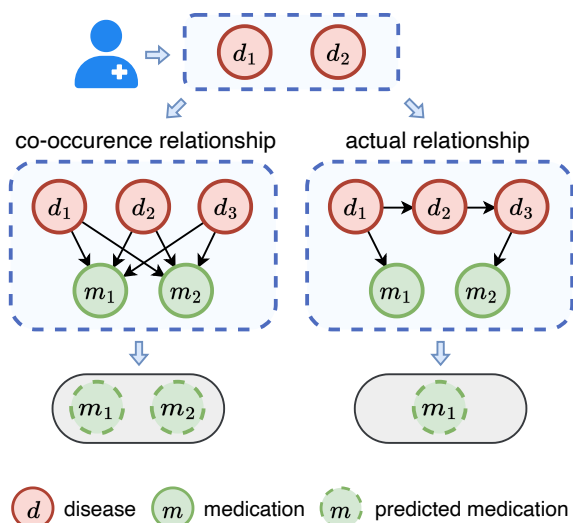
Figure 1: A case where a method based on co-occurrence relationships leads to erroneous recommendation results.

Previous co-occurrence-based researches [23, 55, 54] assumed that high co-occurrence implies a direct relationship or influence between diseases and medications. However, this is not always the case. As shown in Figure 1, the results learned using the co-occurrence method indicate a correlation between each disease (d_1, d_2, d_3) and each medication (m_1, m_2). In reality, the correct relationships should be that d_1 causes d_2, which in turn causes d_3, while m_1 specifically targets d_1, and m_2 specifically targets d_3. When models encounter a new combination of diseases (d_1, d_2), the co-occurrence method mistakenly recommends both m_1 and m_2, resulting in inaccurate and potentially harmful outcomes.

Hence, these studies [23, 55, 54] addressing multi-disease issues relied solely on co-occurrence to determine the relationships between diseases and medications, failing to identify the clear relationships between them. This bias can be amplified through feedback, affecting the accuracy and safety of the recommendations.

(2) **Unintegrated Dual-Granularity Information.** In medication recommendation research, there are two main approaches: one is to directly learn the coarse-grained representation of medications from data, and the other, given that most medications consist of multiple molecules [3], maps medications to individual molecules and treats them as substitutes for the medication itself, emphasizing the importance of the molecular (fine-grained) level. In particular, using only coarse-grained information [46, 21] fails to capture the chemical and molecular structure information, making it difficult to identify medications with similar compound structures and to analyze medication-medication interactions. On the other hand, using only fine-grained information [51, 52], although it can provide detailed chemical information, the ultimate task of medication recommendation is to recommend the entire medication, not just a single

molecule, and this can introduce noise into the training process, affecting the stability and robustness of the model.

However, current methods typically focus only on either the coarse-grained level or the fine-grained level of medications, failing to effectively integrate information from both granularities. This leads to biases in representation learning and recommendation results.

Therefore, to address the above limitations, we design a **C**ausal **I**nference-driven **D**ual-**G**ranularity **Med**ication Recommendation method named **CIDGMed**. The main contributions of our method are summarized as follows:

- We introduce causal inference techniques that can uncover the causal relationships between diseases/procedures and medications, enhancing the rationality and interpretability of medication recommendations.

- A dual-granularity fusion method is proposed. At the coarse-grained level, it learns the effects of entire medications, while at the fine-grained level, it mines richer relationships from molecular structures.

- We employ a post-processing intervention method by introducing a bias correction model. This model further adjusts the generated recommendations using the established causal relationships during the prediction phase, thereby enhancing the accuracy and safety of the model.

- Extensive experiments on real-world datasets demonstrate superior performance compared to existing state-of-the-art models. Our approach achieves a 2.54% increase in accuracy and a 3.65% enhancement in safety. Additionally, we demonstrate the interpretability of CIDGMed through a case study. Our code is publicly available at GitHub[4].

We provide an overview of the different sections of the paper: (1) Introduction: Introduces the main innovations and the motivation behind this work. (2) Related Work: Summarizes typical studies and current trends in medication recommendations and causal inference in RSs. (3) Problem Definition: Clearly explains the inputs and outputs and the special terms mentioned in this paper. (4) Methods: Presents the core ideas of the model and the specific technical details of its implementation. (5) Experiments: Introduces the experimental background and results, and presents a series of supportive experiments to provide an in-depth analysis of the results. A specific case study of the workflow in this paper can be found in Subsection 5.8. (6) Conclusion: Summarizes the research findings and offers perspectives on future research directions.

2. Related Works

In this section, we describe related work in the areas of medication recommendation and causal inference in RSs.

---

[3] https://atcddd.fhi.no/atc/structure_and_principles/

[4] https://github.com/lixiang-222/CIDGMed

2.1. Medication Recommendation

In recent years, the field of medication recommendation has developed rapidly, and mainstream medication recommendation models can be broadly categorized into three main directions:

The first category mainly focuses on the technical aspect of medication recommendation. The earliest methods [40, 1, 43] use statistical data approaches. With the development of collaborative filtering methods, Medicine-LDA [39] proposed a context-aware collaborative model based on Latent Dirichlet Allocation (LDA) to integrate multiple types of information. CompNet [24] addresses recommendation issues by framing the task as an order-free Markov decision process (MDP) problem. In recent years, deep learning has achieved tremendous success, leading to the emergence of medication recommendation methods using this technology. For example, RETAIN [8] is based on a two-level neural attention model that can detect influential past visits and significant information within those visits. The study [17] develops three different LSTMs to model heterogeneous data interactions for predicting next-period prescriptions. GAMENet [35] leverages Graph Neural Networks (GNN) to enhance medication recommendation systems. Similarly, COGNet [46] utilizes the Transformer architecture for medication recommendations, adopting a translation approach to infer medications from diseases/procedures.

The second category of research focuses more on the relationships between diseases/procedures and medications. It constructs relational networks to explore the potential connections among various medical entities. For example, DPR [60] classifies the interactions between different medications and represents them as graphs, designing a medication package recommendation framework based on graph neural networks to integrate medication interaction information. DPG [61] considers the interaction effects between medications affected by patient conditions, initializes the medication interaction graph based on medical records and domain knowledge, and uses an RNN to capture medication interactions. StratMed [21] utilizes a dual-property network to address the mutual constraints between the safety and accuracy of medication combinations, synergistically enhancing these two properties.

The third category of research underscores the importance of incorporating domain-specific knowledge to enhance the relationships among various medical entities within the dataset. By integrating more expert knowledge beyond the dataset, these approaches aim to improve the accuracy and safety of medication recommendations. For instance, SafeDrug [51] introduces detailed molecular information to characterize the relationships between medications. This approach aims to reduce the risk of adverse drug-drug interactions (DDI) and provide safer medication combinations. By focusing on molecular details, SafeDrug enhances our understanding of how different medications interact at a molecular level, contributing to safer prescription practices. Similarly, MoleRec [52] improves the relationships between medications by simulating specific molecular substructures within the medications. This method allows for a more accurate and precise representation of DDIs, enabling more reliable predictions of potential interactions. MoleRec's approach highlights the importance of detailed molecular simulation in understanding medication interactions. And Carmen [3] incorporates patient records into the molecular representation learning process to enhance the ability to differentiate between molecular differences.

Despite the success of the aforementioned methods, they have certain limitations: (1) they primarily rely on the co-occurrence relationships among various medical entities in patient historical records, failing to uncover the underlying causal relationships; (2) they do not consider the use of both coarse-grained medication information and fine-grained molecular information for collaborative supplementary learning and recommendation, which could significantly enhance the accuracy, safety, and interpretability of medication recommendations.

2.2. Causal Inference in RSs

In the real world, causality drives the system, prompting researchers in recommender systems to leverage causal inference to extract causal relationships and thereby enhance the RS. Formally, causality can be defined as the relationship between cause and effect, where the cause is partly responsible for the effect [44]. Causal inference, on the other hand, refers to the process of identifying and utilizing causal relationships based on experimental or observational data [53]. Causal relationships are primarily applied in three key stages of RSs:

(1) Data preprocessing: In recent years, some methods [11, 57, 10] have utilized causal relationships to address issues such as data debiasing or data augmentation. Specifically, causal theory empowers us to pinpoint the underlying cause of data bias by thoroughly examining the generation process of recommendation data. Hence, through causal inference, some studies can effectively mitigate the impact of bias. (2) Model development: Causal inference contributes to enhancing the interpretability of the model itself [36, 42, 37, 58, 56, 48, 41]. (3) Recommendation outcomes: This is a postprocessing mechanism that incorporates causal relationships into the obtained recommendation results to correct bias issues in the model learning process, thereby enhancing their accuracy and interpretability [11, 47, 38].

Building on the powerful inferential capabilities of causal inference, this paper innovatively introduces it into the field of medication recommendation. Specifically, we employ a causal inference-based approach to mine potential links between medications and diseases/procedures, eliminating spurious correlations arising from co-occurring relationships and providing more precise causal relationships. Additionally, during the recommendation phase of the model, we leverage the learned causal relationships to correct model biases, thereby enhancing the model's performance and interpretability.

3. Problem Definition

3.1. Medication Combination Recommendation

Medical Entity. Medical entities, in this article, include diseases, procedures, and medications, and these are denoted by the symbols $\mathcal{D} = \{d_1, d_2, \ldots\}$, $\mathcal{P} = \{p_1, p_2, \ldots\}$, and $\mathcal{M} = \{m_1, m_2, \ldots\}$ respectively.

Input and Output. This paper uses Electronic Health Records (EHR) as its data source [5], covering a wide range of patient visits and treatment records. Each patient's record is denoted as $\mathcal{H}$, containing longitudinal visit records $\mathcal{H} = \{v_1, v_2, \ldots, v_t\}$, where v_t is the clinical information associated with the t_{th} visit. Each visit record $v_t = \{\mathcal{D}_t, \mathcal{P}_t, \mathcal{M}_t\}$ includes three elements: $\mathcal{D}_t \subset \mathcal{D}$, $\mathcal{P}_t \subset \mathcal{P}$, $\mathcal{M}_t \subset \mathcal{M}$, representing the patient's historical and current visit's medical data. If it is the first visit, prior records will be absent. These elements are encoded using a multi-hot encoding technique. Our model's output, denoted as $\hat{\mathcal{M}}_t$, predicts the medication combination for the visit v_t.

3.2. Safety of Medication Combination

Safety is particularly important in medication recommendation, as certain medication combinations can pose significant health risks. To ensure the safety of these combinations, it is necessary to incorporate DDI-related knowledge and minimize the occurrence of DDIs as much as possible. Inspired by SafeDrug [51], DDI information is extracted from the Adverse Event Reporting System [14]. In our model, DDI information is represented in a matrix format $\mathbf{M}^{ddi} \in \{0,1\}^{|\mathcal{M}| \times |\mathcal{M}|}$, where $\mathbf{M}^{ddi}_{m_i m_j} = 1$ indicates an interaction between m_i and m_j. A higher frequency of DDI indicates a greater potential for safety issues in the recommended results.

3.3. Causal Discovery and Inference

As discussed in Section 2.2, causal discovery and inference are crucial for understanding the interactions between entities [12], helping to estimate genuine causal effects and thereby enhancing interpretability, particularly in complex data scenarios. Therefore, causal discovery and inference are crucial for identifying the underlying causal relationships between different medical entities. For example, through these uncovered relationships, it can be discovered that Amoxicillin can treat Strep Throat. This not only enhances the accuracy and safety of the recommendations but also explains why Amoxicillin is effective for treating Strep Throat, thereby improving interpretability. Such relationships cannot be well identified through traditional methods that rely on co-occurrence relationships, leading to inappropriate recommendations. For instance, co-occurrence-based methods might recommend Ibuprofen for patients with Strep Throat, resulting in irrational suggestions and introducing biases that impact the performance of the recommendations. Specifically, we use statistical algorithms and machine learning to

---

Table 1
Notations and Descriptions.

Notations	Descriptions
$\mathcal{D}, \mathcal{P}, \mathcal{M}$	disease set, procedure set, medication set
$\|\mathcal{D}\|, \|\mathcal{P}\|, \|\mathcal{M}\|$	number of disease / procedure / medication
$\|S\|$	number of molecule
d, p, m	disease, procedure, medication
$\mathcal{H}$	patient's history record
v_t	clinical information of the t-th visit
$\mathbf{M}^{ddi}$	DDI matrix
$\mathbf{M}^{dm}, \mathbf{M}^{pm}$	causal effect matrice between disease / procedure and medication
G, G'	causal graph, enhanced graph
X_i	the i-th entity in Bayesian network
$\mathbf{h}_{d_i}, \mathbf{h}_{p_j}$	embedding of entity d_i / p_j
$\mathcal{R}$	relevance stratification structure
r_l^{dm}, r_l^{pm}	relationship between disease / procedure and medication at the l-th layer
$\mathbf{A}$	medication-molecule relationship matrix
a_{ij}	importance of molecules j in medication m_i.
$\mathbf{h}_{s_j}$	molecule embedding of s_j
$\mathbf{h}_{m_i}$	medication embedding of m_i
$\mathcal{N}^{en}$	set of medical entities
$\mathcal{N}^{mo}$	set of molecules
$\mathcal{E}^c$	relationships between medical entities
$\mathcal{E}^f$	relationships between molecules
$\mathbf{h}_i^{mo,l}$	representation of molecule i at layer l
$\mathbf{h}_i^{en,l}$	representation of entity i at layer l
$\mathbf{h}_{v_t}$	final represen- tation of the current visit v_t
$\mathbf{h}_H$	patient representation
τ_1, τ_2	magnitude of adjustment required
δ_1, δ_2	threshold value of causal matrix
$P(m_i)$	recommendation probability of m_i
$P'(m_i)$	recommendation probability of m_i after bias correction

exclude meaningless relationships in the data through backdoor [29] paths, thereby identifying potential causal relationships. Furthermore, causal inference uses observational data to quantify the impacts of these relationships, which is crucial for decision-making and policy formulation. It is important to note that in this paper, the causal mechanism permeates two aspects: the model and the recommendation results. In the model aspect, we leverage a causal graph to explore the relationships among diseases, procedures, and medications, enhancing the model's inferential capabilities. In the recommendation results aspect, we employ learned causal effects to further correct biases in the recommendation results. These specific details will be provided in the following section.

4. Methods

Figure 2 illustrates our model, which consists of three main components. In this study, we aim to mine and utilize the causal relationships among various medical entities across three distinct stages to achieve personalized medication recommendations. For ease of reading, we list important notations and their explanations in Table 1 of this paper.

First, in the **Relationship Mining** stage, we extract patient information (i.e., diseases, procedures, and medications) from EHRs. Using causal discovery, we learn
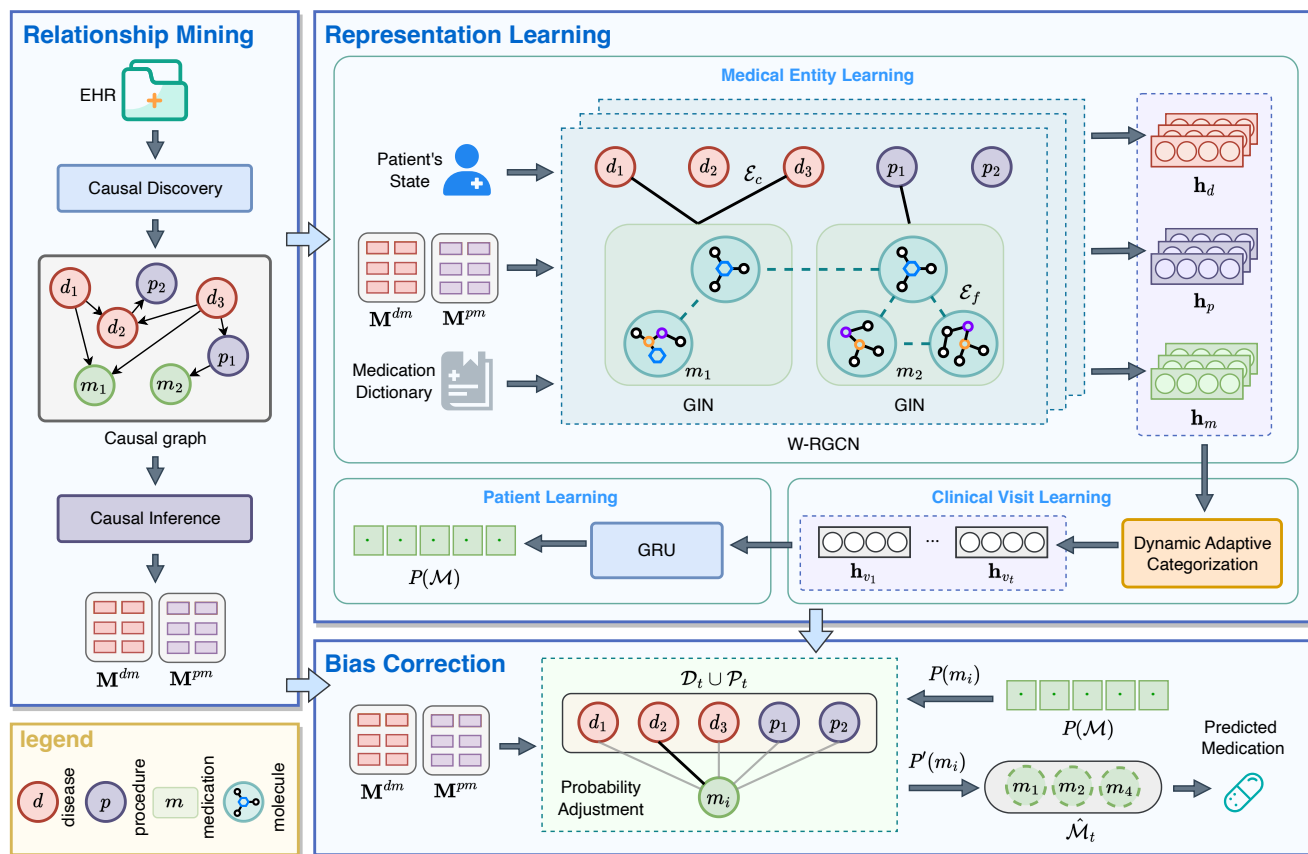

Figure 2: CIDGMed Flowchart: The Relationship Mining module on the left side of the figure utilizes causal discovery and causal inference on the EHR to generate causal graphs and causal effect matrices. The Representation Learning module at the top of the figure builds on this foundation with dual granularity at the medication level and the molecular level to learn patient representations and recommend preliminary medication probabilities. The Bias Correction module at the bottom of the figure corrects the recommended probabilities for each medication based on the causal effects and recommends the final combination of medications.

the causal relationships among these medical entities and construct personalized causal graphs for each patient. This process involves obtaining the causal effect matrices $\mathbf{M}^{dm}$ and $\mathbf{M}^{pm}$, which represent the relationships between diseases/procedures and medications. Next, in the **Representation Learning** stage, we refer to a medication dictionary [6] to establish mappings between each medication and its molecular components. We employ the causal effect matrix constructed during the relationship mining phase (based on the coarse-grained relationships at the medication level) and integrate the relationships between molecules and medications, as well as between molecules themselves (based on the fine-grained relationships at the molecular level), to form a dual-granularity representation message passing method. These representations help the study categorize and aggregate medical entities within the causal graph more precisely, forming a visit-level detailed representation $\mathbf{h}_v$. By incorporating patients' historical data, we ultimately derive the medication recommendation probability $P(\mathcal{M})$. Finally, in the **Bias Correction** stage, we adjust the probabilities for each medication based on the initial causal matrices $\mathbf{M}^{dm}$

and $\mathbf{M}^{pm}$ to correct biases that arise during the model training process, thus recommending medication combinations that are both effective and safe.

4.1. Relationship Mining

At this stage, our objective is to leverage causal discovery to deeply investigate the relationships between medical entities and to quantify these causal relationships as causal effects through causal inference. This will lay the essential groundwork for graph networks in subsequent representation learning and for bias correction methods in the recommendation phase.

Firstly, we explore the relationships among homogeneous entities, that is, entities of the same type, such as disease d_2 being caused by disease d_1. Based on the data distribution U within the EHR, we employ a causal discovery algorithm to construct causal graphs among the same type of medical entities. Given the extensive and discrete nature of medical data, we select the Greedy Intervention Equivalence Search (GIES) algorithm [28] for the causal discovery module. This method not only learns from the data but also iteratively optimizes Bayesian equivalence classes [30], enabling us to accurately determine the most suitable causal graph. Finally, we use an equivalence score

criterion $\mathcal{F}(G, U)$, which generates Bayesian equivalence classes, to assess the quality of the generated causal graph.

Algorithm 1 The Algorithm of Causal Discovery

1: **Input**: data distribution of EHR U
2: **Output**: causal graph G'
3: randomly initialize a graph G
4: **do**:
5:     $\mathcal{F}(G, U) \leftarrow$ compute equivalent classes (eq1)
6:     $G' \leftarrow$ update causal graph (eq2)
7: **while** $G' \neq G$
8: **return** causal graph G'

As depicted in Figure 3, traditional co-occurrence-based methods suggest that both d_1 and d_2 are associated with m_1. However, through causal discovery, we determine that d_2 is a disease caused by d_1, and m_1 is a medication specifically prescribed for d_1, with no direct relevance to d_2. The causal relationships established by this approach, based on the backdoor criterion [29], effectively eliminate spurious associations between medications and diseases. The entire causal discovery process is shown in algorithm 1.

These findings provide a robust foundation for the detailed examination of the relationships between medications and diseases or procedures in future tasks. The analytical approach applied is captured in the following formula:

$$\mathcal{F}(G, U) = \sum_{i=1}^{n} f(X_i, \mathcal{Q}_{X_i}^G), \quad (1)$$

$$G' = \text{GIES}(\mathcal{F}, G), \quad (2)$$

where n represents the number of medical entities (i.e., diseases, procedures, and medications) in the Bayesian network, X_i denotes the i-th entity in the network, and $\mathcal{Q}_{X_i}^G$ stands for the parents of X_i in the graph G, representing the set of entities that directly influence X_i according to the Bayesian network's structure. $f(\cdot)$ is a Bayesian evaluation function that evaluates the relationship between a given node and its parent node in a Bayesian network. The GIES($\cdot$) optimizes and learns an enhanced graph G' from the initial graph G and the equivalence score $\mathcal{F}$.

In addition to causal relationships existing among the same medical entities, there are also causal relationships between different medicinal entities. As shown in Figure 3, m_1 is a medication specifically for d_1. Hence, for heterogeneous relationships, we utilize causal inference to demonstrate the impact of medications on diseases or procedures. This straightforward approach clearly shows whether medications are effective against various diseases or procedures, thereby uncovering their direct relationships. Specifically, the causal graph G' is represented as a binary variable, and a discretized Generalized Linear Model (GLM) [13] is employed to model the causal effects between diseases/procedures and medications. The specific formula is as follows:

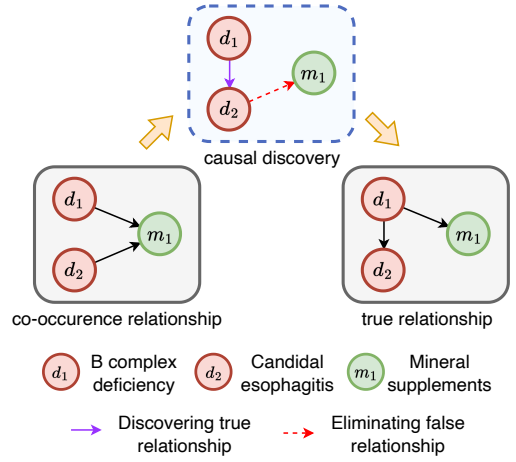$$\text{GLM}(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n, \quad (3)$$

Figure 3: A real example of using causal discovery to correct incorrect relationships from co-occurrence relationships to generate true relationships.

where GLM($\cdot$) denotes the logit link function, μ represents the probability of a binary variable, β_0 is the intercept, and $\beta_1, \beta_2, \ldots, \beta_n$ are the corresponding coefficients. $X_1, X_2, \ldots, X_n$ represent the model's independent variables, including the records of diseases/procedures and medications. In this model, the probability μ of the binary variable indicates the likelihood of a positive response from medications to patients with identified diseases or procedures, elucidating the causal effect of a specific disease or procedure on medication. Ultimately, the causal effect matrices for diseases-medications and procedures-medications can be obtained, denoted as $\mathbf{M}^{dm} \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{M}|}$ and $\mathbf{M}^{pm} \in \mathbb{R}^{|\mathcal{P}| \times |\mathcal{M}|}$, respectively.

4.2. Representation Learning

The main objective of this phase is to obtain detailed entity representations, visit representations, and ultimately comprehensive patient representations, built upon multiple causal-based relational networks, and to recommend preliminary medication probabilities.

4.2.1. Medical Entity Learning

This subsection describes the construction of dual-granularity relationship network and the process of obtaining entity representations.

In the real-world diagnosis and treatment process, doctors determine prescriptions based on the patient's current health status and their historical medical records. In our model, we use the current disease $\mathcal{D}_t$ and procedure $\mathcal{P}_t$ to represent the current health status and regard the diseases $\mathcal{D}_{t-1}$, procedures $\mathcal{P}_{t-1}$ and medications $\mathcal{M}_{t-1}$ from the last visit as the historical medical record. These combined data sources provide a comprehensive overview of the current visit, helping us to simulate the real-world medical consultation process. We initially establish embedding tables $\mathbf{E}_d \in \mathbb{R}^{|\mathcal{D}| \times dim}$ and $\mathbf{E}_p \in \mathbb{R}^{|\mathcal{P}| \times dim}$ for diseases and procedures, with each row corresponding to a specific disease

or procedure.

$$\mathbf{h}_{d_i} = \mathbf{E}_d(d_i), \quad \mathbf{h}_{p_j} = \mathbf{E}_p(p_j), \tag{4}$$

where $d_i \subset \mathcal{D}_t$, $p_j \subset \mathcal{P}_t$ represent specific medical entities, $\mathbf{h}_{d_i} \in \mathbb{R}^{dim}$, $\mathbf{h}_{p_j} \in \mathbb{R}^{dim}$ represent the embedding for entity d_i and p_j, respectively.

Coarse Granularity: First, we establish and apply a coarse-grained relationship network. Based on the previously generated $\mathbf{M}^{dm}$ and $\mathbf{M}^{pm}$, we introduce a relevance stratification strategy [21]. This strategy adaptively layers all relationships in $\mathbf{M}^{dm}$ and $\mathbf{M}^{pm}$ according to their values, using a gradient $\mathcal{K}$ to generate an n-layered relevance stratification structure $\mathcal{R}^{dm} = \{r_1^{dm}, r_2^{dm}, \dots, r_n^{dm}\}$ and $\mathcal{R}^{pm} = \{r_1^{pm}, r_2^{pm}, \dots, r_n^{pm}\}$. The stratification follows a pyramid structure, with more relationships at the lower layers and fewer at the upper layers. By assigning different weights to each layer in subsequent steps, we enhance the learning strength for rare data, addressing the imbalance in the data and obtaining coarse-grained relationships. The specific formulas are as follows:

$$|r_j^{dm}| = |r_1^{dm}|\mathcal{K}^{j-1}, \quad |r_l^{pm}| = |r_1^{pm}|\mathcal{K}^{l-1}, \tag{5}$$

where, r_l^{dm} refers to the lth layer of the medication-disease relationship, $|r_l^{pm}|$ denotes the number of relationships at this layer, $\mathcal{K}$ represents the hierarchical gradient.

Fine Granularity: Next, we mine fine-grained information, i.e., detailed information originating from the molecular structure level. By importing a pre-defined dictionary of medications, we obtain a precise mapping between medications and their molecular structures. Different structures may exist within the same medication, indicating that similar actions may be triggered; hence there is some connection between the molecular structures of the medication.

At the same time, most medications are composed of multiple molecular structures, and the same molecular structure may also exist in different medications, sharing the same representation. Therefore, we develop a learnable medication-molecule relationship matrix ($\mathbf{A} \in \mathbb{R}^{|\mathcal{M}| \times |\mathcal{S}|}$), where $a_{ij} \subset \mathbf{A}$ represents the importance of molecule s_j in medication m_i.

We apply a method similar to that used for constructing entity embeddings to build a molecule embedding table $\mathbf{E}_s \in \mathbb{R}^{|\mathcal{S}| \times dim}$. And, to accurately represent the medications, we construct medication embeddings ($\mathbf{h}_m$) based on their molecular composition structures and the importance of each molecule, thereby constructing fine-grained connections between medications at the structural level.

$$\mathbf{h}_{s_j} = \mathbf{E}_s(s_j), \tag{6}$$

$$\mathbf{h}_{m_i} = \sum_{j=1}^{|\mathcal{S}|} a_{ij} \cdot \mathbf{h}_{s_j}, \tag{7}$$

where $\mathbf{h}_{s_j} \in \mathbb{R}^{dim}$ represents the molecule s_j, $\mathbf{h}_{m_i}$ is the embedding of medication m_i, a_{ij} represents learnable weight.

Dual Granularity Fusion: To integrate relationships of different granularities, we use $\mathcal{G} = (\mathcal{N}^{en}, \mathcal{N}^{mo}, \mathcal{E}^c, \mathcal{E}^f)$, a bipartite graph containing both coarse-grained and fine-grained information. $\mathcal{G}'$ is a bipartite graph where $\mathcal{N}^{en}$ represents the set of all medical entities (coarse-grained medications) contained in the visit records, $\mathcal{N}^{mo}$ represents the set of molecules (fine-grained medications) involved in the medications from the visit records, $\mathcal{E}^c$ represents the relationships between medical entities (coarse-grained relationships), and $\mathcal{E}^f$ represents the relationships between molecules (fine-grained relationships).

This graph starts learning at the fine-grained relationship level, and medications are used as intermediaries, interacting with information from other medical entities through coarse-grained relationships, thereby enriching the representation of medications at different granularities. Specifically:

First, for the fine-grained relationships, we construct fully connected graphs between molecular structures belonging to the same medication and use Graph Isomorphism Networks (GIN) [49] for efficient message passing. This approach updates the representation $\mathbf{h}^{mo} = \{\mathbf{h}_s\}$ of the molecule within the medication, significantly enhancing the fine-grained relationships between the molecules. The specific formula is as follows:

$$\mathbf{h}_i^{mo,l+1} = \sigma \left((1+\epsilon)\mathbf{h}_i^{mo,l} + \sum_{j \in \mathcal{N}_{e^f,i}^{mo}} \mathbf{h}_j^{mo,l} \right), \tag{8}$$

where $i, j \in \mathcal{N}^{mo}$ is a node in the node set $\mathcal{N}^{mo}$, $\mathbf{h}_i^{mo,l+1}$ represents the updated feature representation of molecule node i at layer $l+1$, ϵ is a learnable parameter that controls the influence of the original feature representation, $\mathcal{N}_{e^f,i}^{mo}$ denotes the set of neighbouring molecules of i within the medication, $\mathbf{h}_i^{mo,l}$ represents the original feature representation of molecule i at layer l, and $\mathbf{h}_j^{mo,l}$ represents the original feature representation of neighboring molecule j at layer l.

Next, for the coarse-grained relationships, we perform adaptive adjustments based on the Relational Graph Convolutional Network [34], mapping the n levels of entity relevance $\mathcal{R}^{pm}$ and $\mathcal{R}^{dm}$ onto the $\mathcal{E}^c$ of graph $\mathcal{G}'$. By coarse-grained partitioning of the same relationships, we construct a Weighted-RGCN (W-RGCN) with coarse-grained weights. Through message passing on $\mathcal{G}'$, we update the representations of diseases, procedures, as well as the representations of medications $\mathbf{h}^{en} = \{\mathbf{h}_d, \mathbf{h}_p, \mathbf{h}_m\}$. In the face of coarse-grained relationships, the specific formula is as follows:

$$\mathbf{h}_i^{en,(l+1)} = \sigma \left(\sum_{e^c \in \mathcal{E}^c} \frac{1}{q_{e^c,i}^{en}} \sum_{j \in \mathcal{N}_{e^c,i}^{en}} \mathbf{W}_{e^c}^l \mathbf{h}_j^{en,l} \right), \tag{9}$$

$$\mathbf{W}_{e^c}^l = r_{e^c} + \Delta \mathbf{W}_{e^c}^l, \tag{10}$$

where $i, j \in \mathcal{N}^{en}$ is a medical entity in the entity node set $\mathcal{N}^{en}$, $\mathbf{h}_i^{en,(l+1)}$ is the updated representation of the node i at layer $l+1$, σ is the activation function, $\mathcal{E}^c$ is the set
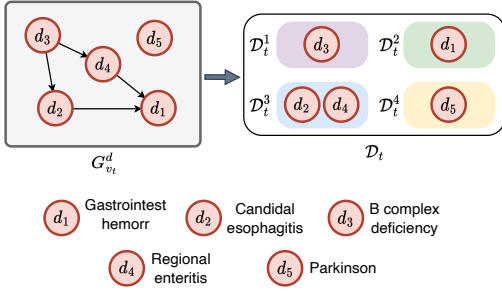
Figure 4: A real example of DAC, we analyze the causal position of each node in the causal graph $G_{v_t}^d$ to gauge its influence on the patient's current state. For instance, if d_3 is identified as a causal disease, it is classified as D_t^1.

of edges related to the medical entities, $q_{e^c,i}^{en}$ is a learnable normalization constant, ensuring the output features are appropriately scaled, $\mathcal{N}_{e^c,i}^{en}$ is the set of neighboring nodes of i related to edge e^c, $\mathbf{W}_{e^c}^l$ is the weight matrix for edge e^c at layer l, and r_{ec} and $\Delta\mathbf{W}_{e^c}^l$ represents the relationship relevance and the weight update matrix for edge e^c at layer l, respectively.

Finally, within graph G', we will propagate the molecular representations obtained from fine-grained relational learning to medications and further propagate the learning on coarse-grained relationships. This fusion of coarse and fine-grained propagation is fed back into the network in a recursive loop, achieving a dual-granularity fusion. Ultimately, we obtain the representations of various medical entities (diseases, procedures, and medications) integrating both granularities, respectively: $\mathbf{h}_d, \mathbf{h}_p$, and $\mathbf{h}_m$.

4.2.2. Clinical Visit Learning

Due to the varying physical constitutions of different patients, the same disease can have different impacts on different individuals. Moreover, different diseases also influence each other, and this varies from person to person. Based on this, the study employs a Dynamic Adaptive Categorization mechanism (DAC) [22] to learn about the effects of diseases in various clinical contexts, thereby enhancing personalized learning for patients.

DAC learns interactions between similar medical entities from causal graphs and integrates them into clinical visit representations. As shown in Figure 4, taking diseases as an example, based on the causal positions of entities in the session graph $G_{v_t}^d$, we divide $\mathcal{D}_t$ into four sets $\mathcal{D}_t^j$: (1) Causal Disease $\mathcal{D}_t^1$: Diseases capable of causing the occurrence of other diseases, acting as the root cause in the pathway, representing the primary diseases in this clinical visit. (2) Effect Disease $\mathcal{D}_t^2$: Diseases influenced by other diseases, serving as the result in the pathway, representing secondary symptoms in the visit. (3) Middle Disease $\mathcal{D}_t^3$: Diseases capable of causing the occurrence of other diseases while being influenced by other diseases, acting as intermediate nodes in the causal pathway. (4) Independent Disease $\mathcal{D}_t^4$:

Diseases existing independently in this visit without a direct causal relationship with other diseases.

Based on DAC, entities can be categorized into four types according to their positions in the causal graph of each user. This paper uses diseases as an example, with medications and treatments being similar. As shown in Figure 4, the four categories of diseases are as follows: (1) Source Diseases ($\mathcal{D}_t^1$), which trigger other diseases and are the cause of some diseases; (2) Intermediate Diseases ($\mathcal{D}_t^3$), which can be caused by other diseases and can also cause other diseases; (3) Resultant Diseases ($\mathcal{D}_t^2$), which can be caused by other diseases; (4) Independent Diseases ($\mathcal{D}_t^4$), which have no causal relationships with other diseases. The specific formula is as follows:

$$\mathcal{D}_t^j = \text{Classify}(d_i, G_{v_t}^d), \qquad (11)$$

where $\text{Classify}(\cdot)$ assigns d_i to a specific category $\mathcal{D}_t^j$.

Since the same disease occupies different positions in the causal graphs of different patients, resulting in different categorizations by DAC, they are classified into different categories. Therefore, this paper enhances user personalization by dynamically adjusting the impact of diseases on patients and assigning different weights to various disease categories. The specific formula is as follows:

$$w_t^j = \frac{\exp(\mathbf{W} \cdot \mathbf{h}_{D_t^j} + b)}{\sum_{k=1}^{4} \exp(\mathbf{W} \cdot \mathbf{h}_{D_t^k} + b)}, \qquad (12)$$

$$\mathbf{h}_{\mathcal{D}_t} = \sum_{i=1}^{|\mathcal{D}_t|} \mathbf{h}_{d_i} \cdot w_t^j, \qquad (13)$$

where w_t^j is the adaptive weight, $\mathbf{W}$ and $\mathbf{b}$ are the trainable weight matrix and bias term, and $\mathbf{h}_{D_t}$ is the final representation of the diseases. Similarly, the final representations for procedures and medications can also be obtained, denoted as $\mathbf{h}_{\mathcal{P}_t}$ and $\mathbf{h}_{\mathcal{M}_t}$ respectively. Therefore, the final representation of the visit is as follows:

$$\mathbf{h}_{v_t} = [\mathbf{h}_{\mathcal{D}_t} || \mathbf{h}_{\mathcal{P}_t} || \mathbf{h}_{\mathcal{M}_t}]. \qquad (14)$$

4.2.3. Patient Learning

Given the robust capability of Gated Recurrent Unit (GRU) [7] to manage complex sequential challenges, this study employs GRUs to analyze historical records, aiming to elucidate the sequential dependencies inherent in the progression of patients' historical diseases, procedures and medications $\{\mathbf{h}_{v_1}, \mathbf{h}_{v_2}, \dots, \mathbf{h}_{v_t}\}$. Then, a Multi-Layer Perceptron (MLP) with powerful nonlinear modelling capabilities is used to obtain the final representation of the patient. The specific formula is as follows:

$$\mathbf{o}_{v_t} = \text{GRU}(\mathbf{o}_{v_{t-1}}, \mathbf{h}_{v_t}), \qquad (15)$$

$$\mathbf{h}_H = \text{MLP}(o_{v_t}) \qquad (16)$$

where $\mathbf{o}_{v_{t-1}}$ is an intermediate variable generated by the GRU($\cdot$), and we represent $\mathbf{o}_{v_0}$ with a zero vector. Finally, we use a nonlinear activation function σ to transform the patient embeddings to the predicted probability $P(\mathcal{M})$ for each medication m_i.

$$P(m_i) = \sigma(\mathbf{h}_H), \qquad (17)$$

where $\mathbf{h}_H$ is the patient representation, and σ is a nonlinear activation function that is used to obtain the predicted probabilities for the medication m_i.

4.3. Bias Correction

Due to the inevitable errors that may arise during the model training process, to mitigate these errors, this paper applies bias correction to the predicted medications using the previously generated causal relationships, thereby more accurately capturing the impact of medications on specific diseases or procedures. Specifically, based on the causal effect matrices $\mathbf{M}^{dm}$ and $\mathbf{M}^{pm}$, for any medication m_i that has a significant therapeutic effect on a specific disease or procedure, we encourage its recommendation and increase its recommendation probability. Conversely, if the medication's therapeutic effect on any disease or procedure during this visit is below a certain quantified treatment effect threshold, we prefer to decrease its probability.

$$P'(m_i) = \begin{cases} P(m_i) + \tau_1, & max\{\mathbf{M}^{dm}_{m_i-D_t}, \mathbf{M}^{pm}_{m_i-P_t}\} \geq \delta_1 \\ P(m_i), & \delta_1 > max\{\mathbf{M}^{dm}_{m_i-D_t}, \mathbf{M}^{pm}_{m_i-P_t}\} \geq \delta_2 \\ P(m_i) - \tau_2, & max\{\mathbf{M}^{dm}_{m_i-D_t}, \mathbf{M}^{pm}_{m_i-P_t}\} < \delta_2 \end{cases}$$
$$(18)$$

where τ_1 and τ_2 represent the magnitudes of adjustment required for the predicted probability of a medication, while δ_1 and δ_2 are the thresholds derived from the causal matrix. When these thresholds are met, it indicates that the medication has a beneficial effect on the disease (or procedure), thus warranting a correction to the original predicted probability $P(m_i)$ of the medication. The adjusted probability, $P'(m_i)$, is the recommended probability after applying bias correction. Finally, based on the recommended probability of each medication, we generate the final medication combination.

4.4. Model Training

We define the medication recommendation process as a multi-label binary classification task and employ both the binary cross-entropy loss function L_{bce} and the multi-label margin loss function L_{multi}. Additionally, a DDI loss, L_{ddi}, is implemented to enhance the safety of medication recommendations by calculating the occurrence probability of medication pairs with potential DDI risk within the medication combination. The specific formulas for the three loss functions are as follows:

$$\mathcal{L}_{bce} = -\sum_{i=1}^{|\mathcal{M}|} m_i \log(\hat{m}_i) + (1 - m_i) \log(1 - \hat{m}_i), \quad (19)$$

$$\mathcal{L}_{multi} = \sum_{i,j:m_i=1,m_j=0} \frac{max(0, 1 - (\hat{m}_i - \hat{m}_j))}{|\mathcal{M}|}, \qquad (20)$$

$$\mathcal{L}_{ddi} = \sum_{i=1}^{|\mathcal{M}|} \sum_{j=1}^{|\mathcal{M}|} \mathbf{M}^{ddi}_{m_i m_j} \cdot \hat{m}_i \cdot \hat{m}_j, \qquad (21)$$

where m_i represents the ground truth, and $\hat{m}_i$ represents the model's predicted value for the medication.

A lower DDI can enhance the safety of recommendations, but pursuing the lowest DDI rate without considering the clinical context may compromise the effectiveness of the prescription. In other words, further blindly reducing DDI would decrease the efficacy of the medication combination, and there is no need to reduce DDI further. Therefore, in constructing the loss function, methods consistent with prior research [52] are employed to ensure a balanced consideration of each loss function.

$$\mathcal{L} = \alpha(\beta\mathcal{L}_{bce} + (1 - \beta)\mathcal{L}_{multi}) + (1 - \alpha)\mathcal{L}_{ddi}, \qquad (22)$$

$$\alpha = \begin{cases} 1, & \mathbf{rate}_{ddi} \leq \gamma \\ max\{0, 1 - \frac{\mathbf{rate}_{ddi}-\gamma}{kp}\}, & \mathbf{rate}_{ddi} > \gamma \end{cases}, \qquad (23)$$

where β represents hyperparameters that control model complexity or regularization strength, α, related to $rate_{ddi}$, serves as a controllable factor that adjusts the influence of the DDI rate on the model's decision-making process. γ, ranging from 0 to 1, signifies a DDI acceptance rate that determines the threshold at which the predicted DDI level is deemed acceptable for clinical use. Lastly, kp acts as a correction factor, adjusting the strength of the information based on proportionality.

The overall optimization process of the algorithm is shown in Algorithm 2.

5. Experiments

In this section, we conduct extensive experiments on our proposed CIDGMed, aiming to answer the following research questions (RQ):

- **RQ1**: Does CIDGMed significantly outperform other state-of-the-art models in terms of accuracy, safety, and time efficiency?

- **RQ2**: Do the key modules proposed in this paper enhance the model's performance?

- **RQ3**: Can the dual-granularity architecture in our proposed model recommend medications better than a single-granularity architecture?

- **RQ4**: What specific structures are captured by our causal inference method, and how do they contribute to the model's performance?

- **RQ5**: Why does our approach enhance both effectiveness and interpretability simultaneously?

Algorithm 2 The Algorithm of CIDGMed

1: **Input**: patient data $\mathcal{H}$, training epoch ep
2: **Output**: predicted value of medication $\hat{m}$
3: randomly initialize all model parameters
4: preprocess the EHR data
5:
6: *#relationship mining*
7: $G' \leftarrow$ GIES generate causal graph (eq2)
8: $\mathbf{M} \leftarrow$ GLM generate casual effect (eq3)
9:
10: **for** i **in** ep:
11:     *#medical entity learning*
12:     $\mathbf{h}_d \leftarrow$ compute disease representation (eq4)
13:     $\mathbf{h}_p \leftarrow$ compute procedure representation (eq4)
14:     $\mathbf{h}_s \leftarrow$ compute molecular representation (eq6)
15:     $\mathbf{h}_m \leftarrow$ aggregate $\mathbf{h}_s$ into medication (eq7)
16:
17:     *#fine granularity fusion*
18:     pass messages between $\mathbf{m}$(eq8)
19:     *#coarse granularity fusion*
20:     pass messages between $\mathbf{d/p}$ and $\mathbf{m}$ (eq9)
21:
22:     *#clinical visit learning*
23:     $h_{\mathcal{H}} \leftarrow$ integrate entity via causal graph G' (eq13)
24:
25:     *#bias correction*
26:     $\hat{m} \leftarrow$ correct bias using causal effect $\mathbf{M}$ (eq18)
27:
28:     update model parameters with loss (eq22)
29: **return** predicted value of medication $\hat{m}$

Table 2
Statistics of the datasets.

Item	MIMIC-III	MIMIC-IV
# patients	6,350	60,125
# visits	15,032	156,810
# diseases	1,958	2,000
# procedures	1,430	1,500
# medications	131	131
avg. # of visits	2.37	2.61
avg. # of medications	11.44	6.66

5.1. Datasets

As shown in Table 2, this paper utilizes the MIMIC-III [19] and MIMIC-IV [18] datasets, which are widely used in clinical research and analysis in the Intensive Care Unit (ICU). The data include comprehensive ICU patient clinical records, physiological monitoring data, laboratory test results, and medication records, among other information. We adopt the same data preprocessing method as described in previous work [22]. For the MIMIC-III dataset, we utilize ICD-9 codes for disease and procedure records, while for the MIMIC-IV dataset, we use both ICD-9 and ICD-10 codes for the relevant data, with medications mapped to ATC-3 codes.

We retain only those visit records that include information on diseases, procedures, and medications.

5.2. Baselines

To validate our model, we select the following high-performing methods as baseline models for comparison.

LR [15] (Logistic Regression) is a linear classification algorithm that estimates the probability of an outcome belonging to a certain category by a linear combination of input features, widely used in probability prediction and data classification tasks.

ECC [33] (Ensemble of Classifier Chains) employs a series of interconnected classifiers to enhance the precision of predictions, where each classifier uses its output as the input for the next classifier. This method is specifically designed for multi-label classification tasks and can effectively improve the overall performance of the model.

RETAIN [8] is an attention-based model tailored for sequence data analysis, adept at integrating temporal dynamics and specific features for accurate disease forecasting and disease. By dynamically capturing critical clinical events of a patient's history, RETAIN offers medication combinations.

GAMENet [35] is a medication recommendation that integrates the strengths of graph neural networks with memory networks and effectively discerns patterns and temporal sequences within medical data, thereby enhancing the precision of its predictions.

SafeDrug [51] leverages the combination of patients' health status and medication-related molecular knowledge. This approach, by reducing the impact of DDIs, can recommend safer medication combinations.

MICRON [50] focuses on customizing medication recommendation plans based on the dynamic changes in patient's health status. It does not produce new recommendations but updates medication combinations according to patients' new symptoms to enhance therapeutic effects while reducing potential side effects.

COGNet [46] employs the Transformer architecture for medication recommendations, using a translation approach to infer medications from illnesses. It also features a copy mechanism to integrate beneficial medications from past prescriptions into new recommendations.

MoleRec [52] delves into the importance of specific molecular substructures in medications. This approach enhances the precision of medication recommendations by leveraging finer molecular representations.

CausalMed [22] utilizes causal discovery based on patient status to identify primary and secondary diseases, thereby enhancing personalized patient representation.

5.3. Evaluation Metrics

We delve into the performance evaluation of our method using four principal metrics [51]: Jaccard index, DDI rate, F1 score, and PRAUC. These metrics provide comprehensive insights into the effectiveness and safety of our approach.

Jaccard (Jaccard Similarity Score) is employed to gauge the similarity between two sets. In medication recommendation, a higher Jaccard score indicates that the predicted

prescription is more consistent with the actual medication regimen, indicating higher accuracy.

$$Jaccard(t) = \frac{|\{i : \hat{m}_i = 1\}| \cap |\{i : m_i = 1\}|}{|\{i : \hat{m}_i = 1\}| \cup |\{i : m_i = 1\}|}, \quad (24)$$

$$Jaccard = \frac{1}{N_h} \sum_{t=1}^{N_h} Jaccard(t), \quad (25)$$

where $\hat{m}_i$ represents the predicted outcome, m_i represents the real label, and N_h represents the total number of visits for patient h.

DDI (Drug-Drug Interaction Rate) measures the occurrence of interactions within the recommended combinations, a lower rate indicates higher safety of the medications.

$$DDI = \frac{\sum_{i=1}^{N} \sum_{k,l \in \{j : \hat{m}_j(t)=1\}} 1\{a_{kl}^{ddi} = 1\}}{\sum_{i=1}^{N} \sum_{k,l \in \{j : m_j(t)=1\}} 1}, \quad (26)$$

where $m(t)$ and $\hat{m}(t)$ denote the real and predicted multi-label at the visit t, $m_j(t)$ denotes the jth entry of $m(t)$, a^{ddi} is the prior DDI relation matrix and 1 is an indicator function which returns 1 when $a^{ddi} = 1$, otherwise 0.

F1 (F1-score) combines precision and recall, reflecting the model's ability to accurately identify correct medications while ensuring comprehensive coverage.

$$Precision(t) = \frac{|\{i : \hat{m}_i = 1\} \cap \{i : m_i = 1\}|}{|\{i : \hat{m}_i = 1\}|}, \quad (27)$$

$$Recall(t) = \frac{|\{i : \hat{m}_i = 1\} \cap \{i : m_i = 1\}|}{|\{i : m_i = 1\}|}, \quad (28)$$

$$F1(t) = \frac{2}{\frac{1}{Precision(t)} + \frac{1}{Recall(t)}}, \quad (29)$$

$$F1 = \frac{1}{N_h} \sum_{i=1}^{N_h} F1(i). \quad (30)$$

PRAUC (Precision-Recall Area Under Curve) assesses model performance across different recall levels, indicating the ability to maintain precision with increasing recall.

$$PRAUC_t = \sum_{k=1}^{|M|} Precision_{k_t} \triangle Recall_{k_t}, \quad (31)$$

$$\triangle Recall_{k_t} = Recall_{k_t} - Recall_{k-1_t}, \quad (32)$$

where $|M|$ denotes the number of medications, k is the rank in the sequence of the retrieved medications, and $Precision_k(t)$ represents the precision at cut-of k in the ordered retrieval list and $\triangle Recall_{k_t}$ denotes the change in recall of a medication's ranking from $k-1$ to k. We averaged the PRAUC across all of the patient's visits.

$$PRAUC = \frac{1}{N_h} \sum_{i=1}^{N_h} PRAUC(t). \quad (33)$$

Avg.#Med (Average number of medications) indicates the average medications per recommendation. A higher value suggests more complex combinations, potentially increasing the risk of adverse reactions, while a lower value indicates safer, more minimalistic treatment plans. This metric should be used as a reference only and not as a strict evaluative criterion.

$$Avg.\#Med = \frac{1}{N_h} \sum_{i=1}^{N_h} |\hat{M}(i)|, \quad (34)$$

where $|\hat{M}(i)|$) denotes the number of predicted medications in visit i of patient h.

5.4. Implementation Details
5.4.1. Configuration and Parameter
For all entity embeddings, the optimal dimension is set to 64. The GIN has only 1 layer and W-RGCN has 2 layers. The activation function of the MLP in this paper is the Sigmoid function, with a dropout rate of 0.5. In the bias correction module, the upper bound δ_1 is set to 0.97, and the lower bound δ_2 is set to 0.90. In the loss function, β is set to 0.95, kp is set to 0.05, and the acceptance rate γ is set to 0.06. Training epoch ep=20, using the Adam optimizer with a learning rate of $lr = 0.0005$ and a regularization factor $Reg = 0.05$, and no batch is used in the training process.

5.4.2. Sampling Approach
Due to the limited availability of publicly accessible EHR data, we adopt bootstrapping sampling in this phase, following the approach recommended in [51]. This technique is particularly effective in scenarios where sample sizes are small, as detailed in [31] and [9].

5.4.3. Experimental Environment and Scalability
The experimental environment employed in this study consists of an Ubuntu 22.04 operating system, a 12-core CPU with 64.3% utilization on a single core, 3.4 GB of memory usage out of 30 GB available, and a 24 GB NVIDIA RTX 3090 GPU. The software dependencies include PyTorch version 2.0.0 and CUDA version 11.7.

5.5. Performance Comparison (RQ1)
In this section, we compare our model with baseline models, focusing on the accuracy, safety, and time efficiency of each.

5.5.1. Effectiveness Analysis
Table 3 and 4 details the results of the effectiveness comparison. Utilizing traditional machine learning approaches, LR and ECC demonstrate modest accuracy and, despite prescribing fewer medications, still encounter a higher DDI rate. In contrast, sequence-based models like LEAP, which employ advanced deep learning techniques, fail to surpass these traditional methods in effectiveness, highlighting the potential limitations of generative models in this domain. Furthermore, RETAIN, which introduces sequence models in medication recommendation, neglects the interactions

Table 3
The performance of each model on the MIMIC-III test set regarding accuracy and safety. The best and the runner-up results are highlighted in bold and underlined respectively under t-tests, at the level of 95% confidence level.

Model	Jaccard↑	DDI↓	F1↑	PRAUC↑	Avg.#Med
LR	0.4924±0.0011	0.0830±0.0025	0.6490±0.0019	0.7548±0.0009	16.0489±0.0015
ECC	0.4856±0.0017	0.0817±0.0018	0.6438±0.0012	0.7590±0.0024	16.2578±0.0007
RETAIN	0.4871±0.0008	0.0879±0.0022	0.6473±0.0027	0.7600±0.0010	19.4222±0.0017
LEAP	0.4526±0.0007	0.0762±0.0015	0.6147±0.0021	0.6555±0.0014	18.6240±0.0019
GAMENet	0.4994±0.0013	0.0890±0.0006	0.6560±0.0016	0.7656±0.0023	27.7703±0.0018
SafeDrug	0.5154±0.0015	**0.0655±0.0021**	0.6722±0.0011	0.7627±0.0008	19.4111±0.0022
MICRON	0.5219±0.0009	0.0727±0.0028	0.6761±0.0013	0.7489±0.0016	19.2505±0.0007
COGNet	0.5312±0.0018	0.0839±0.0005	0.6744±0.0023	0.7708±0.0014	27.6335±0.0011
MoleRec	0.5293±0.0020	0.0726±0.0024	0.6834±0.0008	0.7746±0.0012	22.0125±0.0006
CausalMed	0.5389±0.0011	0.0709±0.0007	0.6916±0.0022	0.7826±0.0010	20.5419±0.0028
CIDGMed	**0.5526±0.0016**	0.0684±0.0014	**0.7033±0.0009**	**0.7955±0.0026**	22.4693±0.0013

Table 4
The performance of each model on the MIMIC-IV test set regarding accuracy and safety. The best and the runner-up results are highlighted in bold and underlined respectively under t-tests, at the level of 95% confidence level.

| Model	Jaccard↑	DDI↓	F1↑	PRAUC↑	Avg.#Med
LR	0.4569±0.0011	0.0783±0.0013	0.6064±0.0016	0.6613±0.0019	8.5746±0.0012
ECC	0.4327±0.0014	0.0764±0.0011	0.6129±0.0018	0.6530±0.0017	8.7934±0.0010
RETAIN	0.4234±0.0017	0.0936±0.0015	0.5785±0.0013	0.6801±0.0012	10.9576±0.0018
LEAP	0.4254±0.0013	0.0688±0.0011	0.5794±0.0017	0.6059±0.0018	11.3606±0.0012
GAMENet	0.4565±0.0018	0.0898±0.0012	0.6103±0.0019	0.6829±0.0017	18.5895±0.0015
SafeDrug	0.4487±0.0012	**0.0604±0.0010**	0.6014±0.0017	0.6948±0.0018	13.6943±0.0014
MICRON	0.4640±0.0017	0.0691±0.0015	0.6167±0.0016	0.6919±0.0014	12.7701±0.0011
COGNet	0.4775±0.0014	0.0911±0.0013	0.6233±0.0019	0.6524±0.0018	18.7235±0.0015
MoleRec	0.4744±0.0013	0.0722±0.0014	0.6262±0.0018	0.7124±0.0017	13.4806±0.0015
CausalMed	0.4899±0.0014	0.0677±0.0017	0.6412±0.0013	0.7338±0.0019	14.4295±0.0012
CIDGMed	**0.5019±0.0012**	0.0644±0.0016	**0.6524±0.0014**	**0.7449±0.0018**	18.4956±0.0015
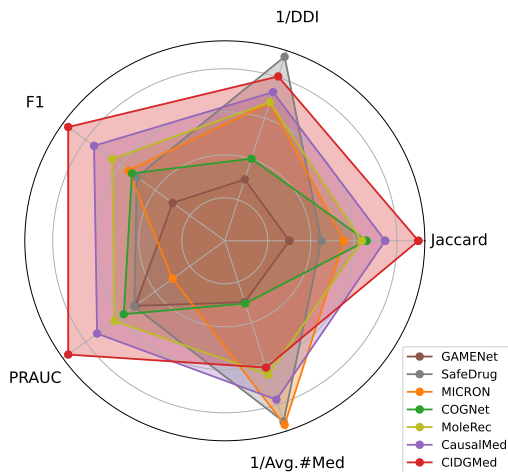
Figure 5: Comparison with recent outstanding works across all metrics in MIMIC-III.

between medications, consequently leading to an increased DDI rate.

We conduct a comprehensive comparison and discussion of several recent and high-performing baselines, as shown

in Figure 5. It is important to note that there are substantial differences in the metrics across different models. For ease of presentation, these metrics undergo standard normalization. Among the five indicators, the Jaccard index, F1 score, and PRAUC are used for accuracy testing, while DDI and Avg.#Med are used for safety testing. Considering that lower values of DDI and the Average Number of medications indicate improved safety, their reciprocals are represented in Figure 5.

First, GAMENet, COGNet, and MICRON do not perform well in terms of accuracy and safety. Although COGNet shows a decent Jaccard index, its performance on other accuracy metrics is poor. MICRON performs reasonably well on the Avg.#Med metric, but its DDI metric is unsatisfactory. This is because, despite considering patient history, these models do not fully explore the fine-grained relationships between diseases/procedures and medications, resulting in some errors in the recommendation outcomes and affecting the models' performance.

Secondly, SafeDrug and MoleRec conduct detailed analyses of fine-grained molecular-level data of medications. While their accuracy needs improvement, their safety is

Table 5
The performance of recent excellent models in training and inference efficiency.

Model	Convergence Epoch	Training Time /Epoch(s)	Total Training Time(s)	Inference Time(s)
GAMENet	39	45.31	1767.09	19.27
SafeDrug	54	38.32	2069.28	20.15
MICRON	40	17.48	699.20	14.48
COGNet	103	38.85	4001.55	142.91
MoleRec	25	249.32	6233.00	32.10
CausalMed	33	164.77	5437.41	18.29
CIDGMed	10	329.41	3294.10	21.85

guaranteed. This proves that fine-grained molecular structures of medications can help models avoid the adverse effects of medication interactions during recommendations. However, they only use a single-granularity approach and do not consider or explore the direct impact of the medication as a whole (coarse-granularity) on the disease or procedure. CausalMed is a patient-centred representation that emphasizes sequencing between diseases through causal discovery with a significant improvement in accuracy, but fails to take into account medication-patient associations at the molecular level of action. Consequently, their effectiveness is not as good as the method proposed in this paper, CIDGMed.

Therefore, our CIDGMed achieves excellent results in both accuracy and safety, further demonstrating the effectiveness of our approach by exploring causal relationships and combining dual-granularity information.

5.5.2. Time Efficiency Analysis

As shown in Table 5, this paper compares and studies the time efficiency of CIDGMed in terms of average training time per epoch, convergence epoch, total training time, and inference time against representative recommendation algorithms. To ensure fairness, all experiments are conducted in the experimental environment mentioned in 5.4.3. From the table 5 and Figure 5, we observe that CIDGMed achieves a better balance between time efficiency, recommendation accuracy, and safety.

Additionally, we present the time complexity of CIDGMed in detail. Specifically, In the relationship mining module, the main task is to utilize causal discovery to unearth the relationships between various entities and quantify these relationships through causal inference, with a time complexity of $O(n \cdot dim^2)$, where n represents the number of samples and dim represents the dimension of the features.

Next, the representation learning stage employs attention mechanisms, GIN, and W-W-RGCN for computation, along with GRU and MLP. The time complexity of this stage is primarily influenced by the length of the input sequence (n) and the dimension of the model (dim). The time complexity of the attention mechanism is $O(n^2 \cdot dim)$, while that of GIN and W-RGCN is $O(n^2)$. Additionally, the computational cost of integrating embeddings through adaptive classifiers and attention mechanisms is $O((n^2 + 1) \cdot dim)$. The time

complexity of GRU and MLP is related to the time steps T of each GRU unit, expressed as $O(T \cdot dim^2 + dim^2)$. Therefore, the overall time complexity of the representation learning stage can be expressed as $O(T \cdot dim^2 + dim^2 + n^2 \cdot dim + dim + n^2)$.

Finally, in the bias correction stage, the data originates from the causal inference. The main task of this stage is to evaluate and adjust probability values, with a time complexity of $O(n)$. Ultimately, the total complexity of the entire model is:

$$O((n + T + 1)dim^2 + (n^2 + 1)dim + n^2 + n). \quad (35)$$

In analyzing the time consumption of models, we observe that models with higher time costs often employ complex network structures of up to 3 to 5 layers, especially in the application of GNNs. In contrast, our design graph network construction method requires fewer layers, thereby significantly reducing the time overhead. Moreover, the bias correction mechanism we propose utilizes a method based on statistical principles and rules, which greatly alleviates the computational burden.

When assessing through both individual metrics and an overall view, CIDGMed significantly outshines other baseline models. Utilizes the best current model CausalMed as a comparative benchmark, CIDGMed notably lowers safety risks by reducing the DDI rate by 3.65%, boosts accuracy metric Jaccard by 2.54%, and also cuts down on training time by 39.42%, showcasing its efficiency and effectiveness in medication recommendation.

5.6. Ablation Study (RQ2&RQ3&RQ4)

To assess the contribution of different components of the proposed model CIDGMed, we test their importance by removing key modules.

CIDGMed *w/o* **C**: removes the coarse-grained relationship learning based on causal inference in the representation learning module. The relationships between diseases/procedures and medications use traditional co-occurrence relationships, meaning there is an edge between each pair of diseases/procedures and medications, with edge weights derived from co-occurrence rates.

CIDGMed *w/o* **F**: removes fine-grained relationships from the molecular structure level in the representation learning module. This eliminates structural connections between medications. Medication representations are obtained through random initialization.

CIDGMed *w/o* **C+F**: simultaneously removes the representation learning methods for both coarse-grained and fine-grained relationships based on causal inference. Instead, medication representations are obtained purely through random initialization while constructing the graph network using co-occurrence rates.

CIDGMed *w/o* **BC**: removes the bias correction module. The model does not perform bias correction based on causal relationships and uses the initial probabilities obtained from representation learning to recommend medication combinations.

Table 6
The performance of each ablation model on the test set regarding accuracy and safety. The best and the runner-up results are highlighted in bold and underlined respectively under t-tests, at the level of 95% confidence level.

Model	MIMIC-III				MIMIC-IV			
	Jaccard↑	DDI↓	F1↑	PRAUC↑	Jaccard↑	DDI↓	F1↑	PRAUC↑
CIDGMed *w/o* C	0.5484	0.0689	0.6954	0.7902	<u>0.4874</u>	0.0673	0.6382	0.7274
CIDGMed *w/o* F	<u>0.5494</u>	0.0726	<u>0.6992</u>	<u>0.7905</u>	0.4863	0.0701	<u>0.6389</u>	<u>0.7318</u>
CIDGMed *w/o* C+F	0.5477	0.0724	0.6924	0.7895	0.4856	0.0702	0.6366	0.7254
CIDGMed *w/o* BC	0.5357	**0.0672**	0.6888	0.7815	0.4798	<u>0.0637</u>	0.6307	0.7244
CIDGMed *w/o* C+F+BC	0.5336	0.0721	0.6875	0.7781	0.4836	0.0649	0.6346	0.7209
CIDGMed	**0.5526**	<u>0.0684</u>	**0.7033**	**0.7955**	**0.5019**	**0.0644**	**0.6524**	**0.7449**

CIDGMed *w/o* C+F+BC: removes most of the innovations from this paper. The model constructs graphs and recommends medications solely based on co-occurrence rates, without considering causal relationships or performing bias correction.

The results of the ablation study are shown in Table 6. Specifically, **CIDGMed *w/o* C** significantly decreases recommendation accuracy, indicating that the coarse-grained method based on causal inference can uncover the true causal relationships between medication entities, playing an essential role in improving overall accuracy. **CIDGMed *w/o* F** significantly decreases safety while slightly reducing recommendation accuracy, indicating that the fine-grained relationships based on molecular structure provide a crucial representation foundation for the safety relationships between medications and play a significant role in reducing safety hazards. The results of **CIDGMed *w/o* C+F** show that removing both coarse-grained and fine-grained representations significantly decreases performance. Thus, combining these representations compensates for deficiencies in safety and accuracy, leading to better and more balanced outcomes.

The results of **CIDGMed *w/o* BC** show a substantial decrease in accuracy, clearly indicating that the bias correction module significantly improves accuracy. This improvement is due to the module combining the causal relationships between diseases/procedures and medications, performing post-intervention processing during the recommendation phase to correct biases in the model learning process, ultimately enhancing accuracy. Additionally, safety shows a slight improvement.

The result of **CIDGMed *w/o* C+F+BC** shows that all metrics decline, clearly indicating a synergistic effect among our proposed innovations, and combining them results in significantly better improvements.

In summary, the results of multiple ablation experiments demonstrate the rationality and effectiveness of our modules, indicating that the design of the causal-based module is indispensable as it significantly enhances the model's accuracy and safety. Additionally, the dual-grained framework ensures more accurate and safer medication combinations, with the coarse-grained approach better serving model accuracy, while the fine-grained approach enhances the safety

of medication recommendations. We also observe that in the field of medication recommendation, safety and accuracy often constrain each other, and achieving optimal results typically involves balancing both aspects effectively.

5.7. Parameter Sensitivity

To thoroughly explore the impact of model parameters on performance, we conduct a sensitivity analysis on five key parameters using the MIMIC-III dataset. The results, shown in Figure 6, demonstrate consistent trends in both MIMIC-III and MIMIC-IV validations. However, due to space limitations, we only present the detailed validation results for MIMIC-III.

Figure 6a shows the model's performance under different embedding dimensions. When the embedding dimension is set to 64, the model's accuracy significantly surpasses other settings, and its safety performance is nearly identical to that at 128. As mentioned earlier, increasing the embedding dimension raises the model's time cost quadratically. Thus, considering accuracy, safety, and time cost, we select 64 as the optimal embedding dimension.

Secondly, the paper focuses on two core parameters when constructing the W-RGCN graph network: the number of layers and the number of node categories. The impact of the W-RGCN layers is illustrated in Figure 6b, showing that the model performs extremely poorly in accuracy when the number of layers is 1. While the accuracy performance does not differ much between 2, 3, and 4 layers, lower numbers of layers bring better safety. This is because an increase in layers can lead to overfitting, causing the model to overly pursue accuracy in combinations while neglecting medication safety, potentially leading to a slightly higher DDI index. Additionally, an increase in layers significantly raises the time cost. Hence, after comprehensive consideration, we chose 2 layers as the best setting.

The results regarding the number of relationships in W-RGCN categories are shown in Figure 6c. Finer division leads to fewer categories, and many different relationships are classified into the same category, making it difficult to express the differences between them. On the other hand, if there are too few categories, each category might lack enough samples, hindering the model's ability to be adequately trained and capture a generalized representation at
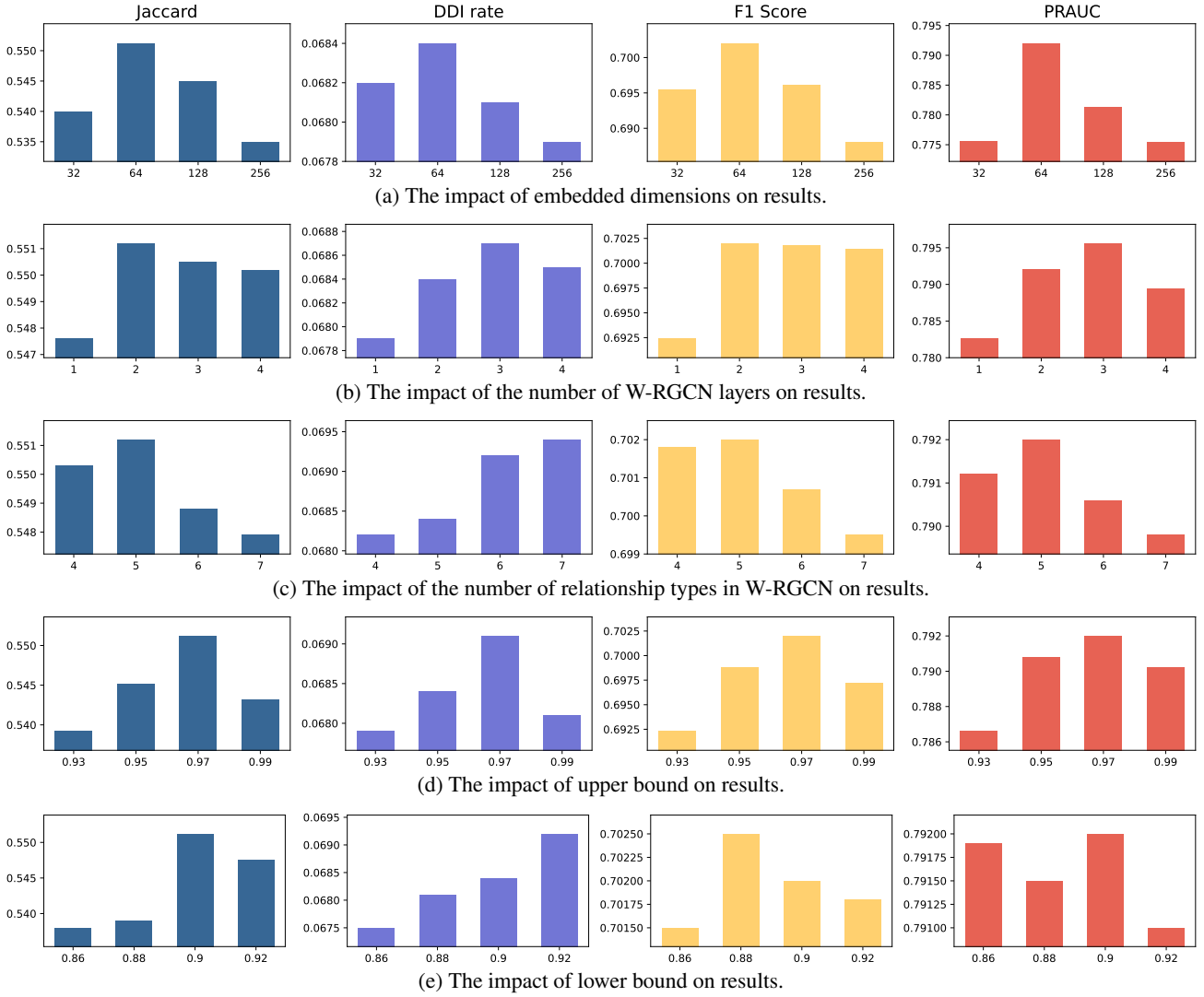
Figure 6: Experiments on parameter sensitivity on MIMIC-III dataset.

that level. Therefore, dividing into 5 categories was chosen as the optimal setting.

Regarding the setting of upper and lower bounds, it is crucial for the function of the probability adjustment module. It is worth mentioning that, in this experiment, both the upper and lower bounds were tested by incrementing from 0.5 to 1 in steps of 0.01. Ultimately, the four most representative values were selected for presentation. Figures 6e and 6d demonstrate that when the causal effect of a medication on a specific disease/procedure is below 0.90, such medication can be considered to have a low association with the disease, implying that the data correlation between the medication and disease is not directly caused by their causal relationship but may be due to co-occurrence phenomena caused by other entities slightly related to them. Conversely, when the causal effect is above 0.97, we regard it as having a direct causal relationship.

5.8. Case Study (RQ4&RQ5)

Figure 7 illustrates the workflow of this paper, especially the explanation of the bias correction process. To further elucidate the effectiveness and rationality of the model based on causal inference, we randomly selected some real cases in the MIMIC-III dataset.

As shown in figure 8, we randomly select two patients (*Patient*1 and *Patient*2) from our dataset to demonstrate the recommendation process based on the method proposed in this paper.

5.8.1. Causal Inference

We use traditional co-occurrence analysis methods and causal inference techniques, respectively, to explore the associations between these diseases and treatments. The casual inference part in figure 8 shows the relationships between diseases/ procedures and candidate medications. Taking patient 1 as an example, in the traditional co-occurrence-based relationship graph, almost every disease/procedure has a direct connection with medications, resulting in an overly
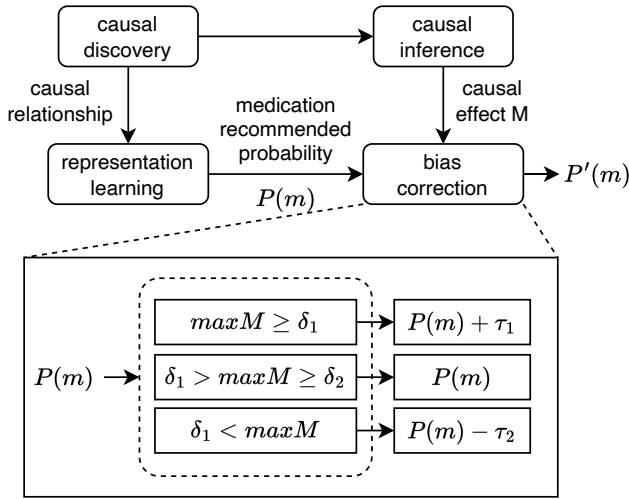
Figure 7: Detailed workflow for bias correction.

Table 7

The statistical analysis of the error in our method across both datasets.

Dataset	FN rate	FP rate
MIMIC-III	0.2392 ± 0.0050	0.0652 ± 0.0006
MIMIC-IV	0.2489 ± 0.0031	0.0712 ± 0.0008

complex and ambiguous network. Thus, the co-occurrence method cannot effectively capture the true underlying relationships between different entities, severely affecting the accuracy and safety of the recommendations and making it difficult to explain the reasons for the recommendations. In contrast, causal inference significantly clarifies the relationships between medications and diseases/procedures. For instance, in the co-occurrence relationships, m_9 is connected to d_2, d_3, d_1, and p_1, but causal inference reveals that m_9 (other mineral supplements) is only related to d_3 (B complex deficiency).

5.8.2. Bias Correction

We use the case of two patients to illustrate how bias correction is performed in this study. The detailed process is illustrated in Figure 8. Specifically, by reviewing their current disease conditions and procedure records, we first illustrate how the proposed method originally recommends medications (before bias correction) for each patient. Next, the bias correction module then utilizes the causal matrix to refine and adjust these original recommendations. This step-by-step process provides a clear demonstration of how the system works, from the original recommendation to the correction phase, ultimately enhancing the accuracy of the medication recommendations.

Through an in-depth analysis of the recommendation process using the bias correction module for $Patient1$ and $Patient2$, we summarize it into the following three scenarios.

Scenario 1: causal effect greater than the upper bound of 0.97. These medications have a highly relevant causal effect with the disease or procedure (greater than 0.97), indicating a strong connection to the patient's condition. As a result, they are considered more suitable for the patient, and their recommendation probability is increased accordingly. For example: medications m_1, m_5, and m_7 in $Patient1$, as well as m_2, m_3, and m_6 in $Patient2$.

Scenario 2: causal effect less than the lower bound of 0.90. These medications exhibit no significant causal effect with the disease or procedure (less than 0.90), making them less likely to be relevant to the patient's condition. As a result, their recommendation probability is reduced to avoid unnecessary or ineffective treatments. For example: medications m_2, m_3, and m_8 in $Patient1$.

Scenario 3: causal effect between 0.90 and 0.97. These medications have a causal effect with the disease or procedure between 0.90 and 0.97, so no adjustment in their recommendation probability is needed, and the recommendation remains unchanged. For example: medications m_4 and m_9 in $Patient1$, and m_4 in $Patient2$.

In summary, bias correction of medications based on the causal matrix between diseases/procedures and medications can significantly improve the performance of medication recommendations. Additionally, the in-depth analysis presented above enhances the interpretability of the proposed method, further validating its effectiveness.

5.9. Error Analysis and Limitation

We perform an error analysis of the model's recommendation results to better understand its limitations and identify areas for improvement. The prescription errors observed can be classified into two categories:

The first type of error is **False Negative (FN)**, where the model fails to recommend necessary medications, leading to omissions. This occurs mainly due to the infrequent appearance of certain medications in the records, which is also known as data sparsity in recommendation systems, making it difficult for the model to learn their relationships with associated diseases or procedures.

The second type of error is **False Positive (FP)**, where the model recommends medications that should not have been prescribed. This mainly occurs because certain disease or procedure and medication combinations are imbalanced in the training data, causing the model to favour recommending frequently occurring medications, leading to over-recommendation.

For these two cases, we conduct a statistical analysis of error, as shown in table 7. We find that the FN value is slightly higher than the FP not only due to the data sparsity but also because medication recommendation involves suggesting a set of medications. If any medication in the set is not recommended, it is considered an FN. This differs from other domains, where only a single item is recommended, resulting in a higher FN value in our case. Although the FP is higher than the FN, it remains within a reasonable range, as both the accuracy and safety of our model are greatly
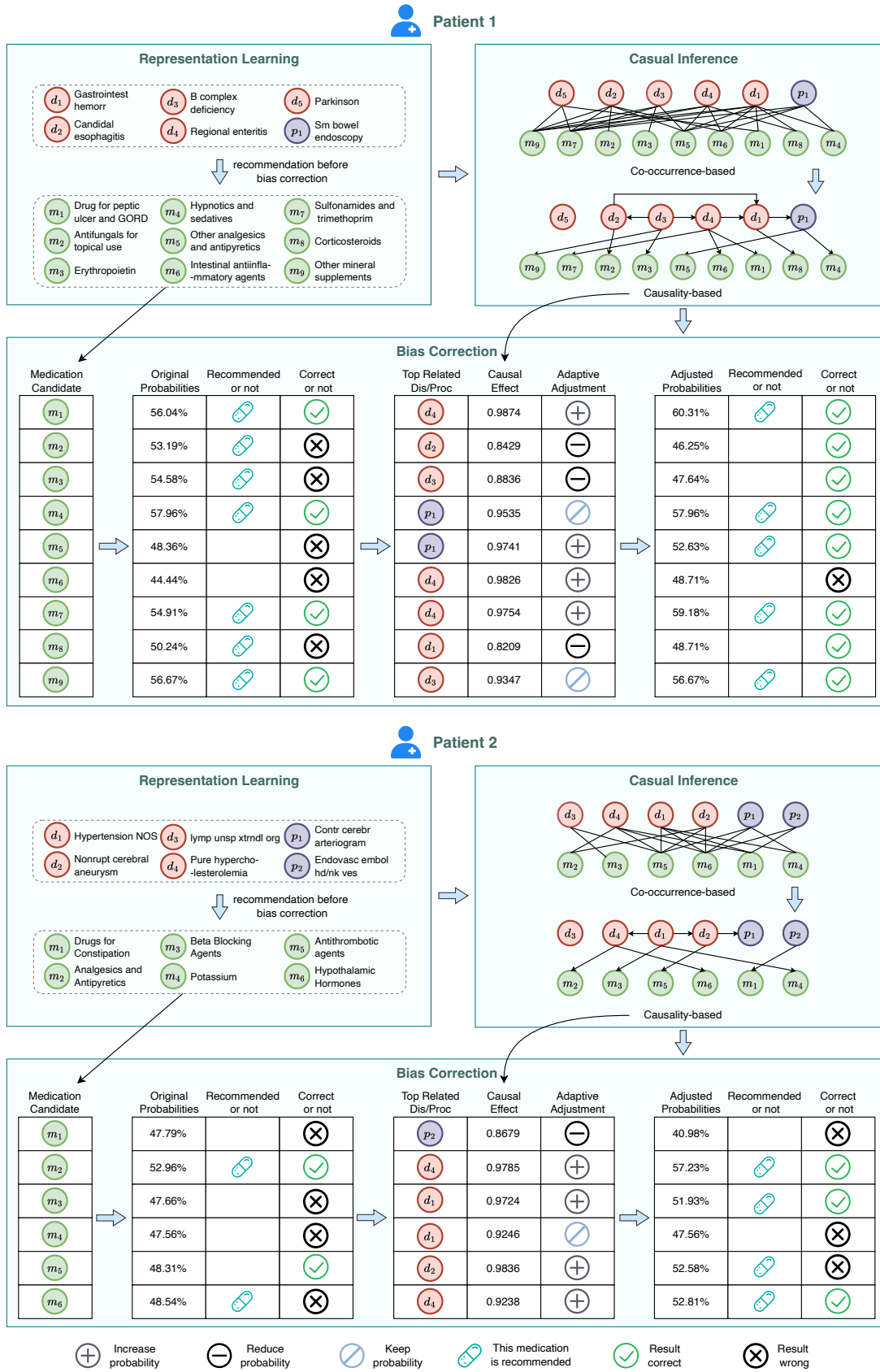
Figure 8: A detailed explanation of the process in this paper, using the real cases of two patients as examples.

improved compared to other baselines. In the future, we aim to further resolve the issue of over-recommendation.

6. Conclusion and Future Work

This paper introduces our developed medication recommendation system, CIDGMed. By applying causal inference, CIDGMed uncovers the causal relationships between diseases/procedures and medications at both the coarse-grained and fine-grained levels, achieving dual-granularity feature fusion. Furthermore, CIDGMed employs a post-processing intervention method, bias correction, during model recommendation. Through a series of rigorous experiments on publicly available clinical datasets, we significantly improve the model's accuracy and safety. Additionally, the case study analysis further validates the rationality of CIDGMed. In summary, the results fully demonstrate the superior performance of our method in terms of accuracy, safety, time efficiency, and rationality.

However, it has not yet been evaluated in real medical environments due to ethical considerations and data protection concerns. In the future, we aim to deploy this method in practical settings and engage with domain experts to refine our approach. Their feedback will be crucial for optimizing our framework to meet the specific needs of real-world medical applications, thereby enhancing its practical value and effectiveness.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Shunpan Liang: Supervision, Funding acquisition. **Xiang Li:** Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Investigation, Writing – original draft, Writing - Review & Editing, Visualization. **Shi Mu:** Writing – original draft, Writing - Review & Editing. **Chen Li:** Supervision, Writing – original draft, Writing - Review & Editing. **Yu Lei:** Supervision. **Yulei Hou:** Resources. **Tengfei Ma:** Supervision.

Acknowledgments

References

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

[2] Bougiatiotis, K., Aisopos, F., Nentidis, A., Krithara, A., Paliouras, G., 2020. Drug-drug interaction prediction on a biomedical literature knowledge graph, in: Artificial Intelligence in Medicine: 18th International Conference on Artificial Intelligence in Medicine, AIME 2020, Minneapolis, MN, USA, August 25–28, 2020, Proceedings 18, Springer. pp. 122–132.

[3] Chen, Q., Li, X., Geng, K., Wang, M., 2023. Context-aware safe medication recommendations with molecular graph and ddi graph embedding, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 7053–7060.

[4] Chen, T., Wong, R.C.W., 2020. Handling information loss of graph neural networks for session-based recommendation, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1172–1180.

[5] Chen, Y., Ma, T., Yang, X., Wang, J., Song, B., Zeng, X., 2021. MUF-FIN: multi-scale feature fusion for drug-drug interaction prediction. Bioinformatics 37, 2651–2658.

[6] Chiang, W.H., Shen, L., Li, L., Ning, X., 2020. Drug-drug interaction prediction based on co-medication patterns and graph matching. International Journal of Computational Biology and Drug Design 13, 36–57.

[7] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 .

[8] Choi, E., Bahadori, M.T., Sun, J., Kulas, J., Schuetz, A., Stewart, W., 2016. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. Advances in Neural Information Processing Systems 29.

[9] Dauji, S., Bhargava, K., 2016. Estimation of concrete characteristic strength from limited data by bootstrap. Journal of Asian Concrete Federation 2, 81–94.

[10] Gao, C., Wang, S., Li, S., Chen, J., He, X., Lei, W., Li, B., Zhang, Y., Jiang, P., 2023. Cirs: Bursting filter bubbles by counterfactual interactive recommender system. ACM Transactions on Information Systems 42, 1–27.

[11] Gao, C., Zheng, Y., Wang, W., Feng, F., He, X., Li, Y., 2024. Causal inference in recommender systems: A survey and future directions. ACM Transactions on Information Systems 42, 1–32.

[12] Hasan, U., Hossain, E., Gani, M.O., 2023. A survey on causal discovery methods for temporal and non-temporal data. arXiv preprint arXiv:2303.15027 .

[13] Hastie, T.J., Pregibon, D., 2017. Generalized linear models, in: Statistical models in S. Routledge, pp. 195–247.

[14] Hoffman, K.B., Dimbil, M., Erdman, C.B., Tatonetti, N.P., Overstreet, B.M., 2014. The weber effect and the united states food and drug administration's adverse event reporting system (faers): analysis of sixty-two drugs approved from 2006 to 2010. Drug safety 37, 283–294.

[15] Indra, S., Wikarsa, L., Turang, R., 2016. Using logistic regression method to classify tweets into the selected topics, in: 2016 international conference on advanced computer science and information systems (icacsis), IEEE. pp. 385–390.

[16] Jin, B., Yang, H., Sun, L., Liu, C., Qu, Y., Tong, J., 2018a. A treatment engine by predicting next-period prescriptions, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1608–1616.

[17] Jin, B., Yang, H., Sun, L., Liu, C., Qu, Y., Tong, J., 2018b. A treatment engine by predicting next-period prescriptions, in: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 1608–1616.

[18] Johnson, A.E., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T.J., Hao, S., Moody, B., Gow, B., et al., 2023. Mimic-iv, a freely accessible electronic health record dataset. Scientific data 10, 1.

[19] Johnson, A.E., Pollard, T.J., Shen, L., Lehman, L.w.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., Mark, R.G., 2016. Mimic-iii, a freely accessible critical care database. Scientific

data 3, 1–9.

[20] Le, H., Tran, T., Venkatesh, S., 2018. Dual memory neural computer for asynchronous two-view sequential learning, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1637–1645.

[21] Li, X., Liang, S., Hou, Y., Ma, T., 2023a. Stratmed: Relevance stratification between biomedical entities for sparsity on medication recommendation. Knowledge-Based Systems , 111239.

[22] Li, X., Liang, S., Lei, Y., Li, C., Hou, Y., Zheng, D., Ma, T., 2024. Causalmed: Causality-based personalized medication recommendation centered on patient health state, in: Proceedings of the 33nd ACM International Conference on Information and Knowledge Management, p. 1276–1285.

[23] Li, X., Zhang, Y., Li, X., Wei, H., Lu, M., 2023b. Dgcl: Distance-wise and graph contrastive learning for medication recommendation. Journal of Biomedical Informatics 139, 104301.

[24] Liang, X., Yang, J., Lu, G., Zhang, D., 2021. Compnet: Competitive neural network for palmprint recognition using learnable gabor kernels. IEEE Signal Processing Letters 28, 1739–1743.

[25] Lin, X., Quan, Z., Wang, Z.J., Ma, T., Zeng, X., 2020. Kgnn: Knowledge graph neural network for drug-drug interaction prediction., in: IJCAI, pp. 2739–2745.

[26] Ma, C., Kang, P., Liu, X., 2019. Hierarchical gating networks for sequential recommendation, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 825–833.

[27] Ma, T., Lin, X., Song, B., Yu, P.S., Zeng, X., 2023. Kg-mtl: Knowledge graph enhanced multi-task learning for molecular interaction. IEEE Transactions on Knowledge and Data Engineering 35, 7068–7081. doi:10.1109/TKDE.2022.3188154.

[28] Nandy, P., Hauser, A., Maathuis, M.H., 2018. Understanding consistency in hybrid causal structure learning. Ann. Stat .

[29] Pearl, J., Mackenzie, D., 2018. The book of why: the new science of cause and effect. Basic books.

[30] Pearl, J., et al., 2000. Models, reasoning and inference. Cambridge, UK: CambridgeUniversityPress 19, 3.

[31] Picheny, V., Kim, N.H., Haftka, R.T., 2010. Application of bootstrap method in conservative estimation of reliability with limited samples. Structural and Multidisciplinary Optimization 41, 205–217.

[32] Quadrana, M., Karatzoglou, A., Hidasi, B., Cremonesi, P., 2017. Personalizing session-based recommendations with hierarchical recurrent neural networks, in: Proceedings of the Eleventh ACM Conference on Recommender Systems, pp. 130–137.

[33] Read, J., Pfahringer, B., Holmes, G., Frank, E., 2011. Classifier chains for multi-label classification. Machine learning 85, 333–359.

[34] Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I., Welling, M., 2018. Modeling relational data with graph convolutional networks, in: The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15, Springer. pp. 593–607.

[35] Shang, J., Xiao, C., Ma, T., Li, H., Sun, J., 2019. Gamenet: Graph augmented memory networks for recommending medication combination, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 1126–1133.

[36] Si, Z., Han, X., Zhang, X., Xu, J., Yin, Y., Song, Y., Wen, J.R., 2022a. A model-agnostic causal learning framework for recommendation using search data, in: Proceedings of the ACM Web Conference 2022, pp. 224–233.

[37] Si, Z., Han, X., Zhang, X., Xu, J., Yin, Y., Song, Y., Wen, J.R., 2022b. A model-agnostic causal learning framework for recommendation using search data, in: Proceedings of the ACM Web Conference 2022, pp. 224–233.

[38] Tran, H.X., Le, T.D., Li, J., Liu, L., Liu, J., Zhao, Y., Waters, T., 2021. Recommending the most effective intervention to improve employment for job seekers with disability, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 3616–3626.

[39] Waghmare, G.H., Posinasetty, B., Shabaz, M., Rahin, S.A., Choudhary, A., Seena, K., 2024. Supervised context-aware latent dirichlet allocation-based drug recommendation model, in: Next Generation Computing and Information Systems: Proceedings of the 2nd International Conference on Next Generation Computing and Information Systems (ICNGCIS 2023), December 18-19, 2023, Jammu, J&K, India, CRC Press. p. 25.

[40] Wang, M., Liu, M., Liu, J., Wang, S., Long, G., Qian, B., 2017. Safe medicine recommendation via medical knowledge graph embedding. arxiv. Information Retrieval .

[41] Wang, W., Lin, X., Feng, F., He, X., Lin, M., Chua, T.S., 2022a. Causal representation learning for out-of-distribution recommendation, in: Proceedings of the ACM Web Conference 2022, pp. 3562–3571.

[42] Wang, Z., Chen, X., Zhou, R., Dai, Q., Dong, Z., Wen, J.R., 2023. Sequential recommendation with user causal behavior discovery, in: 2023 IEEE 39th International Conference on Data Engineering (ICDE), IEEE. pp. 28–40.

[43] Wang, Z., Liang, Y., Liu, Z., 2022b. Ffbdnet: Feature fusion and bipartite decision networks for recommending medication combination, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer. pp. 419–436.

[44] Wu, P., Li, H., Deng, Y., Hu, W., Dai, Q., Dong, Z., Sun, J., Zhang, R., Zhou, X.H., 2022a. On the opportunity of causal learning in recommendation systems: Foundation, estimation, prediction and challenges. arXiv preprint arXiv:2201.06716 .

[45] Wu, Q., Gao, Y., Gao, X., Weng, P., Chen, G., 2019. Dual sequential prediction models linking sequential recommendation and information dissemination, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 447–457.

[46] Wu, R., Qiu, Z., Jiang, J., Qi, G., Wu, X., 2022b. Conditional generation net for medication recommendation, in: Proceedings of the ACM Web Conference 2022, pp. 935–945.

[47] Xian, Y., Fu, Z., Muthukrishnan, S., De Melo, G., Zhang, Y., 2019. Reinforcement knowledge graph reasoning for explainable recommendation, in: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, pp. 285–294.

[48] Xiao, T., Wang, S., 2022. Towards unbiased and robust causal ranking for recommender systems, in: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, pp. 1158–1167.

[49] Xu, K., Hu, W., Leskovec, J., Jegelka, S., 2018. How powerful are graph neural networks? arXiv preprint arXiv:1810.00826 .

[50] Yang, C., Xiao, C., Glass, L., Sun, J., 2021a. Change matters: Medication change prediction with recurrent residual networks, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence 2021, pp. 3728–3734.

[51] Yang, C., Xiao, C., Ma, F., Glass, L., Sun, J., 2021b. Safedrug: Dual molecular graph encoders for safe drug recommendations, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, pp. 3735–3741.

[52] Yang, N., Zeng, K., Wu, Q., Yan, J., 2023. Molerec: Combinatorial drug recommendation with substructure-aware molecular representation learning, in: Proceedings of the ACM Web Conference 2023, pp. 4075–4085.

[53] Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., Zhang, A., 2021. A survey on causal inference. ACM Transactions on Knowledge Discovery from Data (TKDD) 15, 1–46.

[54] Yu, X., Li, X., Zhao, F., Yan, X., Zheng, X., Li, T., 2024. Aka-safemed: A safe medication recommendation based on attention mechanism and knowledge augmentation. Information Sciences , 120577.

[55] Zhang, H., Yang, X., Bai, L., Liang, J., 2023. Enhancing drug recommendations via heterogeneous graph representation learning in ehr networks. IEEE Transactions on Knowledge and Data Engineering .

[56] Zhang, X., Jia, H., Su, H., Wang, W., Xu, J., Wen, J.R., 2021a. Counterfactual reward modification for streaming recommendation with delayed feedback, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 41–50.

[57] Zhang, Y., Feng, F., He, X., Wei, T., Song, C., Ling, G., Zhang, Y., 2021b. Causal intervention for leveraging popularity bias in recommendation, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 11–20.

[58] Zheng, Y., Gao, C., Chang, J., Niu, Y., Song, Y., Jin, D., Li, Y., 2022. Disentangling long and short-term interests for recommendation, in: Proceedings of the ACM Web Conference 2022, pp. 2256–2267.

[59] Zheng, Z., Wang, C., Xu, T., Shen, D., Qin, P., Huai, B., Liu, T., Chen, E., 2021a. Drug package recommendation via interaction-aware graph induction, in: Proceedings of the Web Conference 2021, pp. 1284–1295.

[60] Zheng, Z., Wang, C., Xu, T., Shen, D., Qin, P., Huai, B., Liu, T., Chen, E., 2021b. Drug package recommendation via interaction-aware graph induction, in: Proceedings of the Web Conference 2021, pp. 1284–1295.

[61] Zheng, Z., Wang, C., Xu, T., Shen, D., Qin, P., Zhao, X., Huai, B., Wu, X., Chen, E., 2023. Interaction-aware drug package recommendation via policy gradient. ACM Transactions on Information Systems 41, 1–32.