Explainable Reasoning Path Inference of Anti-Cancer Drug Sensitivity on Genomic Knowledge Graph via Macro-Micro Agent Collaborative Reinforcement Learning

Minhua Feng, Liping Tang, Juntao Liang, Song Huang, Jianfeng Ma, Zhimin Zheng, Ranran Guo, Wen Shi, *Member, IEEE*, and Jianing Xi, *Senior Member, IEEE*

***Abstract*—Artificial intelligence (AI) based anticancer drug recommendation systems have emerged as powerful tools for precision dosing. Although existing methods have advanced in terms of predictive accuracy, they encounter three significant obstacles, including the "black-box" problem resulting in unexplainable reasoning, the computational difficulty for graph-based structures, and the combinatorial explosion during multi-step reasoning. To tackle these issues, we introduce a novel Macro-Micro agent Drug sensitivity inference (MarMirDrug). Specifically, our methodology enhances interpretability via knowledge graphs (KG). To reduce computational overhead, our MarMirDrug also transforms the graph structures of KG into low-dimensional embeddings. To manage the combinatorial explosion of graph paths that occurs with increasing inference steps, we incorporate a macro-micro dual-agent reinforcement learning algorithm, and combines reinforcement learning with KG-based reasoning to infer path reasoning. In computational experimental outcomes, the efficiency of our embedding model surpasses that of the baseline model, achieving an average improvement of 117.65% for hit score. Also, our dual-agent framework exhibits superior performance and interpretability in drug response prediction, with AUC values outperforming conventional baseline methods. In summary, our approach integrates interpretability, computational efficiency, and predictive accuracy, providing novel contributions to precision medicine.**

***Index Terms*—Decision support, machine learning, data mining, knowledge graph, cancer drug response.**

I. Introduction

CANCER has emerged as a global public health crisis, posing a significant threat to human health worldwide [1]. Drug therapy remains a cornerstone of cancer treatment, yet conventional approaches often fail to account for the substantial variability in drug sensitivity among individual patients due to tumor heterogeneity [2]. This variability leads to marked differences in treatment responses and directly impacts therapeutic outcomes. In the absence of reliable predictive tools, clinicians frequently resort to empirical methods or trial-and-error approaches, which are not only inefficient but also carry substantial risks. The emergence of precision medicine offers a transformative solution to this challenge [3]. By integrating molecular diagnostics with advanced data analytics, precision medicine enables the development of tailored treatment strategies that optimize therapeutic efficacy while improving patients' quality of life [4]. Within this paradigm, AI based anticancer drug recommendation systems have emerged as powerful tools for precision dosing, enabling the formulation of more scientific, accurate, and personalized treatment plans that enhance therapeutic outcomes while minimizing adverse effects.

Recent years have witnessed significant advancements in drug recommendation systems for precision dosing. Current approaches to anticancer drug recommendation primarily utilize statistical learning [5], [6], similarity analysis [7], and deep learning [8] techniques, achieving notable success in prediction accuracy. However, these methods are frequently hampered by the "black-box" problem, characterized by limited transparency and interpretability in their decision-making processes. In the medical domain, particularly in precision medicine, the lack of interpretability in AI systems presents a critical challenge that demands urgent attention [9]. Specifically, effective drug recommendation systems must integrate comprehensive patient-specific information with detailed drug-related data to generate personalized treatment suggestions. The

Received 18 March 2025; revised 31 August 2025; accepted 4 September 2025. Date of publication 8 September 2025; date of current version 10 December 2025. This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515010851 and Grant 2022A1515110001, in part by the National Natural Science Foundation of China under Grant 62202117, in part by the Tertiary Education Scientific research project of Guangzhou Municipal Education Bureau under Grant 2024312264, in part by Guangzhou Basic and Applied Basic Research Foundation under Grant SL2023A04J02440, and in part by the Special Foundation in Department of Higher Education of Guangdong under Grant 2022ZDX2053. *(Minhua Feng and Liping Tang contributed equally to this work.) (Corresponding authors: Wen Shi, Jianing Xi.)*

Minhua Feng, Ranran Guo, and Jianing Xi are with the Guangzhou Institute of Cancer Research, Affiliated Cancer Hospital, Guangzhou Medical University, Guangzhou 510095, China, and also with the School of Biomedical Engineering, Guangzhou Medical University, Guangzhou 511436, China (e-mail: xjn@gzhmu.edu.cn).

Liping Tang, Juntao Liang, Song Huang, Jianfeng Ma, and Zhimin Zheng are with the School of Biomedical Engineering, Guangzhou Medical University, Guangzhou 511436, China.

Wen Shi is with the Affiliated Traditional Chinese Medicine Hospital, Guangzhou Medical University, Guangzhou 511436, China, and also with the School of Biomedical Engineering, Guangzhou Medical University, Guangzhou 511436, China (e-mail: shiwen@gzhmu.edu.cn).

The source code and data of MarMirDrug are available at https://github.com/Minhua-F/MarMirDrug_code.

Digital Object Identifier 10.1109/TCBBIO.2025.3607142

complexity of drug molecular structures combined with the diversity of bioinformatics data makes drug response prediction particularly challenging, often necessitating complex AI algorithms to achieve sufficient accuracy. However, when the analytical processes underlying these recommendations remain unexplainable to clinicians, potential discrepancies with clinical experience can erode trust in the AI system, ultimately hindering its practical implementation [10]. Therefore, developing systems that provide interpretable explanations is crucial for clinician trust and adoption, thereby facilitating effective human-machine collaboration and optimizing treatment decisions.

The challenge of enhancing interpretability in anticancer drug recommendation systems represents a critical research frontier. To address the limitations in demonstrating reasoning processes within interpretable path, researchers have increasingly turned to knowledge graphs (KG) as a means of representing drug action pathways. As a structured knowledge representation method, KGs provide clear representations of complex relationships within drug mechanisms and disease progression [11]. However, the computational analysis of graph paths presents substantial challenges, particularly when applied to large-scale biomedical datasets. The inherent complexity of graph structures makes direct computation problematic, and as the number of inference steps increases, the potential path combinations grow exponentially. This not only significantly increases computational complexity, but also makes the extraction of meaningful information from vast datasets exceptionally challenging [12]. Recently, Graph Neural Networks (GNNs) excel at node-level inference and representation learning, but they still generally lack explicit support for path-level interpretability and struggle to control the combinatorial complexity of long-hop reasoning [13]. Consequently, developing methods that effectively visualize reasoning processes while efficiently managing the combinatorial explosion of paths is crucial for advancing precision oncology.

In this paper, we introduce a novel algorithm named Macro-Micro agent Drug sensitivity inference (MarMirDrug). To address the limitations of current AI-based drug recommendation models in terms of interpretability and predictive accuracy, this study proposes a novel precision drug recommendation system for anticancer drug sensitivity based on KGs. To enhance interpretability, we employ KG technology, which provides reasoning processes for the "black-box" limitations. To overcome the computational challenges associated with graph structures, we transform graph entities into vector representations suitable for computational analysis [14]. To manage the combinatorial explosion of graph paths that occurs with increasing inference steps, we incorporate a macro-micro dual-agent reinforcement learning algorithm. Our approach combines reinforcement learning with KG-based reasoning to infer path reasoning. Our comparative analysis demonstrates that MarMirDrug outperforms traditional single-agent reinforcement learning approaches in both interpretability and computational efficiency. The proposed algorithm effectively mitigates the explosive growth of path combinations, significantly enhancing system efficiency and practical applicability. The method achieves superior reasoning performance and less computational cost on the Genomics of Drug Sensitivity in Cancer (GDSC) dataset based KG [15], demonstrating its effectiveness. This approach offers new perspectives and insights for explainable medical AI reasoning, potentially revolutionizing precision oncology drug recommendation systems.

II. Related Work

To conceptualize the drug recommendation process into a reasoning path inference paradigm, we can formalize it as follows: Let $S_X = \{x_1, x_2, \ldots, x_n\}$ represent the set of patient samples, where each x_i corresponds to a distinct patient. Let $S_y = \{y_1, y_2, \ldots, y_n\}$ denote the set of drugs, with each y_i representing a specific drug. The recommendation relationships between patients and drugs are defined by $S_R \subseteq S_X \times S_y$, where each element (x_i, y_i) indicates a recommendation relationship between patient x_i and drug y_i. Additionally, let $S_{Path} \subseteq S_X \times S_y$ represent the set of valid paths within the drug recommendation graph, where each element $(x_i, n_1, n_2, \ldots, n_k, y_i)$ signifies a valid path from patient x_i to drug y_i, where n_1 to n_k denote the intermediate nodes across the path.

The primary objective of a drug recommendation system is to identify a set of patient-drug pairs (x, y) that satisfy the following conditions:

$$x \in S_x,\ y \in S_y;\ \exists\,(x, y) \in S_R \tag{1}$$

Here the existential quantifier $\exists(x, y) \in S_R$ is used to emphasize that the rule R is triggered only for some specific (x, y) in the relation set S_R.

The system must also adhere to the sensitivity condition: if patient x exhibits sensitivity to drug y, indicating the drug's efficacy for the patient, then (x, y) is considered a successfully matched pair. To ensure interpretability, the identification only of such a pair is insufficient; the path connecting x to y must also reside within the set of valid paths, S_{Path}, as expressed by:

$$\exists\,(x, n_1, n_2, \ldots, n_k, y) \in S_{Path}, \tag{2}$$

where $(n_1, n_2, \ldots, n_k)$ represents the sequence of nodes linking patient x to drug y. This path may incorporate information about the drug's mechanism of action, the patient's pathological characteristics, potential side effects, and other relevant factors. The existence of such a path ensures that the recommendation is not merely based on statistical associations, but rather on the intrinsic relationships between the drug and the patient.

Recent advancements in cancer drug recommendation systems have focused on leveraging statistical learning, similarity-based approaches, and deep learning techniques. For instance, Liu et al.'s ensemble model, which combines matrix completion with ridge regression, achieves high predictive accuracy in anticancer drug response and identifies biologically essential genes [5]. Similarly, Tao et al.'s CADRE model employs similarity-based collaborative filtering with contextual attention, enhancing prediction accuracy and facilitating biomarker discovery [7]. More sophisticated methods, such as Su et al.'s SRDFM, a deep learning-based model, excel in ranking anticancer drugs for personalized therapy, demonstrating effectiveness in both single and combination treatments [8]. While these methods achieve

remarkable predictive accuracy, their "black-box" nature limits interpretability, hindering their adoption in clinical settings.

The lack of interpretability in drug recommendation systems poses a significant challenge. Even if artificial intelligence models excel in generating accurate recommendations, physicians remain unable to understand the underlying reasoning processes, undermining trust in the system. To address this, KGs have been employed to enhance interpretability. KGs represent knowledge using nodes, edges, and attributes to denote entities, properties, and relationships, respectively. For example, Caro-Martínez et al.'s graph-based model improves explainability in recommender systems by using interaction graphs to provide clear, interpretable recommendations [16]. He et al.'s CEKGR model enriches KGs with community semantics, illuminating the rationale behind user preferences [17]. Additionally, Wang et al.'s AS-KGAN model leverages high-order relationships in KGs to refine personalized recommendations [18]. While these methods enhance interpretability, they still face computational challenges associated with graph processing.

To mitigate the computational challenges of graph processing, researchers have turned to graph embeddings, which map graphs to low-dimensional vectors. For example, Zulaika et al.'s SparseRESCAL model optimizes KG embeddings through regularized online tensor factorization, achieving computationally efficient and interpretable representations [19]. Forouzandeh et al.'s model streamlines complex user behavior analysis into computationally manageable vectors for precise recommendations [20]. Similarly, Shu and Huang's Multi-Rec model efficiently distills complex user-item relationships for superior recommendation performance [21]. Neural Graph Collaborative Filtering further integrates collaborative signals with graph structure for enhanced recommendation quality [22]. KGAT leverages attention-based neighbor aggregation to boost node representations [13]. However, embeddings typically support only single-step reasoning, limiting their interpretability for multi-step inference.

To enable multi-step reasoning, reinforcement learning techniques have been introduced [23]. For instance, recent work on explainable reasoning over knowledge graphs for recommendation emphasizes transparent path inference [24]. Also, KPRLN model leverages reinforcement learning for long-step reasoning, enabling deeper understanding of user preferences in KG-based recommendations [25]. Yang et al.'s Hierarchical Reinforcement Learning model integrates KG reasoning with conversational recommendations, incorporating dialogue history for accurate predictions [26]. Li et al.'s TMER-RL model dynamically explores item-item paths, tracing user-item interactions for precise and explainable recommendations [27]. The most recent advancement, Gao et al.'s KG-Predict model, rather than personalized drug recommendation, performs prediction on drug-disease interactions [28]. However, while these reinforcement learning approaches excel in long-step reasoning, they often struggle with combinatorial explosion as the number of steps increases, leading to computational inefficiencies [12].

It is worth noting that only a few of the referred works have been directly applied to drug recommendation scenarios. Nevertheless, all the studies are still highly relevant at the methodological level and share structural consistency with our knowledge graph reasoning framework. In summary, the field of drug recommendation urgently requires models that balance interpretability, computational efficiency, and the ability to support long-step reasoning. Such models would address the limitations of existing approaches and pave the way for more reliable and transparent drug recommendation systems.

III. Materials and Methods

A. Framework Overview

To develop an explainable KG-driven reasoning system for predicting anti-cancer drug sensitivity, we organize the task into three core components. First, we begin with Knowledge Acquisition and Representation, which transforms raw drug response data from databases into a structured KG format. This step involves collecting and processing data to construct the KG. Second, to address the computational complexity of graph-structured data, we use a graph embedding-based model to convert complex graph structures into low-dimensional vector representations. This provides a computationally efficient foundation for subsequent tasks. Finally, to reduce the computational cost associated with combinatorial explosion during the graph search process, the third step introduces a Macro-Micro Agent Collaborative Reinforcement Learning algorithm for drug sensitivity prediction. This approach tackles KG long-step reasoning tasks, effectively mitigating the combinatorial explosion problem through the collaborative mechanism of dual agents. By integrating these three components, our framework establishes an efficient and interpretable reasoning system for anti-cancer drug sensitivity prediction, offering novel technical support for precision medicine.

B. Data Acquisition

To perform drug sensitivity inference, both drug response data and genomic mutation data are required. The GDSC database provides a vast amount of data, including drug sensitivity measurements across various samples and information on the relationships between drug responses and genomic mutations. However, for patients with rare genomic mutations, GDSC data alone may not be sufficient, as the dataset might lack associations between these rare mutations and drug responses. To address this limitation, additional genomic information from datasets such as the gene-gene interaction database iRefIndex can be utilized [29]. By leveraging the functional connections between genes, iRefIndex helps infer the potential impact of rare gene mutations. GDSC and iRefIndex are aligned through Ensembl IDs to ensure full compatibility. Densifying the graph with drug–gene and gene–gene interactions help address sparse mutation data (details in supplementary materials). Therefore, integrating GDSC with iRefIndex enables the construction of a comprehensive KG for more robust and accurate drug sensitivity prediction.

C. Explainability: Pharmacogenomics KG

1) Bipartite Graphs for Feature Matrix: During data collection, we obtain three matrices from different types of information: drug response and gene mutation data from GDSC, and gene interaction data from iRefIndex. These matrices include: 1) The drug sensitivity matrix, which captures drug response states (sensitivity or resistance); 2) The gene mutation matrix, which records various types of mutations from mutated genes in each sample; 3) The adjacency matrix, which represents interactions between genes, forming a gene interaction network. However, these matrices cannot be directly converted into a KG due to their conflicting formats as either feature matrices or adjacency matrices. Interestingly, we observe that both the feature matrices and adjacency matrices are sparse. Specifically, the elements of the sensitivity matrix indicate whether a drug-sample pair is sensitive or resistant. The mutation matrix reveals mutations between genes and samples. Also, the adjacency matrix represents interactions between different genes. All three matrices can be interpreted as bipartite graphs, enabling the construction of a comprehensive drug sensitivity KG.

2) Triplet Construction for KGs: Although we have translated the three matrices into bipartite graphs, they have not yet been integrated into a unified KG. During the KG construction process, we identified that the bipartite graphs connecting samples, drugs, and genes share common sample nodes. This allows these graphs to be fused into a larger graph structure. Similarly, the mutation bipartite graph and the gene interaction graph share common gene nodes, enabling their integration into a unified graph with diverse edge types to form the KG. Directly consolidating these graphs is impractical due to storage constraints. Therefore, we adopt a triplet storage format. After merging genes and samples, we assign a unified ID to represent them as a consolidated entity while preserving the distinction between individual entities. The relationships between entities are documented in a structured subject-predicate-object format.

Assuming a KG $G = (E, R, T)$, where E denotes the set of entities, R denotes the set of relationships, and T denotes the set of triplets, we represent the three bipartite graphs as triplets in the KG:

$$\mathcal{T} = \left\{ (e_i, r_{ij}, e_j) \mid (i,j) \in X \wedge r_{ij} = \sum_{k \in \mathcal{R}} I\{X_{ij} = k\} \cdot k \right\}, \tag{3}$$

where $e_i, e_j \in E$ represent entities; $r_{ij} \in R$ represents relationships, determined by evaluating elements in the matrix X, where X denotes the drug response matrix, with rows representing patients and columns representing drugs; X_{ij} represents response state elements in the matrix X, combined with the set R to determine their relationships; $I\{X_{ij} = k\}$ is an indicator function that determines whether elements in the feature matrix belong to a specific relationship k. Specifically, the indicator function is used to distinguish between sensitive and resistant links between a patient and a drug. Specifically, if the IC50 value of drug d_j on patient x_i is below a predefined threshold, the indicator outputs 1 (i.e., sensitive), otherwise it outputs 0 (i.e., resistant). For the drug sensitivity or resistance bipartite graph, k is divided into two categories: sensitive or resistant. For the gene mutation bipartite graph, the number of mutations corresponds to the number of k. In the gene interaction bipartite graph, k either represents an interaction or its absence. Formula (3) demonstrates how to transform a feature matrix into a triplet representation of a KG. Through this process, the sets of entities E, relationships R, and triplets T are constructed, forming the core structure of the KG.

Ultimately, we convert the three distinct bipartite graphs into triplets. Concatenating these triplets results in a highly efficient data structure, facilitating the construction of a KG. We store the diverse types of information in triplet form, which inherently constitutes a KG. This graph is now ready for use in subsequent reasoning tasks.

D. Computability: KG Embedding

1) KG Embedding: After successfully constructing the KG, directly leveraging the graph for multi-step reasoning can still incur significant computational costs. To address this, we convert the graph structure into a vector format, which significantly improves the efficiency of subsequent reasoning tasks. The primary goal of embedding is to streamline complex data representations into a more compact and manageable format while retaining the essential features and structure of the original data. By converting the graph into vectors that approximate the connections in the original graph, we achieve a more efficient representation. However, two key challenges remain.

Challenge 1. Handling Unknown Triples: The original graph structure contains numerous known triples, but it cannot distinguish whether unknown triples do not exist or have simply not been sampled. To address this, we generate negative samples, a common training approach that involves creating non-existent triples using specific strategies, a process known as Negative Sampling. Negative Sampling not only addresses the issue of missing negative samples but also helps construct the objective function [30], enabling the model to more effectively distinguish between positive and negative examples. We define the objective function as:

$$L = \sum_{(h,r,t) \in S,\ (h',r,t') \in S'} [f_r(h,t) - f_r(h',t')]_+, \tag{4}$$

where h, r, t denote the head entity, relation, and tail entity, respectively, in the set of positive triples S, and h', t' denote the head and tail entities in negative triples S', which are generated by replacing h or t while retaining r; $[x]_+$ denotes a positive function, where $[x]_+ = x$ when $x > 0$ and $[x]_+ = 0$ when $x \leq 0$. In this formulation, the vector representation of the tail entity in a positive triple should be closer to the sum of the vector representations of the head entity and the relationship vector. Conversely, in a negative triple, the tail entity vector should be farther from this sum (details in Supplementary materials). Therefore, during the optimization of the objective function L, we aim to minimize its value for positive triples and maximize it for negative triples.

Challenge 2. Resolving Complex Relationships: The second challenge involves handling intricate relationships, such

as one-to-many, many-to-one, and many-to-many. Selecting an appropriate translation model for KG embedding is crucial. However, the score function $f_r(h,t)$ varies depending on the chosen translation model. For example, the score function of the TransE model is $f_r(h,t) = \| h + r - t \|$, where $h + r \approx t$ [30]. However, TransE can only handle one-to-one relationships and is unsuitable for one-to-many or many-to-one relationships [30]. The score function of the TransH model [31] can be written as $f_r(h,t) = \| (h - w_r^\top h w_r) + d_r - (t - w_r^\top t w_r) \|_2^2$ [31], where $(h - w_r^\top h w_r) + d_r \approx t - w_r^\top t w_r$. TransH can handle one-to-many, many-to-one, and many-to-many relationships. However, it assumes that entities and relations are vectors in the same semantic space, making it difficult to distinguish similar entities within the same space.

To address this, we map entities from the entity space to the relation space using a projection matrix M_r. Specifically, M_{rh} and M_{rt} are mapping matrices for entities h and t, respectively. The score function of the TransR model [32] is

$$f_r(h,t) = - \| M_{rh} h + r - M_{rt} t \|_2^2 . \tag{5}$$

However, due to its complexity, TransR is not suitable for large-scale KGs. To balance complexity and the ability to handle intricate relationships, we adopt the TransD model [33], whose score function is:

$$M_{rh_i} = r_p h_{ip}^\top + I^{m \times n} \tag{6}$$

$$M_{rt_i} = r_p t_{ip}^\top + I^{m \times n} \tag{7}$$

Here $I^{m \times n}$ is the identity matrix; h_{ip}, t_{ip} (for $i = 1, 2, 3$) and the relationship r_p are projection vectors.

2) Implementation of Embedding Process: In KG embedding, TransD [33] enhances the TransE model by introducing two separate spaces: one for entities and another for relations. In this framework, TransD independently embeds entities and relations within their respective spaces. The key innovation of TransD lies in modifying entity embeddings through the incorporation of relation embeddings. The scoring function calculates the distances between entities, considering the relations, and uses dynamic mapping to fine-tune these embeddings. This mapping ensures that entity embeddings align with the relational space. During training, the model iteratively refines these embeddings, reducing the distance between accurate and erroneous triples. Over time, this process enhances precision and more effectively captures intricate relationships.

E. Reasoning: Micro-Macro Joint Reinforcement Learning

1) Multi-Step Reasoning: The above process transforms entities and relationships into easily computable vector forms. For the facilitation of extended reasoning, we employ reinforcement learning algorithms. Within our KG, the vectors representing entities are conceptualized as states. We define the neighboring relationships of an entity with its adjacent entities as actions. We allocate a reward exclusively for the accurate final answer by using a soft reward strategy [34]. The agent initiates the reasoning process, considering the comprehensive KG as the fundamental environment. This strategy enables the execution of multi-step reasoning through reinforcement learning. Consequently, the pivotal stages of the reasoning process become clear to the user, thus providing interpretability to our approach. The schematic diagram of our MarMirDrug is shown in Fig. 1.

2) Long and Short-Range Reasoning: KG Embedding Data Mapped to Clusters: In our reasoning mechanism, we initially discount irrelevant areas to focus on relevant regions. We use pre-trained entity embeddings along with the K-Means algorithm to segment the KG into N clusters. Subsequently, we build a cluster connectivity graph that connects clusters if at least one entity-level edge is present. As the number of clusters is less than the number of entities, reasoning within this cluster graph helps us identify valid regions and avoids the problem of combinatorial explosion.

Macroscopic Intelligent Agent: Cluster-Level Exploration. To efficiently search and locate entities within the KG, the macroscopic intelligent agent first excludes ineffective clusters. The state of macroscopic intelligent agent is represented as a triplet $s_t^c = (c_t, c_s) \in S^c$, where $c_t \in G^c$denotes the cluster accessed at step t, encapsulating information relevant to the current state. c_s signifies the initial cluster containing the source entity, representing the global background shared across all states. At each step t, the possible actions of the macroscopic intelligent agent are composed of adjacent clusters to c_t in G_c.

The macroscopic intelligent agent traverses the KG data cluster by cluster. Since cluster-level paths are generally shorter than entity-level paths, a special "stop" action has been added to each action space A_t^c, $A_t^c = \{c' | (c_t, c') \in G^c\}$, to facilitate synchronization between the macroscopic and microscopic intelligent agents. Each action $a_t^c \in A_t^c$ is decided based on predictions from the macroscopic intelligent agent's policy network π_θ^c. After completing the necessary steps, if the macroscopic intelligent agent arrives at the cluster c_T containing the correct entity $e_T \in c_T$, it receives a terminal reward of 1; otherwise, the reward is 0. This reward structure aids the macroscopic intelligent agent in efficiently pinpointing the cluster with the correct answer, thereby narrowing the search space for the microscopic intelligent agent.

Microscopic Intelligent Agent: Entity-Level Exploration. The objective of the microscopic intelligent agent is to navigate within the cluster identified by the Macroscopic Intelligent Agent. This cluster contains the correct answer. The aim is to obtain the accurate answer, specifically the target entity. Each state is $s_t^e = (e_t, (e_s, r_q))$. A state is defined as a triplet composed of the currently accessed entity e_t, the query entity e_s , and the relationship r_q. These states form the foundation for the microscopic intelligent agent's path search within the KG. At each step t, to determine its action, the microscopic intelligent agent uses its policy network π_θ^e to predict and select among all outgoing edges from the current entity. The state set is $A_t^e = \{(r', e') | (e_t, r', e') \in G\}$. Within the maximum step limit, if the agent reaches a correct target entity, it receives a terminal reward of 1. Otherwise, the terminal reward is 0. Thus, the microscopic intelligent agent accurately locates the target entity.

3) Dual-Agent Collaboration: Cooperative Policy Networks: Agents make decisions based on the output of policy networks. To achieve precise navigation and decision-making of

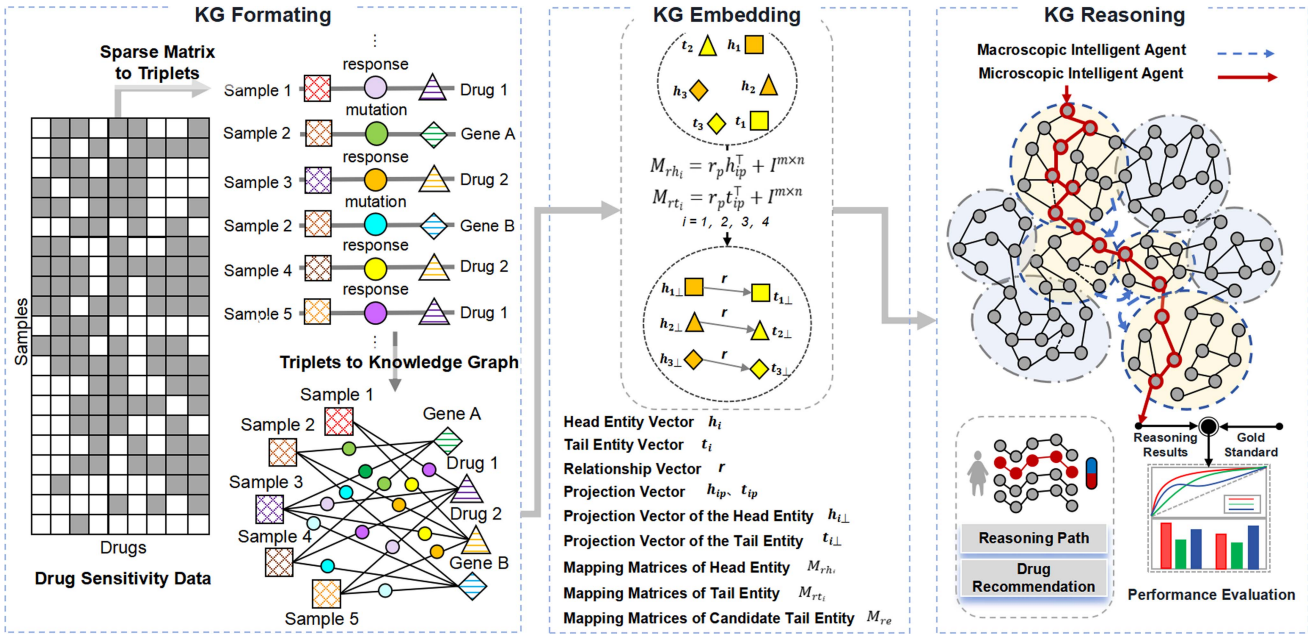


Fig. 1. The schematic diagram of our proposed Macro-Micro agent Drug sensitivity inference (MarMirDrug).

agents within the KG, we employ two separate policy networks π_θ^e and π_θ^c to simulate the action selection of microscopic and macroscopic intelligent agents. We assign dense *embedding* vectors $e \in \mathbb{R}^d$, $r \in \mathbb{R}^d$, and $c \in \mathbb{R}^{2d}$ to each entity, relationship, and cluster in G^c, respectively, for representing the states and actions in reinforcement learning of microscopic and macroscopic intelligent agents. Detailly, the embedding of cluster G^c is computed by averaging the embeddings of all nodes within the cluster. For the microscopic intelligent agent, each action consists of the next outgoing relationship and entity, denoted as $a_t^e = [r_t; e_t] \in A_t^e$, as the concatenation of the relationship embedding r_t and the embedding of the terminal node e_t.

For the macroscopic intelligent agent, actions $a_t^c = c_t \in \mathbb{R}^{2d}$ correspond to the next outgoing cluster, represented directly using the cluster embedding. For π_θ^e and π_θ^c, initially, two independent LSTMs are used to encode their search histories $h_t^c = (c_1, \ldots, c_t) \in H^c$, $h_t^e = (e_s, r_1, e_1, \ldots, r_t, e_t) \in H^e$, based on recursive dynamics,

$$h_t^c = \mathrm{LSTM}_c\left(W^c\left[h_{t-1}^c; h_{t-1}^e\right], \mathrm{a}_{t-1}^c\right), t > 0, \tag{8}$$

$$h_t^e = \mathrm{LSTM}_e\left(W^e\left[h_{t-1}^e; h_{t-1}^c\right], \mathrm{a}_{t-1}^e\right), t > 0, \tag{9}$$

where modifying their internal structures to allow sharing of states between macroscopic intelligent agent's $\mathrm{LSTM}_c$and microscopic intelligent agent's LSTM_e. Specifically, at each $t > 0$ step, we compute the concatenation of two original states h_t^c, $h_t^e \in \mathbb{R}^{2d}$ as the new state $[h_t^c; h_t^e]$, $[h_t^e; h_t^c] \in \mathbb{R}^{4d}$. Furthermore, to prevent exponential growth in dimensionality with increasing steps, we applied two transformation matrices $\mathrm{W^c}$, $\mathrm{W^e} \in \mathbb{R}^{2\mathrm{d}\times 4\mathrm{d}}$ to reduce the dimensionality of the new states. Through modification, each hidden state h_t^c/h_t^e of the agent is conditioned on its own previous state h_{t-1}^c/h_{t-1}^e, the previous state $h_{t-1}^e/h_{t-1}^c$of the other agent, and its previous action a_{t-1}^c/a_{t-1}^e. This ensures that important path information is shared between the macroscopic and microscopic intelligent agents, enhancing the efficacy of their action selections.

We apply a two-layer feedforward network to predict the next cluster for the macroscopic intelligent agent and the next relationship-entity edge for the microscopic intelligent agent. It performs ReLU non-linear analysis on the concatenation of their final LSTM states and the current reinforcement learning state embeddings,

$$\pi_\theta^c\left(a_t^c | s_t^c\right) = \sigma\left(A_t^c \times W_2^c \mathrm{ReLU}\left(W_1^c\left[c_t; h_t^c\right]\right)\right) \tag{10}$$

$$\pi_\theta^e\left(a_t^e | s_t^e\right) = \sigma\left(A_t^e \times W_2^e \mathrm{ReLU}\left(W_1^e\left[e_t; r_q; h_t^e\right]\right)\right) \tag{11}$$

where W_1^c, $W_2^c \in \mathbb{R}^{4d\times 4d}$ and W_1^e, $W_2^e \in \mathbb{R}^{6d\times 6d}$ are matrices of learnable network weights, and the symbols $A_t^c \in \mathbb{R}^{|A_t^c|\times 4d}$, $A_t^e \in \mathbb{R}^{|A_t^e|\times 6d}$ represent embeddings of all possible next actions for the macroscopic and microscopic intelligent agents, respectively. σ denotes the softmax operator. This method ensures efficient path exploration and decision-making in the KG for both microscopic and macroscopic intelligent agents, enhancing their action selection capabilities through shared critical path information.

Mutual Reinforcement Reward: Conventional reward schemes for macroscopic and microscopic intelligent agents focus mainly on reaching target clusters or entities, presenting two principal challenges: 1) Navigating denser cluster graphs challenges macroscopic intelligent agents in aligning with entity-level paths. 2) Microscopic intelligent agents fail to gather stage-specific information from macroscopic intelligent agents. To tackle these challenges, we propose a new mutual reinforcement reward mechanism. This mechanism ensures that the trajectories of macroscopic intelligent agents correspond with entity-level paths and enables the transfer of stage-specific

cues from macroscopic to microscopic intelligent agents. The final reward for both agent types includes their inherent rewards, increased by an additional weighted reward factor based on their counterparts.

$$R_c\left(s_t^c\right) = \underbrace{r_c\left(s_T^c\right)}_{default\ reward} + \underbrace{\Phi\left(s_t^c, s_t^e\right) \cdot r_e\left(s_T^e\right)}_{partner\ reward}, t \in [1, T] \tag{12}$$

$$R_e\left(s_t^e\right) = \underbrace{r_e\left(s_T^e\right)}_{default\ reward} + \underbrace{\Phi\left(s_t^e, s_t^c\right) \cdot r_c\left(s_T^c\right)}_{partner\ reward}, t \in [1, T] \tag{13}$$

where $r_c(s_T^c)$ and $r_e(s_T^e)$ (details in supplementary materials) denote the default final rewards for the macroscopic and microscopic intelligent agents. Additionally, $\Phi(s_t^c, s_t^e)$ (details in supplementary materials) is an evaluation function that measures the consistency of actions taken by both agents. In practice, $\Phi(s_t^c, s_t^e)$ is computed as the cosine similarity between the pretrained embeddings of the current traversed cluster and entity.

For the macroscopic intelligent agent, the partner reward applies only if the microscopic intelligent agent reaches the target entity and the visited cluster is near the entity at step t. For the microscopic intelligent agent, the reward is given only when the macroscopic intelligent agent reaches the target cluster and the consistency weight is sufficient. This setup ensures that both agents earn rewards only when their partners achieve the correct target. The metric coefficient $\Phi(s_t^c, s_t^e)$ adjusts the partner rewards based on the overlap between cluster-level and entity-level paths. The new mutual reinforcement reward mechanism enables effective collaboration between macroscopic and microscopic intelligent agents. It ensures that their trajectories are aligned with entity-level paths and that they provide essential phase-specific cues to each other when needed.

IV. Results

A. Experimental Setup

To validate the interpretability of our method in drug-related reasoning, we conduct comprehensive experiments using a pharmacogenomic KG. The graph, constructed from GDSC data [15], comprises 39825 nodes representing three entity types: genes, samples, and drugs. These entities are interconnected through 14 distinct relationship types, forming a complex network of 510382 triples. The details of hyperparameter setting of our methods are in supplementary materials. For experimental validation, we implement a five-fold cross-validation strategy, utilizing four folds for training and the remaining fold for testing across all scenarios. We employ Hit score metrics to evaluate the model's performance across all drugs in the graph. Additionally, to assess individual drug prediction accuracy, we perform ROC curve analysis on specific drug responses. This experimental framework effectively evaluates the model's capability in predicting personalized drug responses, establishing a robust foundation for its potential applications in precision medicine.

B. Comparison of KG Embedding

1) Embedding Impact on Performance: The direct processing of graph structures presents significant computational challenges due to their inherent complexity, which often results in substantial resource consumption during analysis. To address this limitation, we transform the graph into a vector-based representation for more efficient processing. By leveraging embedding vectors instead of the original graph structure, our MarMirDrug framework can be easily applied in solving path inference tasks. To evaluate the effectiveness of these embeddings in predicting drug sensitivity, we conduct a comparative analysis of four embedding techniques. As illustrated in Fig. 2, all four techniques exhibit stable performance as the number of training epochs increases. In the five-fold cross-validation, TransE achieve comparable hit@10 scores and mean ranks in comparison to those of TransH, serving as baseline models. While TransD and TransR show similar predictive performance, TransD is more computationally efficient (detailed analysis is provided in the following subsection). The selected TransD model significantly outperforms the baselines (TransE), achieving a 117.65% improvement in average hit@10 score and reducing the mean rank by 57.24% in the five-fold cross-validation. Further comparisons also illustrate that, although state-of-the-art embedding methods ComplEx [35], RotatE [36], GraphSAGE [37] demonstrate strong performances, TransD maintains comparable or superior accuracy while significantly reducing parameter complexity. (details in supplementary materials). These results demonstrate that TransD is the effective approach and is selected for subsequent inference tasks.

2) Computational Efficiency: The computational efficiency of graph embedding is critical for optimizing resource utilization. To evaluate this aspect, we conduct a comprehensive comparison of training time and memory consumption between TransD and TransR, both of which demonstrate strong performance in benchmark tests. As illustrated in Fig. 3, while both models maintain low overall loss during training and validation, TransD demonstrates superior performance through its implementation of early stopping criteria, effectively alleviating overfitting [33]. This phenomenon makes TransD particularly suitable for practical applications such as drug sensitivity prediction. Our experimental results reveal two advantages of TransD: first, it achieves faster convergence compared to TransR, significantly reducing the number of iterations required to reach stable performance and thereby enhancing time efficiency. Second, as shown in Supplementary Fig. S1, TransD demonstrates substantially lower memory consumption, using approximately half the memory required by TransR. These findings collectively establish TransD's superiority over TransR in both computational speed and memory efficiency, making it a more practical choice for large-scale applications.

C. Comparison of Multi-step Reasoning

1) Performance Trend Across Episodes: The aforementioned graph embedding serves as an effective vector representation of the state space, enabling the depiction of reasoning paths through state transitions. Based on our embedding analysis, we

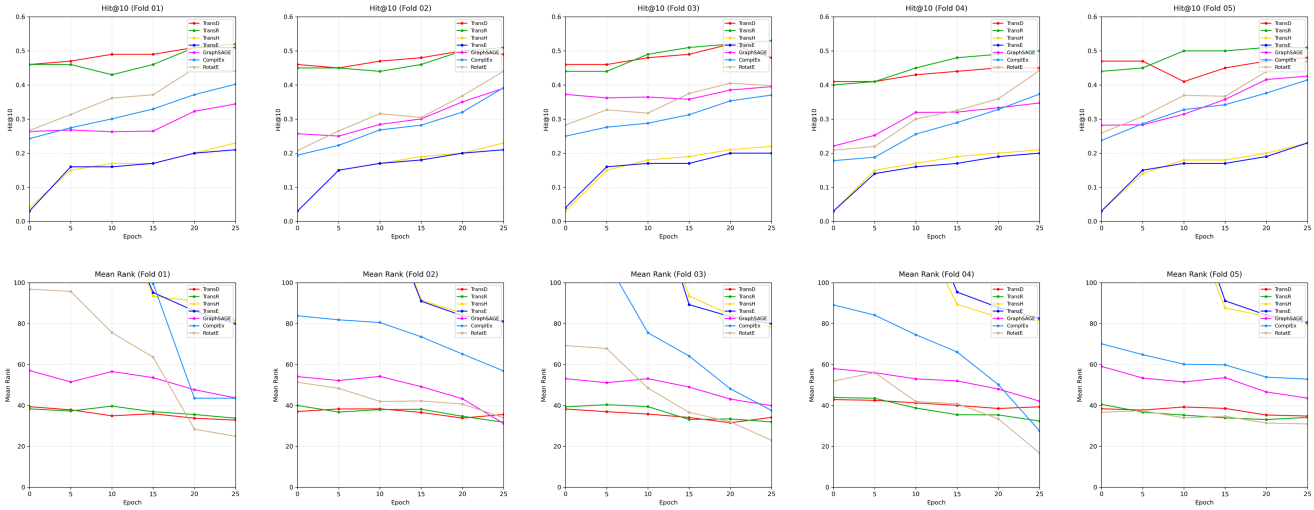


Fig. 2. Comparison of Hit@10 and mean rank for seven embedding benchmarks in five-fold cross-validation.

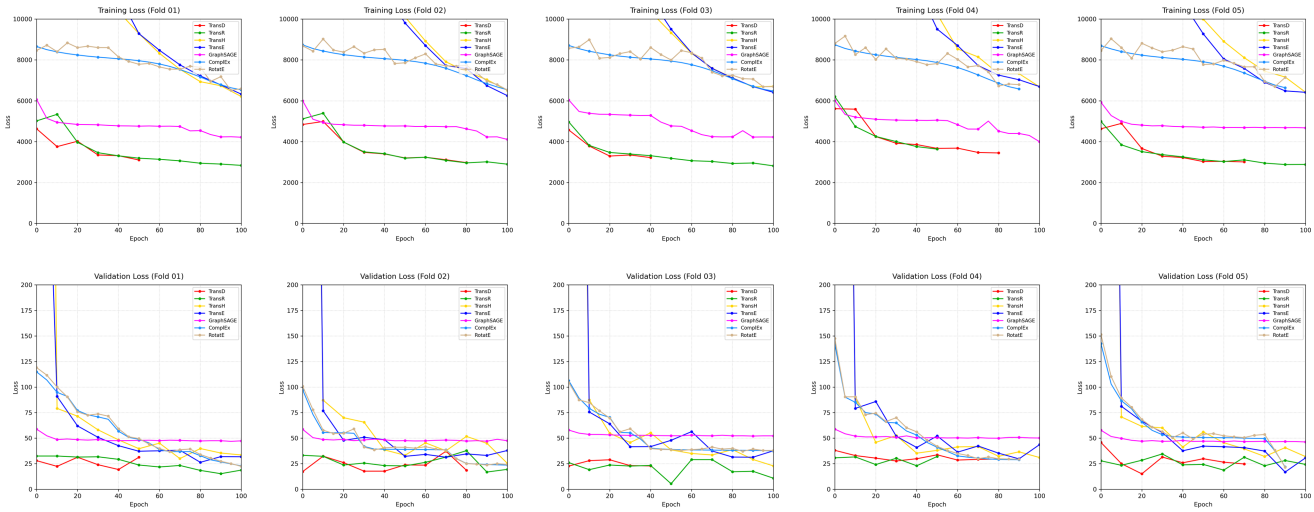


Fig. 3. Comparison of training total loss and validating total loss for seven embedding benchmarks in five-fold cross-validation. Due to early stopping, TransD training ended earlier when its validation performance converged.

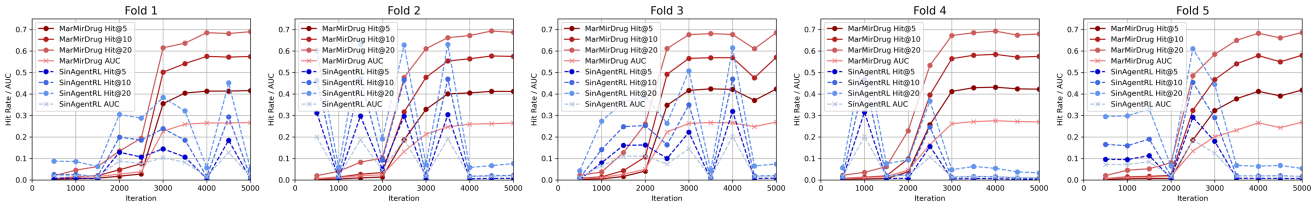


Fig. 4. Comparison of Hit scores and AUCs between MarMirDrug and SinAgentRL in five-fold cross-validation.

select TransD as the vector space representation for reinforcement learning states. In explainable reasoning, unlike conventional single-agent reinforcement learning [23] (SinAgentRL, also the framework of our single-agent ablation), our proposed MarMirDrug employs a novel dual-agent framework. To systematically evaluate their performance, we compare SinAgentRL-based models with our MarMirDrug framework, tracking their performance trends across training episodes. As illustrated in Fig. 4, the reinforcement learning models demonstrate distinct performance patterns throughout the training process. Our MarMirDrug framework exhibits a significant performance improvement around the 3,000th iteration, reaching its peak near the 5,000th iteration. In contrast, SinAgentRL achieves its optimal performance between 2000 and 3500 iterations, followed

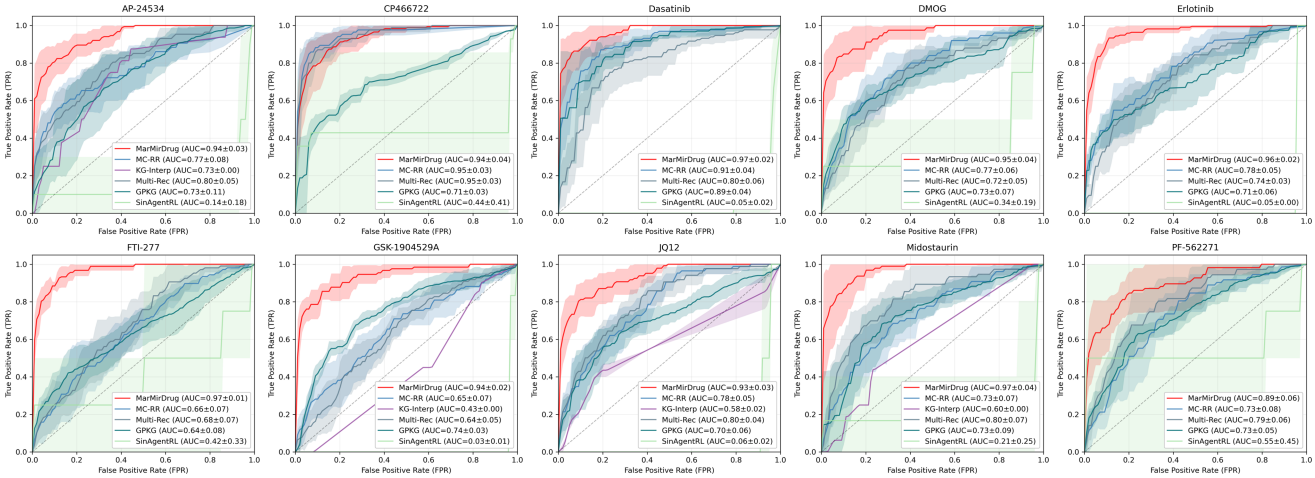


Fig. 5. ROC comparisons of MarMirDrug, SinAgentRL, and four existing methods across selected drugs by AUC with standard deviation shading.

by a rapid performance degradation. Here SinAgentRL shows rapid degradation, which may be due to overfitting suboptimal paths or getting trapped in early policy convergence. Notably, MarMirDrug consistently outperforms SinAgentRL after 2500 iterations. Based on these observations, we establish 5000 iterations as the selected training threshold, effectively balancing model performance and computational efficiency.

2) Evaluation on Drug Response Inference: Compared to the Hit Score metric, ROC curve analysis provides a more comprehensive evaluation framework for individual drug prediction performance. Within this framework, we conduct a comparative evaluation not only between SinAgentRL and our MarMirDrug approach, but also four representative related works, i.e., similarity-based method MC-RR [5], KG interpretability related method KG-Interp [16], graph embedding based computational efficient method Multi-Rec [21], and multi-step predicting framework based GPKG [28]. This experimental comparison can quantitatively demonstrate our method's advantages in interpretability, computational efficiency, and multi-step reasoning. Recognizing the challenge posed by imbalanced positive and negative sample distributions across drugs, we select the top-ten balanced drugs for detailed analysis. For these drugs, we compute ROC curves and calculate their corresponding Area Under the Curve (AUC) values for the compared algorithms. In Fig. 5, taking AP-24534 drug recommendation as an example, all methods except the ablation SinAgentRL demonstrate notable strengths. Specifically, MC-RR, KG-Interp, Multi-Rec, and GPKG achieve mean AUCs of 0.77, 0.73, 0.80, and 0.77 across folds, respectively, by leveraging similarity matching, interpretable pruning, graph-based representation, and multi-step reasoning. In contrast, our MarMirDrug framework further improves performance, achieving a mean AUC of 0.94 through a two-stage reasoning strategy. The average ROC curves a cross all selected drugs are provided in Supplementary Fig. S2. In contrast to GNN based methods Multi-Rec and GPKG, MarMirDrug employs dual-agent framework to explicitly model multi-hop reasoning paths while mitigating path explosion. This strategy effectively constrains the search space, enabling MarMirDrug

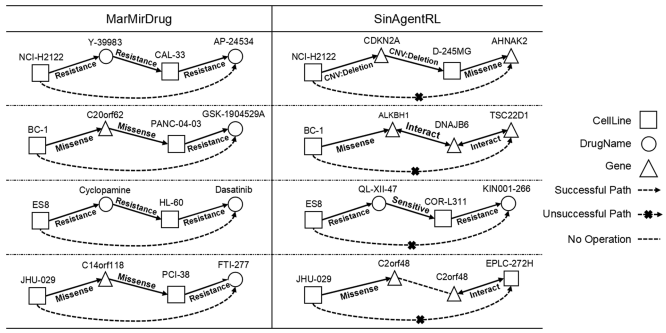


Fig. 6. Case study of reasoning paths of MarMirDrug and SinAgentRL.

to achieve an impressive average AUC of approximately 94% . Moreover, we further applied our method to chemotherapy drug response data from TCGA [38], and the results yielded multiple clinically interpretable reasoning paths (see supplementary materials and Supplementary Table S1). These results demonstrate MarMirDrug's superior performance in drug recommendation tasks.

D. Case Study

To further validate the explainability of MarMirDrug, we select four reasoning paths as examples in case study, which successfully reach the target within our framework (Fig. 6). Notably, these paths consistently fail to achieve target results when implemented in SinAgentRL. MarMirDrug's path discovery operates under two fundamental principles: 1) when samples A and B demonstrate similar drug sensitivity/resistance patterns, we infer comparable pharmacological responses between them; 2) when samples A and B share common gene mutations, we predict analogous drug responses based on their genetic similarity. In contrast, the single-agent approach lacks this structured reasoning framework, often resulting in inconsistent and potentially erroneous path discoveries. Taking the first MarMirDrug path as an example, it connects a cell line to a drug via shared

sensitivity or resistance profiles. NCI-H2122 is resistant to drug Y-39983, and CAL-33 also shows resistance to Y-39983. Since CAL-33 is also resistant to AP-24534, MarMirDrug infers that NCI-H2122 will likewise resist AP-24534. This mirrors expert practice, that clinicians compare a patient's drug response profile with similar cases to guide treatment predictions. Thus, the reasoning paths can improve the framework's practical utility in clinical decision-making.

V. Discussion and Conclusion

This study focuses on enhancing the interpretability of anticancer drug sensitivity prediction in precision medicine. While existing methods have made strides in improving prediction accuracy, they still face three unignorable challenges: 1) the "black-box" problem, which results in a lack of transparency in the reasoning process; 2) the high computational complexity of graph structures when handling large-scale biomedical data; 3) the combinatorial explosion issue in multi-step reasoning. To address these challenges, we propose a dual-agent reinforcement learning algorithm based on KG embedding. Our approach leverages KG to enhance interpretability, employs embedding techniques to transform complex graph structures into low-dimensional vectors for reduced computational complexity, and introduces a dual-agent algorithm to mitigate the combinatorial explosion problem. Experimental results demonstrate significant improvements, where our dual-agent algorithm MarMirDrug shows higher accuracy and interpretability in drug response prediction. This method combines interpretability, efficiency, and accuracy, offering new insights for precision medicine.

The success of this study stems from three key advantages. First, we employ a KG to construct reasoning paths, which, compared to traditional deep learning "black-box" models, provides a clear illustration of the relationship across drug mechanism entities, offering high interpretability. Second, we utilize the embedding model for KG, which not only achieves higher prediction accuracy but also significantly reduces computational complexity. Third, we introduce a macro-micro dual-agent collaborative reinforcement learning mechanism, effectively addressing the combinatorial explosion problem in multi-step reasoning that is prevalent in traditional single-agent approaches. This significantly enhances reasoning efficiency and accuracy. The path inference process, though theoretically exponential in complexity, is made tractable through fixed-depth reasoning and a dual-agent framework that effectively limits the action space (details in supplementary materials). Instead of attention based neighbor aggregation for node representations [13], MarMirDrug employs a dual-agent reinforcement learning framework to model and prune multi-hop reasoning paths, yielding transparent stepwise explanations and controlling the combinatorial explosion. These advantages make our method more efficient and scalable in processing large-scale biomedical data, providing robust technical support for precision medicine research.

Regarding biological significance, compared with the general dual-agent KG reasoning framework proposed in [12], which is primarily designed for general-purpose knowledge graphs such as product recommendation and multi-hop relational reasoning across diverse entities, our method introduces domain-specific adaptations tailored for biomedical knowledge graphs: 1) gene–gene interactions are incorporated to supplement structural information among genes and enhance semantic continuity along the paths; 2) the sparsity of drug–gene relationships is alleviated, improving the overall connectivity for reasoning; 3) with a more connected graph structure, entity clustering becomes more effective in compressing the search space and mitigating path explosion. Together, these three aspects enhance both the reasoning performance and biological interpretability of our model in drug sensitivity tasks.

Despite these achievements, our study also has some limitations. First, the data sources are currently limited to the GDSC database, and the model's generalizability requires further validation. Future work will need to expand data sources to enhance robustness. Second, the drug-related information is primarily limited to genomic data, lacking multi-omics data such as transcriptomics and proteomics. Integrating more biological information in the future will improve predictive capabilities. For example, recent studies [39], [40] further highlight that both inter- and intra-molecular structures can be exploited to enhance drug prediction performance and explainability. Third, the computational efficiency of the algorithm needs further optimization to enhance its clinical applicability. Also, GDSC exhibits class imbalance, which we mitigate by uniformly allocating each drug's samples across five folds, whereas iRefIndex serves only as side information and thus involves no such imbalance (details in supplementary materials). Although there is room for improvement in data samples, functional dimensions, and computational efficiency, this study has made important progress in the interpretability of anticancer drug sensitivity prediction, contributing three key innovations: interpretability, efficient computation, and long-path reasoning. These advancements provide new insights for precision medicine. Future research will focus on data expansion, multi-omics integration, and computational optimization to drive technological progress in the field. Overall, this study offers a novel technical pathway for anticancer drug sensitivity prediction and lays a crucial foundation for the further development of precision medicine.

References

[1] A. El-Hussein, S. L. Manoto, S. Ombinda-Lemboumba, Z. A. Alrowaili, and P. Mthunzi-Kufa, "A review of chemotherapy and photodynamic therapy for lung cancer treatment," *Anti-Cancer Agents Med. Chem.*, vol. 21, no. 2, pp. 149–161, 2021.

[2] G. Gambardella, G. Viscido, B. Tumaini, A. Isacchi, R. Bosotti, and D. di Bernardo, "A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response," *Nature Commun.*, vol. 13, no. 1, Mar. 2022, Art. no. 1714.

[3] J. de Jong et al., "Towards realizing the vision of precision medicine: AI based prediction of clinical drug response," *Brain*, vol. 144, pp. 1738–1750, Jun. 2021.

[4] J. Mateo et al., "Delivering precision oncology to patients with cancer," *Nature Med.*, vol. 28, no. 4, pp. 658–665, Apr. 2022.

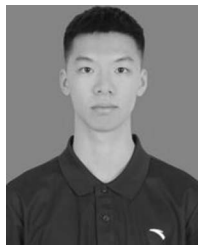
[5] C. Y. Liu et al., "An improved anticancer drug-response prediction based on an ensemble method integrating matrix completion and ridge regression," *Mol. Ther.-Nucleic Acids*, vol. 21, pp. 676–686, Sep. 2020.

[6] E. W. Huang, A. Bhope, J. Lim, S. Sinha, and A. Emad, "Tissue-guided LASSO for prediction of clinical drug response using preclinical samples," *Plos Comput. Biol.*, vol. 16, no. 1, Jan. 2020, Art. no. e1007607.

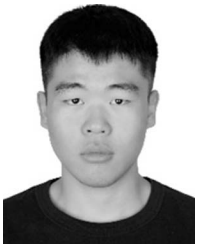
[7] Y. Tao, S. Ren, M. Q. Ding, R. Schwartz, and X. Lu, "Predicting drug sensitivity of cancer cell lines via collaborative filtering with contextual attention," in *Proc. 5th Mach. Learn. Healthcare Conf.*, 2020, pp. 660–684.
[8] R. Su, Y. X. Huang, D. G. Zhang, G. B. Xiao, and L. Y. Wei, "SRDFM: Siamese response deep factorization machine to improve anti-cancer drug recommendation," *Brief. Bioinf.*, vol. 23, no. 2, Mar. 2022, Art. no. bbab534.
[9] J. Rueda, J. D. Rodríguez, I. P. Jounou, J. Hortal-Carmona, T. Ausín, and D. Rodríguez-Arias, ""Just" accuracy? Procedural fairness demands explainability in AI-based medical resource allocations," *AI Soc.*, vol. 2022, pp. 1–12, Dec. 21, 2022.
[10] R. ElShawi, Y. Sherif, M. Al-Mallah, and S. Sakr, "Interpretability in healthcare: A comparative study of local machine learning interpretability techniques," *Comput. Intell.*, vol. 37, no. 4, pp. 1633–1650, Nov. 2021.
[11] F. Gong, M. Wang, H. F. Wang, S. Wang, and M. Y. Liu, "SMR: Medical knowledge graph embedding for safe medicine recommendation," *Big Data Res.*, vol. 23, Feb. 2021, Art. no. 100174.
[12] D. Zhang, Z. Yuan, H. Liu, X. Lin, and H. Xiong, "Learning to walk with dual agents for knowledge graph reasoning," in *Proc. AAAI Conf. Artif. Intell.*, 2021.
[13] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "KGAT: Knowledge graph attention network for recommendation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 950–958.
[14] A. Gogleva et al., "Knowledge graph-based recommendation framework identifies drivers of resistance in EGFR mutant non-small cell lung cancer," *Nature Commun.*, vol. 13, no. 1, Mar. 2022, Art. no. pp. 1667.
[15] W. J. Yang et al., "Genomics of drug sensitivity in cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells," *Nucleic Acids Res.*, vol. 41, no. D1, pp. D955–D961, Jan. 2013.
[16] M. Caro-Martínez, G. Jiménez-Díaz, and J. A. Recio-Garcia, "A graph-based approach for minimising the knowledge requirement of explainable recommender systems," *Knowl. Inf. Syst.*, vol. 65, no. 10, pp. 4379–4409, 2023.
[17] Z. Y. He, C. D. Wang, J. Wang, J. H. Lai, and Y. Tang, "Community enhanced knowledge graph for recommendation," *IEEE Trans. Comput. Social Syst.*, vol. 11, no. 5, pp. 5789–5802, Oct. 2024.
[18] C. Wang, H. Zhang, L. Li, and D. Li, "Knowledge graph attention network with attribute significance for personalized recommendation," *Neural Process. Lett.*, vol. 55, no. 4, pp. 5013–5029, 2023.
[19] U. Zulaika, A. Almeida, and D. López-de-Ipiña, "Regularized online tensor factorization for sparse knowledge graph embeddings," *Neural Comput. Appl.*, vol. 35, no. 1, pp. 787–797, 2023.
[20] S. Forouzandeh, K. Berahmand, and M. Rostami, "Presentation of a recommender system with ensemble learning and graph embedding: A case on MovieLens," *Multimedia Tools Appl.*, vol. 80, no. 5, pp. 7805–7832, 2021.
[21] H. Shu and J. Huang, "Multi-task feature and structure learning for user-preference based knowledge-aware recommendation," *Neurocomputing*, vol. 532, pp. 43–55, 2023.
[22] X. Wang, X. He, and T.-S. Chua, "Learning and reasoning on graph for recommendation," in *Proc. 13th Int. Conf. Web Search Data Mining*, 2020, pp. 890–893.
[23] R. Das et al., "Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning," in *Proc. 6th Int. Conf. Learn. Representations*, 2018.
[24] X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T.-S. Chua, "Explainable reasoning over knowledge graphs for recommendation," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 5329–5336.
[25] D. Wu, M. Tang, S. Zhang, A. You, and W. Gao, "KPRLN: Deep knowledge preference-aware reinforcement learning network for recommendation," *Complex Intell. Syst.*, vol. 9, no. 6, pp. 6645–6659, 2023.
[26] Y. C. Yang, C. T. Chen, T. Y. Lu, and S. H. Huang, "Hierarchical reinforcement learning for conversational recommendation with knowledge graph reasoning and heterogeneous questions," *IEEE Trans. Serv. Comput.*, vol. 16, no. 5, pp. 3439–3452, Sep./Oct. 2023.
[27] Y. Li, H. Chen, Y. Li, L. Li, P. S. Yu, and G. Xu, "Reinforcement learning based path exploration for sequential explainable recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 11801–11814, Nov. 2023.
[28] Z. Gao, P. Ding, and R. Xu, "KG-Predict: A knowledge graph computational framework for drug repurposing," *J. Biomed. Inform.*, vol. 132, Aug. 2022, Art. no. 104133.
[29] S. Razick, G. Magklaras, and I. M. Donaldson, "iRefIndex: A consolidated protein interaction database with provenance," *BMC Bioinf.*, vol. 9, Sep. 2008, Art. no. 405.
[30] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2013, pp. 2787–2795.
[31] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Proc. 28th AAAI Conf. Artif. Intell.*, Québec City, Québec, Canada, 2014, pp. 1112–1119.
[32] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proc. 29th AAAI Conf. Artif. Intell.*, Austin, TX, 2015, pp. 2181–2187.
[33] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, "Knowledge graph embedding via dynamic mapping matrix," in *Proc. Annu. Meeting Assoc. Comput. Linguistics*, 2015, pp. 106–118.
[34] A. Ma, Y. H. Yu, C. Shi, S. Zhen, L. Pang, and T. S. Chua, "PMHR: Path-based multi-hop reasoning incorporating rule-enhanced reinforcement learning and KG embeddings," *Electronics*, vol. 13, no. 23, Dec. 2024, Art. no. 2024.
[35] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2071–2080.
[36] Z. Sun, Z. Deng, J. Nie, and J. Tang, *RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space*. Alameda, CA, USA: OpenReview.net, 2019.
[37] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.
[38] T. C. G. A. Network, "Comprehensive molecular portraits of human breast tumours," *Nature*, vol. 490, no. 7418, pp. 61–70, 2012.
[39] S. Wang et al., TRGH-PPI: Effective and Generalized Prediction of Protein-protein Interactions through Transformer and Graph, 2025.
[40] W. J. Du et al., "Molecular merged hypergraph neural network for explainable solvation Gibbs free energy prediction," *Research*, vol. 8, Aug. 15, 2025, Art. no. 0740.

Minhua Feng received the bachelor's degree from Guangzhou Medical University, in 2025. Now she is currently working toward the master's degree with Shenzhen University. Her research interests include medical artificial intelligence and bioinformatics.

Liping Tang received the bachelor's degree from Guangzhou Medical University, in 2025. Now she is currently working toward the master's degree with Shenzhen University. Her research interests include medical artificial intelligence and bioinformatics.



Juntao Liang received the bachelor's degree from Guangzhou Medical University, in 2025. Now he is an algorithm engineer with Guangzhou Laboratory. His research interests include medical artificial intelligence and bioinformatics.



Song Huang is working toward the bachelor's degree with Guangzhou Medical University. His research interests include biomedical signal processing, biometrics, and bioinformatics.

Jianfeng Ma received the bachelor's degree from Guangzhou Medical University, in 2025. Now he is currently working toward master's degree with Sun Yat-sen University. His research interests include medical artificial intelligence, bioinformatics, and flexible electronics.

Zhimin Zheng received the bachelor's degree from Guangzhou Medical University, in 2025. She is now an assistant engineer wth Guangzhou Zenith Lab LTD. Her research interests include medical artificial intelligence and bioinformatics.

Ranran Guo received the BS degree from Sichuan University, in 2014, and the PhD degree from Fudan University, in 2019. She is currently a lecturer with Guangzhou Medical University, and her primary research focuses on the design of functional nanomaterials and their application in immunomodulation for diseases such as cancer.

Wen Shi (Member, IEEE) received the bachelor's degree from Sun Yat-sen University, in 2016, and the PhD degree from the South China University of Technology, in 2021. She is now a lecturer with Guangzhou Medical University. Her current research interests include evolutionary computation algorithms and their applications on medical optimization problems.

Jianing Xi (Senior Member, IEEE) received the BS degree and the PhD degree from University of Science and Technology of China, in 2013 and 2018 respectively. In 2018-2019, he was in Xidian University, China. In 2019-2022, he is in Northwestern Polytechnical University, China. Currently, he is an associate professor with Guangzhou Medical University. His research interests include bioinformatics.