MIND-DDI: Adaptive Graph-Neural-Symbolic Joint Optimization for Multi-Omics Interpretable Drug-Drug Interaction Prediction

Linxin Xiao
DCST, Tsinghua University
Beijing, China
xlx21@mails.tsinghua.edu.cn

Xin Wang*
DCST, BNRist, Tsinghua University
Beijing, China
xin_wang@tsinghua.edu.cn

Yang Yao
DCST, Tsinghua University
Beijing, China
yaoyang21@tsinghua.org.cn

Wenwu Zhu*
DCST, BNRist, Tsinghua University
Beijing, China
wwzhu@tsinghua.edu.cn

Abstract

Multi-omics drug-drug interaction (DDI) prediction forecasts DDI outcomes from heterogeneous multi-omics data and complex structured rules, where multi-omics interpretability of hierarchical biological factor interactions is extremely crucial, which remains unexplored in literature. In this paper, we are the first to study the problem of *Multi-Omics Interpretable DDI Prediction*, which is highly non-trivial with three challenges: i) how to identify interaction-relevant biological entities from large-scale knowledge sources, which lay the foundation for interpretability; ii) how to capture multi-omics cross-modal connections that elucidate underlying mechanisms under diverse interaction contexts; and iii) how to explicitly enable mechanism-aware reasoning over structured biological rules. To tackle the challenges, we propose a **M**ulti-omics **I**nterpretable prediction framework with adaptive graph-**N**eural-symbolic joint optimization to han**D**le **DDI** prediction (**MIND-DDI**), which follows a novel interpretable learning paradigm that tightly couples biomedical graph sampling, hierarchical neural representation learning, and symbolic rule modeling under a unified optimization scheme. Specifically, we propose a curriculum-guided biological subgraph sampler for progressively constructing informative subgraphs, then introduce a hierarchical multi-omics neural architecture search module to capture multi-omics alignment as well as cross-modal interactions, and finally present a mechanism-aware pathway-based symbolic reasoner for multi-hop reasoning over a unified neural-symbolic space. Extensive experiments show the superiority of **MIND-DDI** in both predictive accuracy and interpretability.

1 Introduction

Drug–drug interaction (DDI), driven by complex interaction patterns between drug structures and complex biological systems [1], plays a crucial role in clinical practice and drug development, given that the concomitant administration of multiple drugs may lead to altered therapeutic efficacy or unexpected adverse effects [2, 3]. Multi-omics DDI prediction aims to forecast DDI outcomes based on heterogeneous and multi-modal information that is derived from multi-omics data and structured rule-based mechanisms across multiple omics.

*Corresponding authors: Xin Wang and Wenwu Zhu. DCST denotes Department of Computer Science and Technology. BNRist denotes Beijing National Research Center for Information Science and Technology.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

In addition to prediction accuracy, interpretability, which enables transparent and logically grounded reasoning to support trustworthy decision-making [4, 5], serves as an urgent demand for (multi-omics) DDI prediction. As such, *Multi-Omics Interpretable DDI* prediction, which aims to explain how hierarchical multi-omics information interacts with each other and jointly makes contributions when predicting DDI, is becoming an increasingly important problem.

However, among the many recent advances on computational DDI modeling with machine learning approaches such as graph neural networks [6–8], e.g., representing drugs through molecular graph structures [9, 10] and incorporating diverse biomedical knowledge [11, 12], no one has proposed to model hierarchical multi-omics dependencies and structured interaction mechanisms, thus suffering from low multi-omics DDI predictive accuracy and deteriorated multi-omics DDI interpretability.

In this paper, we are the first to study the problem of *Multi-Omics Interpretable DDI Prediction*, to the best of our knowledge. Nevertheless, conducting multi-omics interpretable drug-drug interaction prediction is highly non-trivial, posing several challenges:

- How to identify interaction-relevant biological entities from large-scale biomedical graphs, so that the constructed subgraphs may provide a faithful basis for multi-omics interpretability?
- How to adaptively construct heterogeneous and multi-modal drug-related graphs via capturing multi-omics cross-modality connections and fine-grained omics interactions within a unified yet flexible neural architecture?
- How to discover biologically meaningful symbolic spaces that encode multi-omics drug mechanisms and further uncover new mechanism-driven relationships through interpretable reasoning over these spaces?

To tackle these challenges, we propose a **M**ulti-omics **I**nterpretable prediction framework with adaptive graph-**N**eural-symbolic joint optimization to han**D**le **D**rug-**D**rug **I**nteraction prediction (**MIND-DDI**), following a novel interpretable learning paradigm which employs neural-symbolic reasoning and unifies neural representation learning with structured, rule-based relational inference [13, 14].

In concrete terms, we first propose a *Curriculum-guided Biological Subgraph Sampler*, which progressively identifies informative subgraphs from large-scale biomedical graphs, effectively filtering out noisy and redundant information, thereby grounding model predictions in interpretable biological evidence. We then introduce *Hierarchical Multi-omics Neural Architecture Search* to capture and align the complex interplay among heterogeneous multi-omics modalities, while dynamically selecting hierarchically optimal neural architectures tailored for each drug pair. We further present the *Mechanism-aware Pathway-based Symbolic Reasoner*, where mechanistic rules are unified within a symbolic space and exploited through pathway-based reasoning to automatically uncover multi-hop, cross-modal entailment rules for interpretable multi-omics DDI prediction.

Moreover, the proposed **MIND-DDI** framework is designed to jointly optimize the biomedical graph structure, hierarchical multi-omics graph neural architecture, and the mechanism-aware symbolic rule space within a unified manner. By resorting to our proposed joint optimization, **MIND-DDI** benefits from the following two advantages: i) mitigating the issue caused by the misalignment between neural representations learned from heterogeneous, noisy biological subgraphs and rule-guided symbolic mechanisms that reflect underlying interaction principles, and ii) preventing spurious subgraph correlations irrelevant to the true pharmacological symbolic rules. We conduct extensive experiments on four dataset splits, demonstrating that our method achieves superior performance over existing approaches. Furthermore, we validate the interpretability of **MIND-DDI** from multiple perspectives. In summary, the main contributions of this paper lie in the following points:

- We are the first to study **Multi-Omics Interpretable Drug-Drug Interaction** prediction, and propose a novel **MIND-DDI** framework with theoretical convergence guarantee which jointly models heterogeneous multi-omics structure, neural graph architecture, and symbolic rules containing explicit mechanism-aware reasoning.
- We propose a joint optimization strategy that unifies biomedical graph structures, neural representation learning, and mechanism-aware symbolic reasoning together, applying it to multi-omics interpretable DDI prediction.
- We conduct extensive experiments on four dataset splits to demonstrate the superiority of our **MIND-DDI** with comprehensive interpretability analyses, validating its ability of predic-

tion accuracy and interpretability.Our code is available at `https://github.com/EkkoXiao/MIND-DDI/`.

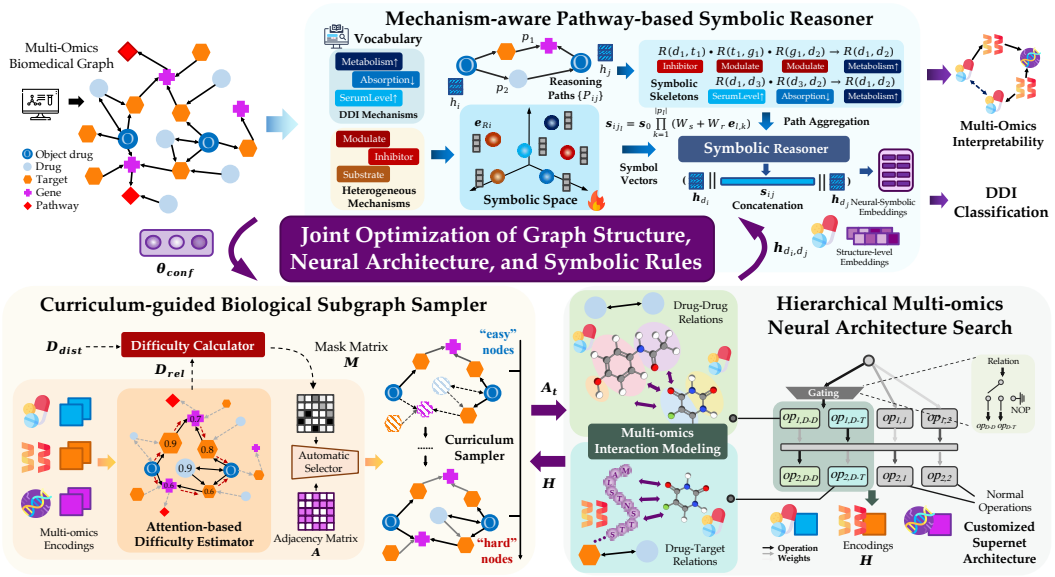

Figure 1: The overall framework of **MIND-DDI**. Our **MIND-DDI** enables multi-omics interpretable DDI prediction through three jointly optimized modules: (1) Curriculum-guided Biological Subgraph Sampler; (2) Hierarchical Multi-omics Neural Architecture Search; and (3) Mechanism-aware Pathway-based Symbolic Reasoner.

2 Preliminaries

In this section, we present the formal problem setup for multi-omics DDI prediction and define the notion of interpretability from a mechanism-aware and symbolic perspective.

Drug-Drug Interaction Prediction We formulate the Drug-Drug Interaction (DDI) prediction task based on a heterogeneous multi-omics biomedical knowledge graph. Let $\mathcal{G}_{KG} = (\mathcal{V}_{KG}, \mathcal{E}_{KG}, \mathcal{R}_{KG})$ denote the knowledge graph, where $\mathcal{V}_{KG}$ is the set of nodes representing heterogeneous biomedical entities; $\mathcal{E}_{KG}$ is the set of edges capturing diverse associations between entities; and $\mathcal{R}_{KG}$ denotes the set of edge types that encode underlying biological mechanisms.

Given a drug pair $(d_i, d_j) \in \mathcal{V}_{KG} \times \mathcal{V}_{KG}$, the goal of DDI prediction is to learn a classification mapping:

$$f : (d_i, d_j, \mathcal{G}_{KG}) \rightarrow \hat{y}_{ij} \in \mathcal{Y}_{DDI}, \tag{1}$$

where $\mathcal{Y}_{DDI}$ is the set of predefined DDI types.

Multi-Omics Interpretability In multi-omics DDI prediction, interpretability refers to the ability to provide mechanistic explanations for how hierarchical and heterogeneous omics features jointly give rise to a DDI. We formalize elementary biological mechanisms as a mechanism vocabulary $\mathbb{V}_{mech}$, where each mechanism $s \in \mathbb{V}$mech represents a distinct biological relation. Let $\mathcal{O} = \{O_X, O_Y, \dots\}$ represent different omics modalities. A multi-omics interpretable model not only predicts the interaction outcome $\hat{y}_{ij}$, but also outputs a set of mechanism-aware multi-omics evidence:

$$\mathcal{S}_{ij} = \left\{ \psi(d_i, v_1^{O_X}) \wedge \psi(v_1^{O_X}, v_2^{O_Y}) \wedge \cdots \implies \psi(d_i, d_j) \right\}, \tag{2}$$

where $v_k \in \mathcal{V}_{KG}$ represents an intermediate biological entity, and $\psi(\cdot,\cdot) \rightarrow s_\ell \in \mathbb{V}_{mech}$. The rule set $\mathcal{S}_{ij}$ explicitly characterizes how cross-modal interactions collectively lead to the predicted DDI and provides a biologically grounded explanation.

3 The MIND-DDI Framework

In this section, we elaborate on the proposed **MIND-DDI** framework, which is composed of three core modules. Sec. 3.1 introduces the *Curriculum-guided Biological Subgraph Sampler*, while Sec. 3.2 and Sec. 3.3 elaborate on *Hierarchical Multi-omics Neural Architecture Search* and *Mechanism-aware Pathway-based Symbolic Reasoner*. The complete algorithmic workflow of the framework is described in Sec. 3.4. An overview of the proposed framework is illustrated in Fig. 1.

3.1 Curriculum-guided Biological Subgraph Sampler

To enable interpretable and efficient reasoning over large-scale knowledge graphs, we propose *Curriculum-guided Biological Subgraph Sampler*, which consists of two components as follows, i) a *Difficulty Calculator* that assigns node-wise difficulty scores through integrating structural proximity and semantic relevance, and ii) a *Curriculum Sampler* that admits nodes of increasing difficulty to construct subgraphs in a curriculum manner.

Difficulty Calculator Specifically, for the target drug pair (d_i, d_j) and a node v in $\mathcal{G}_{KG}$, we first define a distance-based difficulty score that captures its topological proximity to both drugs:

$$D_{\text{dist}}(v) = \sqrt{\frac{1}{1+\text{dist}(v,d_i)} \cdot \frac{1}{1+\text{dist}(v,d_j)}}, \tag{3}$$

where $\text{dist}(\cdot,\cdot)$ denotes the shortest-path distance. This formulation prioritizes nodes that are close to both drugs, which are more likely to participate in meaningful interaction mechanisms.

To further assess semantic relevance, we construct a drug-pair context vector by concatenating the encodings of the two drugs obtained from Sec. 3.2. Let $\mathbf{h}_v$, $\mathbf{h}_{d_i}$ and $\mathbf{h}_{d_j}$ denote the embeddings of node v and the target drug pair. Each node v receives an attention-based relevance score

$$D_{\text{rel}}(v) = \text{softmax}_v\big(\mathbf{W}_h \mathbf{h}_v^{(L)} \cdot \mathbf{W}_b \big[\mathbf{h}_{d_i}^{(L)} \,\|\, \mathbf{h}_{d_j}^{(L)}\big]\big), \tag{4}$$

where $\mathbf{W}_h$ and $\mathbf{W}_b$ are learnable projection matrices. This score quantifies how strongly a node contributes to the joint representation of the target drug pair. Finally, the overall difficulty score is computed as a weighted combination of both factors:

$$D(v) = \lambda D_{\text{dist}}(v) + (1-\lambda)\,\hat{D}_{\text{rel}}(v), \tag{5}$$

where $\hat{D}_{\text{rel}}(v)$ denotes the normalized relevance score, and $\lambda \in (0,1)$ controls the curriculum balance. Nodes with a higher $D(v)$ are considered easier and more explanatory, and are preferentially sampled at early training stages.

Curriculum Sampler Based on the node-wise difficulty scores $\{D(v)\}_{v\in\mathcal{V}_{KG}}$, we design a curriculum sampler that progressively relaxes a difficulty threshold, enabling the model to transition from easy to more challenging structures.

At training epoch t, we introduce a curriculum threshold $\epsilon_t > 0$ to define a binary node mask matrix

$$\mathbf{M}_t = (\mathbb{I}[\mathbf{D} \geq \epsilon_t \mathbf{1}])(\mathbb{I}[\mathbf{D} \geq \epsilon_t \mathbf{1}])^\top, \tag{6}$$

where $\mathbf{D} \in \mathbb{R}^{|\mathcal{V}_{KG}|}$ collects the difficulty scores of all nodes, and $\mathbb{I}(\cdot)$ denotes the element-wise indicator function. Based on this mask, the curriculum-conditioned subgraph structure is obtained by masking the original adjacency matrix $\mathbf{A}$ as $\mathbf{A}_t = \mathbf{A} \odot \mathbf{M}_t$. $\odot$ denotes element-wise multiplication.

To realize curriculum, threshold ϵ_t is scheduled as a monotonically non-increasing function of the training epoch:

$$\epsilon_t = \max\big(\epsilon_{\min},\, f(t)\big), \tag{7}$$

where $f(t)$ is a predefined decreasing function, and $\epsilon_{\min} > 0$ prevents degenerate inclusion of the entire graph. Furthermore, as will be defined in Sec. 3.3, each training sample is associated with a task confidence score $\theta_{\text{conf}} \in [0,1]$. For high-confidence samples, we adopt a more aggressive exploration strategy by lowering the effective threshold:

$$\epsilon_t^{(i,j)} = \begin{cases} \gamma\,\epsilon_t, & \text{if } \theta_{\text{conf}}^{(i,j)} \geq \tau, \\ \epsilon_t, & \text{otherwise}, \end{cases} \tag{8}$$

where $\gamma \in (0,1)$ is a reduction factor and τ is a confidence cutoff. This allows reliable samples to access harder nodes earlier, facilitating deeper multi-omics reasoning, while preserving conservative structures for uncertain cases.

However, computing over the entire $\mathcal{G}_{KG}$ would incur prohibitive computational cost. We therefore propose Proposition 3.1 to show that our strategy naturally restricts reasoning to a bounded neighborhood of drug combinations.

Proposition 3.1 (Bounded Subgraph Convergence)**.** *Let $D(v)$ denote the node-wise difficulty score defined in Eq. (5) and $\{\epsilon_t\}_{t=1}^{T}$ be a monotonically decreasing curriculum schedule with $\lim_{t\to T}\epsilon_t = \epsilon_{\min} > 0$. Assume that the weighting coefficient $\lambda \in (0,1)$ further satisfies $(1-\lambda) < \epsilon_{\min}$, the sampled subgraph induced by $\mathbf{A}_t$ converges to a finite subgraph $\mathcal{G}_{ij}^{\star}$ such that for any node $v \in \mathcal{V}(\mathcal{G}_{ij}^{\star})$, the distance to both central drugs is bounded:*

$$\max\big(\mathrm{dist}(v,d_i),\mathrm{dist}(v,d_j)\big) \le C, \tag{9}$$

where $C < \infty$ is a constant determined by $\epsilon_{\min}$ and the structural component of $D(v)$.

As a consequence, in practical implementations, we extract a local neighborhood graph within a fixed hop range around target drugs to perform difficulty estimation and curriculum-guided sampling instead of the whole $\mathcal{G}_{KG}$ to improve computational efficiency, while preserving the essential context required for accurate and interpretable reasoning.

3.2 Hierarchical Multi-omics Neural Architecture Search

In this Section, we introduce *Hierarchical Multi-omics Neural Architecture Search*, which consists of two components. The *Interaction Modeling* aligns and integrates heterogeneous multi-omics features, while the *Architecture Search* adaptively discovers customized neural propagation strategies specialized for different relational mechanisms.

Interaction Modeling Given the curriculum-sampled subgraph, we explicitly capture cross-omics relational mechanisms, mainly focusing on two biologically critical relation types: drug-drug and drug-target proteins.

For drug-drug relations, each molecule is decomposed into a set of motif-level substructures using BRICS [15], forming a motif graph. A graph convolutional network (GCN) is then applied to extract motif-level representations, yielding a set of drug-specific embeddings $\{\mathbf{Z}_i\}$. These representations are propagated across drug-drug edges via an attention-based mechanism:

$$\mathbf{h}_i^{(d)} = \sum_{j\in\mathcal{N}_d(i)} \alpha_{ij}^{(d)}\,\mathbf{W}_d\mathbf{Z}_j, \tag{10}$$

where $\alpha_{ij}^{(d)}$ denotes attention weights capturing the importance of drug j to drug i.

For drug-target relations, we encode each target protein sequence using ESM2 [16] to obtain a semantic representation $\mathbf{Z}_p = [\mathbf{z}_{p,1},\ldots,\mathbf{z}_{p,L_p}]$ at the sequence level, where L_p denotes the number of fragments. To bridge the molecular structure and protein semantics, we introduce a cross-attention mechanism.

Given the motif-level drug representation $\mathbf{Z}_i$ and protein token embeddings $\{\mathbf{z}_{p,\ell}\}_{\ell=1}^{L_p}$, we compute the relevance between the drug and each protein token, and aggregate protein information as

$$\alpha_{i,\ell} = \frac{\exp\left(\mathbf{Z}_i^{\top}\mathbf{W}_a\mathbf{z}_{p,\ell}/\sqrt{d}\right)}{\sum_{\ell'=1}^{L_p}\exp\left(\mathbf{Z}_i^{\top}\mathbf{W}_a\mathbf{z}_{p,\ell'}/\sqrt{d}\right)}, \quad \mathbf{h}_i^{(t)} = \sum_{\ell=1}^{L_p}\alpha_{i,\ell}\mathbf{z}_{p,\ell}. \tag{11}$$

Together, this module explicitly integrates chemical substructure-level drug interactions with target-aware molecular-proteomic couplings, enabling coherent and interpretable reasoning over heterogeneous omics relations.

Architecture Search Different biomedical contexts often exhibit heterogeneous interaction patterns, making a single, modality-specific modeling strategy or a fixed message-passing rule insufficient to capture the complexity of diverse environments. To address this, we introduce a gated architecture search module in a supernet formulation.

Specifically, given the adjacency matrix $\mathbf{A}_t$, we construct relation-sensitive gated masks that separate drug-drug and drug-target relations, yielding $\mathbf{A}_t^{\text{D-D}}$ and $\mathbf{A}_t^{\text{D-T}}$. In each layer ℓ, the supernet consists of three types of candidate propagation operators: a drug-drug-specific operator $\text{op}_\ell^{\text{D-D}}$, a drug-target-specific operator $\text{op}_\ell^{\text{D-T}}$, and a set of standard relation-agnostic operators $\{\text{op}_\ell^{\text{N}_i}\}$. We adopt a continuous relaxation strategy and aggregate these candidates via learnable architecture weights:

$$\mathbf{H}^{(\ell+1)} = \sum_{k \in \{\text{D–D, D–T, N}_i\}} g_\ell^k \, \text{op}_\ell^k\left(\mathbf{H}^{(\ell)}, \mathbf{A}_t^k\right), \tag{12}$$

where g_ℓ^k are normalized gating coefficients. Through layer-wise iterative aggregation, the model jointly learns node encodings and a customized supernet architecture.

3.3 Mechanism-aware Pathway-based Symbolic Reasoner

In this section, we propose *Mechanism-aware Pathway-based Symbolic Reasoner*. We first perform *Symbolic Formalization* to construct spaces and skeletons from biomedical pathways. We then introduce a *Path Reasoner* that composes these symbolic rules in a closed-loop manner, yielding neural–symbolic representations for DDI prediction.

Symbolic Formalization We first formalize biological mechanistic knowledge into a symbolic space that serves as the foundation of this module. Specifically, we construct a mechanism vocabulary $\mathbb{V}_{\text{mech}}$ from the biomedical knowledge graph, which consists of two subsets: $\mathbb{V}_{\text{mech}}^{\text{DDI}}$ encoding concrete DDI mechanisms (e.g., metabolism↑, absorption↓), and the other representing heterogeneous interaction mechanisms (e.g., inhibition, substrate, induce). Each symbolic mechanism s_ℓ is embedded into a shared high-dimensional symbolic representation space $\mathbf{e}^{s_\ell} \in \mathbb{R}^{d_v}$ via a learnable mapping ϕ.

Given (d_i, d_j), we extract a set of reasoning paths $\mathcal{P}_{ij}$ from the curriculum-sampled subgraph. To abstract away instance-specific noise while preserving mechanistic structure, we map each path to a symbolic skeleton that captures the ordered composition of underlying mechanisms:

$$\mathcal{SK}_{ij} = \left\{ s_1 \circ s_2 \circ \cdots \circ s_{|p|} \implies s_{ij} \;\middle|\; p \in \mathcal{P}_{ij},\ s_\ell \in \mathbb{V}_{\text{mech}} \right\}, \tag{13}$$

where $\circ$ represents an ordered compositional operator. This symbolic skeletonization transforms reasoning paths into a compact set of mechanism-level interaction templates, explicitly encoding how multiple biological mechanisms interact in sequence to entail an interaction.

Path Reasoner Building upon $\mathcal{SK}_{ij}$, we design a mechanism-aware path reasoner that performs an ordered symbolic composition and neural-symbolic alignment. Each symbolic skeleton sk_ℓ is processed in a step-wise manner to obtain an instance-specific symbolic representation. Starting from an initial state $\mathbf{s}_0$, we iteratively compose mechanism-aware symbols along the skeleton:

$$\mathbf{s}_{ij_\ell} = \mathbf{s}_0 \;\odot\; \prod_{k=1}^{|p_\ell|} \left(\mathbf{W}_s + \mathbf{W}_r \mathbf{e}_{\ell,k}^s \right), \tag{14}$$

where $\mathbf{e}_{\ell,k}^s \in \mathbb{R}^{d_v}$ denotes the embedding of the k-th symbolic mechanism in the ℓ-th skeleton.

The symbolic representations of all skeletons associated with (d_i, d_j) are then aggregated via a mean-pooling operator:

$$\mathbf{s}_{ij} = \text{Pooling}\left(\{\mathbf{s}_\ell\}_{\ell=1}^{|\mathcal{SK}_{ij}|} \right). \tag{15}$$

We then concatenate the pooled symbolic embedding with the initial drug encodings $\mathbf{h}_{d_i}$ and $\mathbf{h}_{d_j}$

$$\mathbf{h}_{ij} = \text{Concat}\left(\left[\mathbf{h}_{d_i} \,\|\, \mathbf{h}_{d_j} \,\|\, \mathbf{s}_{ij} \right] \right). \tag{16}$$

The resulting representation $\mathbf{h}_{ij}$ serves as the neural-symbolic embedding for the drug pair.

Finally, $\mathbf{h}_{ij}$ is mapped to mechanism-level logits using a linear classifier. Concretely, we introduce an auxiliary mechanism-consistency objective

$$\mathcal{L} = \text{CE}(\text{softmax}(\mathbf{W}_{\text{m}} \mathbf{h}_{ij} + \mathbf{b}_{\text{m}}),\ \mathbf{y}_{ij}), \tag{17}$$

where CE means cross-entropy loss, and $\mathbf{y}_{ij} \in \mathbb{V}_{\text{mech}}^{\text{DDI}}$ denotes the ground-truth mechanism label of (d_i, d_j) and is shared with the primary classification objective. This enforces that aggregated symbolic skeletons entail mechanistically valid outcomes, thereby closing the causal loop between path-level symbolic reasoning and instance-level DDI prediction.

Moreover, the sharpness of the mechanism distribution reflects the model's confidence for (d_i, d_j). Let $\boldsymbol{\ell}_{ij} = \mathbf{W}_{\mathrm{m}}\mathbf{h}_{ij} + \mathbf{b}_{\mathrm{m}}$. We therefore define the confidence score as

$$\theta_{\mathrm{conf}}^{(i,j)}(d_i, d_j) = \max_{s \in \mathbb{V}_{\mathrm{mech}}^{\mathrm{DDI}}} \mathrm{softmax}(\boldsymbol{\ell}_{ij})_s \,, \tag{18}$$

which is used as a confidence score, as mentioned in Sec. 3.1.

3.4 Putting All Together: The Joint Optimization

Algorithm 1 in Appendix B summarizes the overall procedure, in which the parameters of the three modules, denoted by $\boldsymbol{\Theta}_{\mathrm{cur}}, \boldsymbol{\Theta}_{\mathrm{nas}}, \boldsymbol{\Theta}_{\mathrm{sym}}$, are jointly optimized in an end-to-end manner.

In addition to $\mathcal{L}$, we introduce a lightweight structure-aware loss $\mathcal{L}_s$. Specifically, we derive drug-level structural embeddings $\mathbf{z}_{ij}^{\mathrm{struct}}$ from fine-grained molecular graphs and map them to interaction logits via a linear classifier. The resulting logits are supervised using the same labels as Eq. 17.

To theoretically justify the stability of the proposed end-to-end training scheme, we establish Proposition 3.2, which shows that the unified objective is non-increasing during joint optimization and that the gradient norm approaches zero along the training trajectory. As a result, the proposed training procedure is stable in the sense of converging toward a stationary solution of the total objective.

Proposition 3.2 (Stationarity of Joint Optimization)**.** *Let $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_{\mathrm{cur}}, \boldsymbol{\Theta}_{\mathrm{arch}}, \boldsymbol{\Theta}_{\mathrm{sym}})$ denote the trainable parameters of all modules. Consider the end-to-end objective $\min_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\Theta})$. Assume that $\mathcal{L}$ is lower bounded and has L-Lipschitz continuous gradients. If the joint parameters are updated by gradient descent with a sufficiently small learning rate $0 < \eta_t \leq 1/L$, then the objective values $\{\mathcal{L}(\boldsymbol{\Theta}^{(t)})\}$ are non-increasing and convergent. Moreover, $\liminf_{t\to\infty} \left\|\nabla\mathcal{L}(\boldsymbol{\Theta}^{(t)})\right\| = 0$. If, in addition, the gradient norm converges, then $\lim_{t\to\infty} \left\|\nabla\mathcal{L}(\boldsymbol{\Theta}^{(t)})\right\| = 0$. Consequently, any accumulation point of the generated sequence, if it exists, is a stationary point of $\mathcal{L}$.*

4 Experiments

We conduct experiments on Ryu's dataset [17] and DrugBank (version 5.1.12, March 14, 2024) [18] and adopt a 5-fold cross-validation protocol for both datasets. We further design two artificially constructed cold-start splits on Ryu's dataset to evaluate the generalization ability of different methods. Following the same data partition ratios as the original splits, we construct: (i) a Seen-Unseen setting, where each drug pair in the test set contains exactly one drug that has appeared in the training set, and (ii) an Unseen-Unseen setting, where both drugs in each test pair are completely unseen during training. The benchmark biomedical knowledge graph we use is based on an integration of BioKG [19] and Bioteque [20]. Detailed dataset statistics are shown in Appendix C.1.

Under the above four evaluation settings, we compare our proposed method with a wide range of representative DDI prediction models, including (i) molecule-based methods DeepDDI [17], SSI-DDI [21], GGI-DDI [4], CMRL [22] and CGIB [23]; (ii) knowledge graph-based methods KGNN [24], SumGNN [25], LaGAT [26] and KnowDDI [5]; (iii) combined methods MolecBioNet [27], and TIGER [28]. All methods are evaluated using Accuracy, PR-AUC, F1-score and Cohen's κ. The experimental results are reported in Table 1.

4.1 Performance Analysis

As shown in Table 1, our method consistently achieves the best performance across all metrics on all datasets. This demonstrates the effectiveness of our unified framework via joint optimization. Specifically, under both settings, several key observations can be made as follows:

Standard Setting (i) Representative molecule-centric methods exhibit clear performance limitations, as they primarily rely on molecular graph structures and are unable to leverage the rich contextual information provided by biomedical knowledge graphs. (ii) Methods that jointly incorporate molecular structures and knowledge graphs achieve relatively stronger performance. However, they still treat biological information in a largely homogeneous manner and lack explicit mechanisms to distinguish, align, and reason over multi-omics signals. (iii) Our method exhibits more stable

Table 1: Performance comparison of different DDI prediction methods under both settings. The best results are highlighted in **bold**.

Standard Setting								
Method	**Ryu's Dataset**				**DrugBank**			
	ACC ↑	F1 ↑	PR-AUC ↑	Cohen's κ ↑	ACC ↑	F1 ↑	PR-AUC ↑	Cohen's κ ↑
DeepDDI	86.8±0.5	70.3±0.6	81.1±1.0	84.2±0.6	86.2±0.3	67.6±2.6	81.0±1.6	84.8±0.4
SSI-DDI	84.8±0.6	61.8±1.7	70.4±1.4	81.8±0.7	74.3±0.9	32.7±0.5	43.8±0.9	71.7±1.0
GGI-DDI	88.9±0.6	81.1±1.9	88.0±1.2	86.9±0.7	85.9±0.8	76.2±1.5	83.6±1.6	84.6±0.9
CMRL	93.4±0.2	85.9±0.3	44.6±1.9	76.8±1.4	89.9±0.9	78.9±0.9	40.8±0.1	68.3±1.9
CGIB	88.7±0.8	88.6±0.9	69.7±1.6	81.2±1.5	88.6±0.3	76.5±0.4	56.6±5.6	72.4±4.6
KGNN	91.6±0.3	81.4±2.1	87.2±1.4	90.0±0.4	90.7±0.2	76.4±1.1	83.2±0.7	89.8±0.2
SumGNN	90.4±0.2	84.2±0.5	91.2±1.3	88.6±0.2	86.4±0.2	73.2±1.9	85.5±1.1	85.1±0.2
LaGAT	93.5±0.1	84.3±1.4	89.9±1.2	92.2±0.1	90.0±0.1	74.1±1.0	82.9±0.8	89.1±0.1
KnowDDI	92.3±0.2	88.0±1.0	93.4±1.1	90.9±0.2	89.4±0.1	77.2±1.5	86.8±0.8	88.5±0.1
MolecBioNet	95.2±0.2	91.2±0.8	94.2±0.6	94.3±0.3	93.5±0.2	84.6±1.0	91.0±0.8	92.9±0.2
TIGER	93.8±0.3	88.0±0.7	89.4±1.1	92.6±0.4	91.7±0.1	79.1±0.8	76.3±1.1	91.0±0.1
MIND-DDI	**95.7**±0.1	**93.1**±0.4	**96.3**±0.7	**94.7**±0.2	**93.8**±0.1	**85.5**±0.2	**91.2**±0.1	**93.1**±0.1
Cold-Start Setting								
Method	**Seen-Unseen**				**Unseen-Unseen**			
	ACC ↑	F1 ↑	PR-AUC ↑	Cohen's κ ↑	ACC ↑	F1 ↑	PR-AUC ↑	Cohen's κ ↑
DeepDDI	63.0±2.8	45.9±1.8	43.0±2.1	55.2±3.1	37.2±2.8	19.7±1.8	14.4±1.7	23.8±2.1
SSI-DDI	56.9±2.0	36.0±2.1	35.7±3.0	47.7±2.0	33.7±4.8	14.4±1.7	12.5±1.5	18.1±3.2
GGI-DDI	63.6±2.3	48.6±2.4	48.5±3.0	55.9±2.2	36.4±3.7	17.0±0.8	15.3±1.2	23.9±2.4
CMRL	44.5±2.5	37.1±0.8	48.9±1.1	35.8±1.9	39.5±2.9	1.7±1.4	15.6±0.7	18.3±4.1
CGIB	44.3±2.5	34.4±2.7	43.5±3.1	31.6±3.4	38.6±0.3	2.5±0.5	2.5±1.5	5.6±2.8
KGNN	41.8±2.9	4.4±0.8	6.9±1.4	26.8±2.3	20.7±3.4	1.8±0.1	17.2±2.6	4.7±2.0
SumGNN	58.0±2.7	40.5±2.5	42.1±3.3	48.6±2.7	31.1±5.8	7.8±1.7	9.3±2.4	10.2±4.0
LaGAT	55.7±1.5	24.9±1.1	27.6±1.2	45.5±1.2	22.7±4.4	1.2±0.1	17.1±2.5	0.2±2.1
KnowDDI	51.6±2.3	28.2±1.4	26.9±1.4	40.9±1.7	26.4±6.4	1.5±0.1	1.7±0.4	4.6±1.4
MolecBioNet	65.2±2.0	53.5±1.7	54.9±3.9	57.6±2.1	38.8±4.9	20.8±1.9	17.9±2.9	24.0±4.1
TIGER	54.3±1.9	32.2±2.8	26.6±2.6	44.2±1.8	26.4±6.4	1.5±0.1	1.7±0.4	7.1±1.9
MIND-DDI	**67.7**±3.1	**61.5**±2.7	**65.4**±2.1	**60.8**±3.3	**40.1**±2.6	**20.8**±0.8	**19.0**±0.2	**25.1**±1.0

performance under cross-validation, while achieving an average improvement of approximately 0.5%, suggesting that the proposed framework remains competitive even without accounting for its interpretability benefits.

Cold-Start Setting (i) Compared to the standard setting, **MIND-DDI** exhibits a more pronounced advantage in these scenarios, which indicates that the model is less dependent on memorizing seen drugs and better at inferring unseen interactions. (ii) Specifically, the Seen-Unseen scenario yields substantially larger improvements, with gains of nearly 10% in F1-score and PR-AUC. This suggests that partial exposure to drug information during training provides a strong inductive signal, enabling the model to transfer learned mechanism-aware and multi-omics representations to novel drugs. (iii) Models relying solely on biomedical graphs experience dramatic performance drop. This indicates that knowledge graphs may introduce noise or redundant patterns when modeling unseen drugs. In contrast, our approach mitigates this issue through the curriculum strategy, which progressively incorporates reliable substructures while suppressing noisy signals.

4.2 Interpretability

We analyze the interpretability of our framework using the interaction between *Malathion* and *Pilocarpine* as a case study. Our model derives a set of multi-omics evidence with clear biological semantics, as in Table 2. Specifically, E1 infers the adverse interaction between *Malathion* and *Pilocarpine* via their shared induction of *Gliosis*. E2 provides a more detailed mechanistic account by tracing a multi-step process involving metabolic modulation, serum level alteration, and cross-drug interactions. Collectively, these explanations demonstrate that our framework produces structured, meaningful explanations that are both human-interpretable and biologically actionable.

Beyond qualitative case studies, we quantitatively evaluate the faithfulness and compactness of the learned explanations on the Ryu dataset using Fidelity and Sparsity metrics [29, 30]. Specifically, we progressively remove the most important explanatory nodes ranked by our model under different ratios r_{remove}, and measure the resulting prediction changes. As shown in Table 3, increasing r_{remove} leads to a consistent increase in Fidelity$+$, while Fidelity$-$ remains low. The results demonstrate that

Table 2: Multi-omics evidence explaining the predicted DDI between *Malathion* and *Pilocarpine*. Different colors represent different omics modalities.

Evidence	Multi-omics Evidence Explanation
E1	$R($ Malathion $,$ Gliosis $) \wedge R($ Pilocarpine $,$ Gliosis $) \Longrightarrow R($ Malathion $,$ Pilocarpine $)$ Induce Induce AdverseEffect↑
E2	$R($ Malathion $,$ Caffeine $) \wedge R($ Caffeine $,$ Nicodine $) \wedge R($ Nicodine $,$ Nadolol $)$ Metabolism↓ SerumLevel↑ $\wedge R($ Nadolol $,$ Pilocarpine $) \Longrightarrow R($ Malathion $,$ Pilocarpine $)$ AdverseEffect↑ AdverseEffect↑
E3	$R($ Malathion $,$ Conjunctivitis $) \wedge R($ Nadolol $,$ Conjunctivitis $) \wedge R($ Nadolol $,$ Pilocarpine $)$ Induce Induce AdverseEffect↑ $\Longrightarrow R($ Malathion $,$ Pilocarpine $)$ AdverseEffect↑

the identified symbolic explanations are both faithful and compact. More interpretability results can be found in Appendix D.2.

Table 3: Interpretability evaluation on the Ryu dataset.

r_{remove}	0.14	0.28	0.42	0.56	0.70	0.85
Sparsity	0.09	0.17	0.26	0.34	0.43	0.51
Fidelity+	$1.85_{\pm 0.68}$	$3.92_{\pm 1.05}$	$5.19_{\pm 1.80}$	$6.58_{\pm 1.42}$	$9.47_{\pm 1.80}$	$13.89_{\pm 1.44}$
Fidelity−	$0.94_{\pm 1.02}$	$0.72_{\pm 0.89}$	$0.38_{\pm 0.73}$	$0.43_{\pm 0.07}$	$0.22_{\pm 0.18}$	$0.19_{\pm 0.70}$

4.3 Ablation Study

To evaluate the contribution of each component in our framework, we conduct an ablation study by constructing several model variants, each removing a specific module:

- **w/o Curri**: Removes the Curriculum-guided Biological Subgraph Sampler and directly trains the model on the full biological subgraph.
- **w/o** D_{rel}: Maintains the curriculum strategy but measures node difficulty solely based on D_{dist}, excluding the relational difficulty term D_{rel}.
- **w/o NAS**: Disables the Hierarchical Multi-omics Neural Architecture Search module and instead adopts a fixed GNN architecture.
- **w/o Multi-Omics**: Preserves architecture search while removing multi-omics modeling.
- **w/o Sym**: Eliminates the Mechanism-aware Pathway-based Symbolic Reasoner, resulting in a purely neural model lacking interpretability.
- **w/o** θ_{conf}: Blocks feedback from the symbolic component to the neural modules, decoupling symbolic reasoning from neural representation learning.

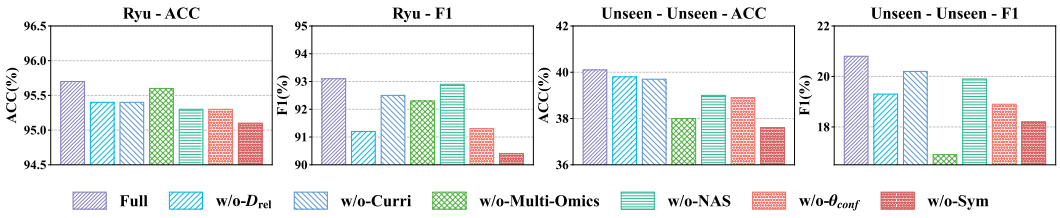

Figure 2: Ablation Study Results.

The results are in Fig. 2. As shown, removing any individual component consistently degrades performance across all metrics. In particular, disabling θ_{conf}, or aggressively removing the symbolic reasoning module altogether, leads to a more notable performance drop, which highlights the importance of our designed mechanism-aware symbolic integration. More pronounced degradation is observed in the unseen–unseen setting, where the **w/o Multi-Omics** variant suffers a substantial performance loss. This indicates that heterogeneous omics information is crucial for generalizing to

novel drugs and capturing novel interactions. Overall, these results demonstrate that our proposed modules play complementary roles. Through joint optimization, the framework effectively balances data sampling, representation learning, and multi-omics reasoning.

5 Related Work

5.1 Drug-Drug Interaction Prediction

Traditional DDI prediction through clinical trials and in vitro assays is reliable but costly and difficult to scale to large interaction spaces [31]. Computational methods have thus been widely explored as efficient, low-cost alternatives [32, 33]. Early approaches relied on drug fingerprints or manually designed chemical features [34, 35], but their limited representational capacity hindered the modeling of complex interaction mechanisms. Recent advances in graph-based learning, especially graph neural networks (GNNs), have substantially improved DDI prediction by exploiting structured relational information [9, 10, 36–38]. Existing graph formulations broadly fall into two paradigms. *Molecular-based* DDI networks, such as SSI-DDI [21], DSN-DDI [39], GGI-DDI [4], CMRL [22], and CGIB [23], model interactions using molecular structures and chemical features. In contrast, *knowledge graph-based* DDI networks, including KGNN [24], SumGNN [25], LaGAT [26], and KnowDDI [5], leverage biomedical knowledge graphs to capture dependencies among drugs and other biological entities. More recently, hybrid approaches such as TIGER [28] and MolecBioNet [27] integrate molecular and biomedical relational information to improve prediction. However, these methods largely focus on single-level interaction modeling without explicitly reasoning over multi-omics entities, limiting their capacity for multi-omics interpretability.

5.2 Neural Symbolic Learning On Graphs

Neural symbolic learning on graphs aims to integrate neural representation learning with symbolic reasoning over graph-structured data. This paradigm has been widely studied in knowledge graph reasoning and question answering, where symbolic rules provide human-interpretable reasoning paths, while neural networks enable robust learning from noisy and incomplete data. Existing approaches can be broadly divided into two categories. In *symbolic-driven neural learning*, symbolic rules are treated as prior knowledge to regularize neural representation learning, typically by constraining the loss function or projecting embeddings into rule-consistent subspaces, as exemplified by FSL [40], KALE [41], RUGE [42], IterE [43], UKGE [44]. *Neural-driven symbolic learning* primarily relies on neural features to parameterize or softly match symbolic rules, enabling differentiable rule induction and weighting, as in RLogic [45], Neural LP [46], NLIL [47], Neural-Num-LP [48], CPL [49], R5 [50]. Despite their success, most existing methods are evaluated on relatively simple tasks, such as logical query answering over knowledge graphs, and typically do not consider the adaptive selection of neural architectures or symbolic rule spaces required to handle complex, heterogeneous data distributions. This limitation is particularly pronounced in biomedical multi-omics settings, where diverse data modalities and hierarchical biological mechanisms demand flexible and interpretable reasoning frameworks, and where corresponding benchmark studies remain largely unexplored.

6 Conclusion

In this work, we introduce **MIND-DDI**, a novel framework for *multi-omics interpretable DDI prediction* that jointly models heterogeneous biological data structure, neural architecture, and explicit mechanism-aware symbolic rules in an interpretable learning paradigm. By integrating curriculum-guided subgraph sampling, hierarchical multi-omics architecture search, and pathway-based symbolic reasoning, our method not only achieves state-of-the-art predictive performance but also provides biologically grounded explanations.

Acknowledgments and Disclosure of Funding

This work was supported by Beijing National Research Center for Information Science and Technology under Grant No.BNR2026TD03005.

References

[1] Wei Jiang, Weicai Ye, Xiaoming Tan, and Yun-Juan Bao. Network-based multi-omics integrative analysis methods in drug discovery: a systematic review. *BioData Mining*, 18(1):27, 2025.

[2] Jin Niu, Robert M Straubinger, and Donald E Mager. Pharmacodynamic drug–drug interactions. *Clinical Pharmacology & Therapeutics*, 105(6):1395–1406, 2019.

[3] A David Rodrigues. *Drug-drug interactions*. CRC Press, 2019.

[4] Hui Yu, Jing Wang, Shi-Yu Zhao, Omayo Silver, Zun Liu, JingTao Yao, and Jian-Yu Shi. GGI-DDI: identification for key molecular substructures by granule learning to interpret predicted drug–drug interactions. *Expert Systems with Applications*, 240:122500, April 2024. doi: https://doi.org/10.1016/j.eswa.2023.122500. URL https://www.sciencedirect.com/science/article/pii/S0957417423030026.

[5] Yaqing Wang, Zaifei Yang, and Quanming Yao. Accurate and interpretable drug-drug interaction prediction enabled by knowledge subgraph learning. *Communications Medicine*, 4: 59, March 2024. doi: 10.1038/s43856-024-00486-y. URL https://doi.org/10.1038/s43856-024-00486-y.

[6] Arnold K Nyamabo, Hui Yu, Zun Liu, and Jian-Yu Shi. Drug–drug interaction prediction with learnable size-adaptive molecular substructures. *Briefings in Bioinformatics*, 23(1):bbab441, January 2022. doi: 10.1093/bib/bbab441. URL https://doi.org/10.1093/bib/bbab441.

[7] Farhan Tanvir, Muhammad Ifte Khairul Islam, and Esra Akbas. Predicting drug-drug interactions using meta-path based similarities. In *2021 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–8, 2021. doi: 10.1109/CIBCB49929.2021.9562802.

[8] Md. Rezaul Karim, Michael Cochez, Joao Bosco Jares, Mamtaz Uddin, Oya Beyan, and Stefan Decker. Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-lstm network. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '19, page 113–123, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450366663. doi: 10.1145/3307339.3342161. URL https://doi.org/10.1145/3307339.3342161.

[9] Changxiang He, Yuru Liu, Hao Li, Hui Zhang, Yaping Mao, Xiaofei Qin, Lele Liu, and Xuedian Zhang. Multi-type feature fusion based on graph neural network for drug-drug interaction prediction. *BMC Bioinformatics*, 23:224, June 2022. doi: 10.1186/s12859-022-04763-2. URL https://doi.org/10.1186/s12859-022-04763-2.

[10] Zhichun Guo, Kehan Guo, Bozhao Nan, Yijun Tian, Roshni G. Iyer, Yihong Ma, Olaf Wiest, Xiangliang Zhang, Wei Wang, Chuxu Zhang, and Nitesh V. Chawla. Graph-based molecular representation learning. In Edith Elkind, editor, *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pages 6638–6646. International Joint Conferences on Artificial Intelligence Organization, 8 2023. doi: 10.24963/ijcai.2023/744. URL https://doi.org/10.24963/ijcai.2023/744. Survey Track.

[11] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, July 2018. doi: 10.1093/bioinformatics/bty294. URL https://doi.org/10.1093/bioinformatics/bty294.

[12] Xiaorui Su, Lun Hu, Zhuhong You, Pengwei Hu, and Bowei Zhao. Attention-based knowledge graph representation learning for predicting drug-drug interactions. *Briefings in Bioinformatics*, 23(3):bbac140, May 2022. doi: 10.1093/bib/bbac140. URL https://doi.org/10.1093/bib/bbac140.

[13] Jing Zhang, Bo Chen, Lingxi Zhang, Xirui Ke, and Haipeng Ding. Neural, symbolic and neural-symbolic reasoning on knowledge graphs. *AI Open*, 2:14–35, 2021.

[14] Dongran Yu, Bo Yang, Dayou Liu, Hui Wang, and Shirui Pan. A survey on neural-symbolic learning systems. *Neural Networks*, 166:105–126, 2023.

[15] Jörg Degen, Christof Wegscheid-Gerlach, Andrea Zaliani, and Matthias Rarey. On the art of compiling and using 'drug-like' chemical fragment spaces. *ChemMedChem*, 3(10):1503–1507, October 2008. doi: https://doi.org/10.1002/cmdc.200800178. URL https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cmdc.200800178.

[16] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574. URL https://www.science.org/doi/abs/10.1126/science.ade2574.

[17] Jae Yong Ryu, Hyun Uk Kim, and Sang Yup Lee. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the National Academy of Sciences*, 115 (18):E4304–E4311, March 2018. doi: 10.1073/pnas.1803294115. URL https://www.pnas.org/doi/abs/10.1073/pnas.1803294115.

[18] Craig Knox, Mike Wilson, Christen M Klinger, Mark Franklin, Eponine Oler, Alex Wilson, Allison Pon, Jordan Cox, Na Eun (Lucy) Chin, Seth A Strawbridge, Marysol Garcia-Patino, Ray Kruger, Aadhavya Sivakumaran, Selena Sanford, Rahil Doshi, Nitya Khetarpal, Omolola Fatokun, Daphnee Doucet, Ashley Zubkowski, Dorsa Yahya Rayat, Hayley Jackson, Karxena Harford, Afia Anjum, Mahi Zakir, Fei Wang, Siyang Tian, Brian Lee, Jaanus Liigand, Harrison Peters, Ruo Qi (Rachel) Wang, Tue Nguyen, Denise So, Matthew Sharp, Rodolfo da Silva, Cyrella Gabriel, Joshua Scantlebury, Marissa Jasinski, David Ackerman, Timothy Jewison, Tanvir Sajed, Vasuk Gautam, and David S Wishart. Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic Acids Research*, 52(D1):D1265–D1275, January 2024. doi: 10.1093/nar/gkad976. URL https://doi.org/10.1093/nar/gkad976.

[19] Brian Walsh, Sameh K. Mohamed, and Vít Nováček. BioKG: a knowledge graph for relational learning on biological data. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, CIKM '20, page 3173–3180, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450368599. doi: 10.1145/3340531.3412776. URL https://doi.org/10.1145/3340531.3412776.

[20] Adrià Fernández-Torras, Miquel Duran-Frigola, Martino Bertoni, Martina Locatelli, and Patrick Aloy. Integrating and formatting biomedical data as pre-calculated knowledge graph embeddings in the Bioteque. *Nature Communications*, 13:5304, September 2022. doi: 10.1038/s41467-022-33026-0. URL https://doi.org/10.1038/s41467-022-33026-0.

[21] Arnold K Nyamabo, Hui Yu, and Jian-Yu Shi. SSI–DDI: substructure–substructure interactions for drug–drug interaction prediction. *Briefings in Bioinformatics*, 22(6):bbab133, November 2021. doi: 10.1093/bib/bbab133. URL https://doi.org/10.1093/bib/bbab133.

[22] Namkyeong Lee, Kanghoon Yoon, Gyoung S Na, Sein Kim, and Chanyoung Park. Shift-robust molecular relational learning with causal substructure. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1200–1212, 2023.

[23] Namkyeong Lee, Dongmin Hyun, Gyoung S Na, Sungwon Kim, Junseok Lee, and Chanyoung Park. Conditional graph information bottleneck for molecular relational learning. In *International Conference on Machine Learning*, pages 18852–18871. PMLR, 2023.

[24] Xuan Lin, Zhe Quan, Zhi-Jie Wang, Tengfei Ma, and Xiangxiang Zeng. KGNN: knowledge graph neural network for drug-drug interaction prediction. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 2739–2745. International Joint Conferences on Artificial Intelligence Organization, 7 2020. doi: 10.24963/ijcai.2020/380. URL https://doi.org/10.24963/ijcai.2020/380. Main track.

[25] Yue Yu, Kexin Huang, Chao Zhang, Lucas M Glass, Jimeng Sun, and Cao Xiao. SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics*, 37(18):2988–2995, September 2021. doi: 10.1093/bioinformatics/btab207. URL https://doi.org/10.1093/bioinformatics/btab207.

[26] Yue Hong, Pengyu Luo, Shuting Jin, and Xiangrong Liu. LaGAT: link-aware graph attention network for drug–drug interaction prediction. *Bioinformatics*, 38(24):5406–5412, December 2022. doi: 10.1093/bioinformatics/btac682. URL https://doi.org/10.1093/bioinformatics/btac682.

[27] Mengjie Chen, Ming Zhang, and Cunquan Qu. Towards interpretable drug-drug interaction prediction: A graph-based approach with molecular and network-level explanations. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, page 203–214, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714542. doi: 10.1145/3711896.3737163. URL https://doi.org/10.1145/3711896.3737163.

[28] Xiaorui Su, Pengwei Hu, Zhu-Hong You, Philip S. Yu, and Lun Hu. Dual-channel learning framework for drug-drug interaction prediction via relation-aware heterogeneous graph transformer. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1):249–256, March 2024. doi: 10.1609/aaai.v38i1.27777. URL https://ojs.aaai.org/index.php/AAAI/article/view/27777.

[29] Kenza Amara, Rex Ying, Zitao Zhang, Zhihao Han, Yinan Shan, Ulrik Brandes, Sebastian Schemm, and Ce Zhang. Graphframex: Towards systematic evaluation of explainability methods for graph neural networks, 2024. URL https://arxiv.org/abs/2206.09677.

[30] Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. Explainability in graph neural networks: A taxonomic survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(5): 5782–5799, 2022.

[31] Steven Whitebread, Jacques Hamon, Dejan Bojanic, and Laszlo Urban. Keynote review: In vitro safety pharmacology profiling: an essential tool for successful drug development. *Drug Discovery Today*, 10(21):1421–1433, November 2005. doi: https://doi.org/10.1016/S1359-6446(05)03632-9. URL https://www.sciencedirect.com/science/article/pii/S1359644605036329.

[32] Reza Ferdousi, Reza Safdari, and Yadollah Omidi. Computational prediction of drug-drug interactions based on drugs functional similarities. *Journal of Biomedical Informatics*, 70:54–64, June 2017. doi: https://doi.org/10.1016/j.jbi.2017.04.021. URL https://www.sciencedirect.com/science/article/pii/S1532046417300953.

[33] Hui Yu, Kui-Tao Mao, Jian-Yu Shi, Hua Huang, Zhi Chen, Kai Dong, and Siu-Ming Yiu. Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC Systems Biology*, 12:14, April 2018. doi: 10.1186/s12918-018-0532-7. URL https://doi.org/10.1186/s12918-018-0532-7.

[34] Santiago Vilar, Rave Harpaz, Eugenio Uriarte, Lourdes Santana, Raul Rabadan, and Carol Friedman. Drug—drug interaction through molecular structure similarity analysis. *Journal of the American Medical Informatics Association*, 19(6):1066–1074, November 2012. doi: 10.1136/amiajnl-2012-000935. URL https://doi.org/10.1136/amiajnl-2012-000935.

[35] Yifan Deng, Xinran Xu, Yang Qiu, Jingbo Xia, Wen Zhang, and Shichao Liu. A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics*, 36(15): 4316–4322, August 2020. doi: 10.1093/bioinformatics/btaa501. URL https://doi.org/10.1093/bioinformatics/btaa501.

[36] Junfeng Fang, Shuai Zhang, Chang Wu, Zhengyi Yang, Zhiyuan Liu, Sihang Li, Kun Wang, Wenjie Du, and Xiang Wang. Moltc: Towards molecular relational modeling in language models. *arXiv preprint arXiv:2402.03781*, 2024.

[37] Linxin Xiao, Xin Wang, Zeyang Zhang, Yang Yao, and Wenwu Zhu. Dynas-ddi: Dynamic pairwise architecture search for generalizable drug-drug interaction llm. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 2216–2225, 2025.

[38] Xin Wang, Linxin Xiao, Yang Yao, and Wenwu Zhu. Ood-graphllm: Graph large language model for out-of-distribution generalized drug synergy prediction. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 12315–12326, 2026.

[39] Zimeng Li, Shichao Zhu, Bin Shao, Xiangxiang Zeng, Tong Wang, and Tie-Yan Liu. DSN-DDI: an accurate and generalized framework for drug–drug interaction prediction by dual-view representation learning. *Briefings in Bioinformatics*, 24(1):bbac597, January 2023. doi: 10.1093/bib/bbac597. URL `https://doi.org/10.1093/bib/bbac597`.

[40] Thomas Demeester, Tim Rocktäschel, and Sebastian Riedel. Lifted rule injection for relation embeddings. *arXiv preprint arXiv:1606.08359*, 2016.

[41] Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 192–202, 2016.

[42] Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Knowledge graph embedding with iterative guidance from soft rules. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[43] Wen Zhang, Bibek Paudel, Liang Wang, Jiaoyan Chen, Hai Zhu, Wei Zhang, Abraham Bernstein, and Huajun Chen. Iteratively learning embeddings and rules for knowledge graph reasoning. In *The world wide web conference*, pages 2366–2377, 2019.

[44] Xuelu Chen, Muhao Chen, Weijia Shi, Yizhou Sun, and Carlo Zaniolo. Embedding uncertain knowledge graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3363–3370, 2019.

[45] Kewei Cheng, Jiahao Liu, Wei Wang, and Yizhou Sun. Rlogic: Recursive logical rule learning from knowledge graphs. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 179–189, 2022.

[46] Fan Yang, Zhilin Yang, and William W Cohen. Differentiable learning of logical rules for knowledge base reasoning. *Advances in neural information processing systems*, 30, 2017.

[47] Yuan Yang and Le Song. Learn to explain efficiently via neural logic inductive learning. *arXiv preprint arXiv:1910.02481*, 2019.

[48] Po-Wei Wang, Daria Stepanova, Csaba Domokos, and J Zico Kolter. Differentiable learning of numerical rules in knowledge graphs. In *International conference on learning representations*, 2019.

[49] Cong Fu, Tong Chen, Meng Qu, Woojeong Jin, and Xiang Ren. Collaborative policy learning for open knowledge graph reasoning. *arXiv preprint arXiv:1909.00230*, 2019.

[50] Shengyao Lu, Bang Liu, Keith G Mills, Shangling Jui, and Di Niu. R5: Rule discovery with reinforced and recurrent relational reasoning. *arXiv preprint arXiv:2205.06454*, 2022.

[51] Ross Taylor, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. Galactica: A large language model for science. *arXiv preprint arXiv:2211.09085*, 2022.

A Proof

A.1 Proof of Proposition 3.1

Proposition 3.1 Let $D(v)$ denote the node-wise difficulty score defined in Eq. (5) and $\{\epsilon_t\}_{t=1}^{T}$ be a monotonically decreasing curriculum schedule with $\lim_{t\to T}\epsilon_t = \epsilon_{\min} > 0$. Assume that the weighting coefficient $\lambda \in (0,1)$ further satisfies $(1-\lambda) < \epsilon_{\min}$, the sampled subgraph induced by $\mathbf{A}_t$ converges to a finite subgraph $\mathcal{G}_{ij}^{\star}$ such that for any node $v \in \mathcal{V}(\mathcal{G}_{ij}^{\star})$, the distance to both the central drugs is bounded:

$$\max\big(\mathrm{dist}(v,d_i), \mathrm{dist}(v,d_j)\big) \le C, \tag{19}$$

where $C < \infty$ is a constant determined by $\epsilon_{\min}$ and the structural component of $D(v)$.

Proof. We first analyze the asymptotic behavior of the distance-based difficulty term Eq.(3). Since each factor is positive and monotonically decreasing with respect to the graph distance, it follows immediately that

$$\lim_{\min(\mathrm{dist}(v,d_i),\, \mathrm{dist}(v,d_j))\to\infty} D_{\mathrm{dist}}(v) = 0. \tag{20}$$

Hence, for any $\delta > 0$, there exists a finite constant $R(\delta) < \infty$ such that

$$\min(\mathrm{dist}(v,d_i),\, \mathrm{dist}(v,d_j)) \ge R(\delta) \implies D_{\mathrm{dist}}(v) \le \delta. \tag{21}$$

Recall the overall node difficulty defined in Eq.(5), where the semantic relevance term is bounded, i.e.,

$$\hat{D}_{\mathrm{rel}}(v) \le C_{\mathrm{rel}} \le 1. \tag{22}$$

Therefore, for all nodes satisfying the above distance condition, we have

$$D(v) \le \lambda\delta + (1-\lambda)C_{\mathrm{rel}}. \tag{23}$$

For the curriculum threshold, we choose

$$\begin{aligned} \delta^{\star} &= \frac{\epsilon_{\min} - (1-\lambda)C_{\mathrm{rel}}}{\lambda} \\ &\ge \frac{\epsilon_{\min} - (1-\lambda)}{\lambda} \\ &> 0. \end{aligned} \tag{24}$$

According to Eq.21, for any $\delta \in (0, \delta^{\star})$, there exists a corresponding $R(\delta) < \infty$ such that

$$\min(\mathrm{dist}(v,d_i),\, \mathrm{dist}(v,d_j)) \ge R(\delta) \implies D(v) < \epsilon_{\min}. \tag{25}$$

Since a node is retained at epoch t only if

$$D(v) \ge \epsilon_t, \tag{26}$$

all nodes whose minimum distance to (d_i, d_j) exceeds $R(\delta)$ are permanently excluded as $t \to T$.

Consequently, there exists a finite constant $C = R(\delta)$ such that any node v retained in the final curriculum mask satisfies

$$\min\big(\mathrm{dist}(v,d_i),\, \mathrm{dist}(v,d_j)\big) \le C. \tag{27}$$

Finally, the curriculum-conditioned adjacency matrix

$$\mathbf{A}_t = \mathbf{A} \odot \mathbf{M}_t, \qquad \mathbf{M}_t = \mathbb{I}[\mathbf{D} \ge \epsilon_t \mathbf{1}]\, \mathbb{I}[\mathbf{D} \ge \epsilon_t \mathbf{1}]^{\top}, \tag{28}$$

retains only edges whose endpoints satisfy the above distance bound. Therefore, the sampled subgraph converges to a finite induced subgraph contained within a bounded neighborhood of (d_i, d_j), completing the proof. □

A.2 Proof of Proposition 3.2

Proposition 3.2 Let $\mathbf{\Theta} = (\mathbf{\Theta}_{\text{cur}}, \mathbf{\Theta}_{\text{arch}}, \mathbf{\Theta}_{\text{sym}})$ denote the trainable parameters of all modules. Consider the end-to-end objective $\min_{\mathbf{\Theta}} \mathcal{L}(\mathbf{\Theta})$. Assume that $\mathcal{L}$ is lower bounded and has L-Lipschitz continuous gradients. If the joint parameters are updated by gradient descent with a sufficiently small learning rate $0 < \eta_t \leq 1/L$, then the objective values $\{\mathcal{L}(\mathbf{\Theta}^{(t)})\}$ are non-increasing and convergent. Moreover, $\liminf_{t\to\infty} \left\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\right\| = 0$. If, in addition, the gradient norm converges, then $\lim_{t\to\infty} \left\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\right\| = 0$. Consequently, any accumulation point of the generated sequence, if it exists, is a stationary point of $\mathcal{L}$.

Proof. Let

$$\mathbf{\Theta} = (\mathbf{\Theta}_{\text{cur}}, \mathbf{\Theta}_{\text{arch}}, \mathbf{\Theta}_{\text{sym}})$$

denote the joint parameters of all modules. The joint training procedure updates the parameters by backpropagating through the end-to-end objective:

$$\mathbf{\Theta}^{(t+1)} = \mathbf{\Theta}^{(t)} - \eta_t \nabla\mathcal{L}(\mathbf{\Theta}^{(t)}), \tag{29}$$

where $0 < \eta_t \leq 1/L$.

Since $\mathcal{L}$ has L-Lipschitz continuous gradients, the standard descent lemma gives

$$\mathcal{L}(\mathbf{\Theta}^{(t+1)}) \leq \mathcal{L}(\mathbf{\Theta}^{(t)}) + \left\langle \nabla\mathcal{L}(\mathbf{\Theta}^{(t)}), \mathbf{\Theta}^{(t+1)} - \mathbf{\Theta}^{(t)} \right\rangle + \frac{L}{2} \left\|\mathbf{\Theta}^{(t+1)} - \mathbf{\Theta}^{(t)}\right\|^2. \tag{30}$$

Substituting the update rule yields

$$\mathcal{L}(\mathbf{\Theta}^{(t+1)}) \leq \mathcal{L}(\mathbf{\Theta}^{(t)}) - \eta_t \left\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\right\|^2 + \frac{L\eta_t^2}{2} \left\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\right\|^2. \tag{31}$$

Equivalently,

$$\mathcal{L}(\mathbf{\Theta}^{(t+1)}) \leq \mathcal{L}(\mathbf{\Theta}^{(t)}) - \eta_t \left(1 - \frac{L\eta_t}{2}\right) \left\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\right\|^2. \tag{32}$$

Since $0 < \eta_t \leq 1/L$, we have $1 - L\eta_t/2 \geq 1/2$, and therefore

$$\mathcal{L}(\mathbf{\Theta}^{(t+1)}) \leq \mathcal{L}(\mathbf{\Theta}^{(t)}) - \frac{\eta_t}{2} \left\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\right\|^2. \tag{33}$$

Thus, the objective value is non-increasing. Since $\mathcal{L}$ is lower bounded, the sequence $\{\mathcal{L}(\mathbf{\Theta}^{(t)})\}$ converges.

Summing the above inequality from $t = 1$ to T gives

$$\sum_{t=1}^{T} \frac{\eta_t}{2} \left\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\right\|^2 \leq \mathcal{L}(\mathbf{\Theta}^{(1)}) - \mathcal{L}(\mathbf{\Theta}^{(T+1)}). \tag{34}$$

Taking $T \to \infty$ and using the lower boundedness of $\mathcal{L}$, we obtain

$$\sum_{t=1}^{\infty} \eta_t \left\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\right\|^2 < \infty. \tag{35}$$

Therefore, the gradient norm cannot remain bounded away from zero. Otherwise, there would exist a constant $\delta > 0$ such that $\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\| \geq \delta$ for all sufficiently large t, which would make the above summation diverge. Hence,

$$\liminf_{t\to\infty} \left\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\right\| = 0. \tag{36}$$

If the gradient norm itself converges, then the above liminf result further implies

$$\lim_{t\to\infty} \left\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t)})\right\| = 0. \tag{37}$$

Finally, let $\mathbf{\Theta}^\star$ be any accumulation point of the generated sequence. Then there exists a subsequence $\{\mathbf{\Theta}^{(t_k)}\}$ such that $\mathbf{\Theta}^{(t_k)} \to \mathbf{\Theta}^\star$ and $\|\nabla\mathcal{L}(\mathbf{\Theta}^{(t_k)})\| \to 0$. By the continuity of $\nabla\mathcal{L}$, we have

$$\nabla\mathcal{L}(\mathbf{\Theta}^\star) = \mathbf{0}. \tag{38}$$

Therefore, $\mathbf{\Theta}^\star$ is a stationary point of $\mathcal{L}$.

The delayed activation of the symbolic module only affects a finite number of early training epochs. After activation, its parameters are included in $\mathbf{\Theta}$ and optimized jointly with the other modules. Therefore, this finite-stage warm-up does not affect the asymptotic stationarity result. □

B Algorithm

Algorithm 1 Joint Optimization of Data Structure, Neural Architecture, and Symbolic Rules

Input: $\mathcal{G}_{KG}$, (d_i, d_j), hyperparameters $\lambda, \epsilon_{\min}, \gamma, \tau$, etc. symbolic start epoch T_{sym}, total epochs T
Output: Label $\hat{y}_{ij}$, explanation $\mathcal{S}_{ij}$
Extract a neighborhood subgraph $\mathcal{G}_{ij}$ from $\mathcal{G}_{KG}$ around drug combination (d_i, d_j) (Proposition 3.1)
Initialize trainable parameters $\{\mathbf{\Theta}_{\text{cur}}, \mathbf{\Theta}_{\text{nas}}, \mathbf{\Theta}_{\text{sym}}\}$
Initialize confidence score $\theta_{\text{conf}}^{(i,j)} \leftarrow 0$
for $t = 1$ **to** T **do**
  Compute node-wise scores $\{D(v)\}$ via Eq. (5)
  Update curriculum threshold ϵ_t and obtain $\mathbf{M}_t$, $\mathbf{A}_t$ using ϵ_t, $\theta_{\text{conf}}^{(i,j)}$ and $\mathbf{A}$ (Sec. 3.1)
  Obtain entity representations $\mathbf{H}$, $\mathbf{h}_{d_i}$, $\mathbf{h}_{d_j}$ via $\mathbf{\Theta}_{\text{nas}}$ on $\mathbf{A}_t$ (Sec. 3.2)
  if $t \geq T_{\text{sym}}$ **then**
    Extract reasoning skeletons $\mathcal{SK}_{ij}$ via Eq. (13)
    Obtain embedding $\mathbf{h}_{ij}$ using $\mathbf{\Theta}_{\text{sym}}$ (Sec. 3.3)
  else
    Set $\mathbf{h}_{ij} \leftarrow \text{Concat}\left(\mathbf{h}_{d_i} \,\|\, \mathbf{h}_{d_j} \,\|\, \mathbf{0}\right)$
  end if
  Update confidence score $\theta_{\text{conf}}^{(i,j)}$ via Eq. (18)
  Compute $\mathcal{L}$ via Eq. (17) and $\mathcal{L}_s$ using fine-grained molecular representations
  Update all parameters by backpropagating $\mathcal{L} + \gamma_s\mathcal{L}_s$
end for
Formalize symbolic explanations $\mathcal{S}_{ij}$ from the top-k highest-confidence elements in $\mathcal{SK}_{ij}$
Output prediction $\hat{y}_{ij}$ from h_{ij}

C Experiment Details

C.1 Dataset Statistics

Table 4 summarizes the data split statistics for each dataset and evaluation setting. For the standard benchmark splits, we adopt a 5-fold cross-validation protocol, where the dataset is evenly partitioned into five folds with identical split sizes across all folds. For the manually constructed evaluation settings, the data are partitioned based on predefined constraints over drug occurrence patterns. Specifically, drug pairs in the test set are selected to satisfy particular drug-level generalization conditions, which may result in varying absolute numbers of samples across folds.

C.2 Baselines

In this section, we provide brief descriptions of the baselines we compare our **MIND-DDI** with:

- **Molecular graph-based methods**
  - **DeepDDI** [17]: A deep learning–based framework that predicts fine-grained drug interaction types using only drug names and structural information.

Table 4: Dataset statistics and data split sizes for each fold under different evaluation settings.

Setting	Dataset / Setting	Split	Fold				
			1	2	3	4	5
Standard Setting	Ryu's Dataset	Train			114,832		
		Valid			38,278		
		Test			38,278		
	DrugBank (v5.1.12)	Train			370,365		
		Valid			123,456		
		Test			123,456		
Cold-Start Setting	Seen-Unseen	Train	121,824	118,059	117,549	121,828	113,462
		Valid	33,872	35,622	35,881	33,869	37,781
		Test	33,872	35,622	35,881	33,869	37,781
	Unseen-Unseen	Train	70,223	64,637	65,946	67,838	58,205
		Valid	18,948	20,931	20,487	19,891	23,089
		Test	18,948	20,931	20,487	19,891	23,089

  - **SSI-DDI** [21]: A deep learning framework that models drug–drug interactions by explicitly decomposing them into pairwise interactions between molecular substructures.
  - **GGI-DDI** [4]: A granular computing–based end-to-end framework that decomposes drugs into interpretable granules corresponding to key substructures or functional groups.
  - **CMRL** [22]: A causal framework that leverages a structural causal model to disentangle causally relevant substructures from spurious correlations.
  - **CGIB** [23]: A graph-based framework that predicts interactions by identifying core subgraphs from each molecule conditioned on its interaction partner.
- **Knowledge graph-based methods**
  - **KGNN** [24]: An end-to-end knowledge graph neural network for DDI prediction that models drugs together with related entities by aggregating their multi-hop neighborhood information.
  - **SumGNN** [25]: A knowledge graph neural network that addresses noise and scale issues in biomedical KGs by anchoring on relevant subgraphs and summarizing them into compact reasoning representations.
  - **LaGAT** [26]: A graph attention framework that enhances DDI prediction by integrating link-aware attention to capture interaction-specific relational semantics between drugs.
  - **KnowDDI** [5]: A graph neural network–based framework that enhances drug representations by adaptively aggregating neighborhood information from large biomedical knowledge graphs.
- **Combined methods**
  - **MolecBioNet** [27]: A multi-scale DDI prediction framework that models drug pairs as unified graphs by integrating molecular-level and biomedical knowledge representations.
  - **TIGER** [28]: A transformer-basedgraph learning framework that models both drug molecular graphs and biomedical knowledge graphs via a dual-channel architecture.

C.3 Metrics

We evaluate classification performance using four complementary metrics: Accuracy, PR-AUC, F1-score, and Cohen's κ. Accuracy measures the overall proportion of correctly classified samples. PR-AUC summarizes the area under the precision-recall curve. It is particularly informative for imbalanced datasets. F1-score, defined as the harmonic mean of precision and recall, balances false positives and false negatives and reflects the trade-off between sensitivity and precision. Cohen's κ evaluates the agreement between predicted and ground-truth labels while accounting for chance agreement.

C.4 Candidate Operations

The candidate operation set $\mathcal{O}$ contains six operators, covering relation-specific propagation, relation-agnostic graph modeling, and node-wise feature transformation. This search space allows the model to adapt the contribution of different information sources and propagation mechanisms at each layer. The candidate operators are described below.

- **DDI Layer.** This operator propagates information through drug-drug relations to model pairwise interaction patterns. By incorporating information from drugs connected through these relations, it captures interaction-dependent features that complement the representation of an individual drug, which is detailedly introduced in Sec. 3.2.
- **DPI Layer.** This operator aggregates information through drug-protein relations to capture target-level biological dependencies. It connects drug representations with their associated proteins, providing a biological basis for modeling interactions beyond drug-drug connectivity alone. Detailed information are introduced in Sec. 3.2.
- **Graph Transformer.** This operator employs sparse multi-head attention followed by a feed-forward network. Attention assigns input-dependent weights to the permitted graph connections, allowing the model to capture long-range dependencies within the sparse attention pattern and combine information from multiple representation subspaces.
- **Graph Convolution.** This operator performs degree-normalized neighborhood aggregation. It incorporates local graph context into node representations while accounting for differences in node degree, providing a standard mechanism for propagating information over graph neighborhoods.
- **GraphSAGE.** This operator combines separately transformed self and neighboring features. Treating these two sources separately allows the model to preserve node-specific information while incorporating contextual information aggregated from neighboring nodes.
- **MLP.** This operator transforms each node independently without graph propagation. It provides a candidate that relies on node features alone, allowing the search process to reduce the contribution of neighborhood information when such information is less useful at a particular layer.

At each layer ℓ, the candidate set consists of the drug–drug-specific operator op^{ℓ}_{D-D}, the drug–target-specific operator op^{ℓ}_{D-T}, and the relation-agnostic operators $\{\mathrm{op}^{\ell}_{N_i}\}$, comprising Graph Transformer, Graph Convolution, GraphSAGE, and MLP.

Following Eq. 12, candidate outputs are continuously aggregated using normalized coefficients g^{ℓ}_{k}. Each coefficient is derived from a learnable architecture parameter and controls the contribution of candidate operator k at layer ℓ. The continuous formulation makes the relative contributions of the candidates differentiable, enabling the architecture parameters to be optimized during training.

C.5 More Details

We optimize the proposed model using the Adam optimizer with an initial learning rate of 1×10^{-3} and a weight decay rate of 1×10^{-5}. An exponential learning rate scheduler with decay factor 0.9 is applied throughout training. The model is trained for a total of $T = 80$ epochs, with the symbolic module activated after $T_{\mathrm{sym}} = 20$ epochs. For the difficulty-aware weighting in Eq. 5, we set $\lambda = 0.5$. In the confidence-based regularization term in Eq. 8, the hyperparameters are set to $\gamma = 0.9$ and $\tau = 0.8$, respectively. The weight of the symbolic loss $\mathcal{L}_s$ is fixed as $\gamma_s = 1.0$. All experiments are able to be conducted on a single NVIDIA A100-SXM4-40GB GPU. With a batch size of 256, the peak GPU memory consumption is approximately 16 GB, indicating that our method incurs relatively low computational and memory overhead.

D Additional Results

D.1 Hyperparameter Analysis

We further analyze the sensitivity of the structure loss weight γ_s on the Ryu and DrugBank datasets. As shown in Fig. 3, both datasets achieve the best performance around $\gamma_s = 1.0$, while smaller values result in moderate degradation and larger values may cause significant drops.

D.2 Interpretability Results

We further present additional interpretability analyses on several drug combinations to demonstrate how our model derives mechanistically coherent relational rules from heterogeneous biomedical evidence. Results are illustrated in Table 5.

Specifically, for *Palmitic Acid* and *Everolimus*, the model identifies a shared *Albumin*-binding pattern, where both *Palmitic Acid* and *Palbociclib* bind *Albumin*, and the known anticoagulant-enhancing

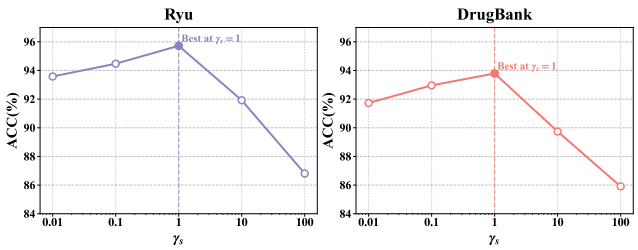

Figure 3: Sensitivity analysis of the structure loss weight γ_s on Ryu and DrugBank. Both datasets show peak performance around $\gamma_s = 1.0$.

Table 5: More multi-omics evidence explaining the predicted DDI. Different colors represent different omics modalities.

Evidence	Multi-omics Evidence Explanation
E1	$R($ Palmitic Acid $,$ Fatty Liver $) \wedge R($ Prednisolone $,$ Fatty Liver $) \wedge R($ Prednisolone $,$ Everolimus $)$ (Induce; Induce; Hypertension↑) $\Longrightarrow R($ Palmitic Acid $,$ Everolimus $)$ (Anticoagulant↑)
E2	$R($ Palmitic Acid $,$ Albumin $) \wedge R($ Palbociclib $,$ Albumin $) \wedge R($ Palbociclib $,$ Everolimus $)$ (Bind; Bind; Anticoagulant↑) $\Longrightarrow R($ Palmitic Acid $,$ Everolimus $)$ (Anticoagulant↑)
E1	$R($ Nicotinamide $,$ Liver Injury $) \wedge R($ Liver Injury $,$ CYP3A4 $) \wedge R($ CYP3A4 $,$ Metoprolol $)$ (Induce; Downregulate; Metabolize) $\wedge R($ Metoprolol $,$ Tamoxifen $) \Longrightarrow R($ Nicotinamide $,$ Tamoxifen $)$ (Metabolism↓; Anticoagulant↑)
E2	$R($ Nicotinamide $,$ CYP2D6 $) \wedge R($ CYP2D6 $,$ Adverse Drug Reaction $) \wedge R($ Tamoxifen $,$ CYP2D6 $)$ (Inhibit; Modulate; Induce) $\Longrightarrow R($ Nicotinamide $,$ Tamoxifen $)$ (Anticoagulant↑)

interaction between *Palbociclib* and *Everolimus* is transferred through this common binding mediator, leading to an inferred increase in anticoagulant-related interaction for *Palmitic Acid–Everolimus*. As to the interaction between *Nicotinamide* and *Tamoxifen*, it is explained via enzyme-mediated modulation, where *Nicotinamide* inhibits *Cytochrome P450 2D6*, an enzyme broadly associated with adverse drug reaction susceptibility, while *Tamoxifen* engages the same metabolic pathway, resulting in perturbed metabolism and an inferred elevation in anticoagulant-related interaction risk. Together, these cases demonstrate that the model captures both protein-binding–driven and metabolism-driven mechanisms to support multi-omics interpretable DDI inference.

To further analyse the case study discussed in Sec. 4.2, we provide an additional visualization of the multi-omics subgraph identified for the drug pair *Malathion* and *Pilocarpine*. This analysis aims to offer a more fine-grained view of how our framework organizes heterogeneous biomedical evidence around a specific drug-drug interaction instance.

Specifically, using the curriculum-based sampler introduced in Sec. 3.1, we estimate the importance of multi-omics nodes associated with this drug pair through the difficulty score defined in Eq. 5. As shown in Fig. 4, the resulting subgraph contains high-importance entities across multiple biological layers, together with their corresponding relation types and mechanism annotations. This supplementary visualization provides additional evidence that the proposed framework can identify structured cross-omics evidence and present it in an interpretable form for individual drug-pair predictions.

D.3 Efficiency

To further assess the practical efficiency of our framework, we report both training and inference runtime comparisons against representative baselines. All results are measured as time per batch with batch size 256. We plot the relationship between accuracy and computation time in Fig. 5. The results show that **MIND-DDI** achieves the best predictive accuracy while maintaining competitive training efficiency, and its moderately higher inference time reflects the additional cost of interpretability-related reasoning procedures.

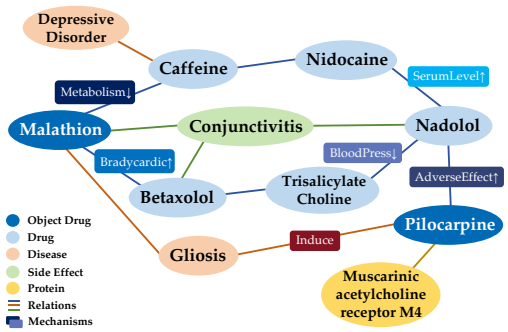

Figure 4: Supplementary visualization of the multi-omics subgraph for the drug pair *Malathion* and *Pilocarpine*. Node colors represent entities from different omics modalities, edge colors denote relation types, and edge annotations indicate the corresponding biological mechanisms.

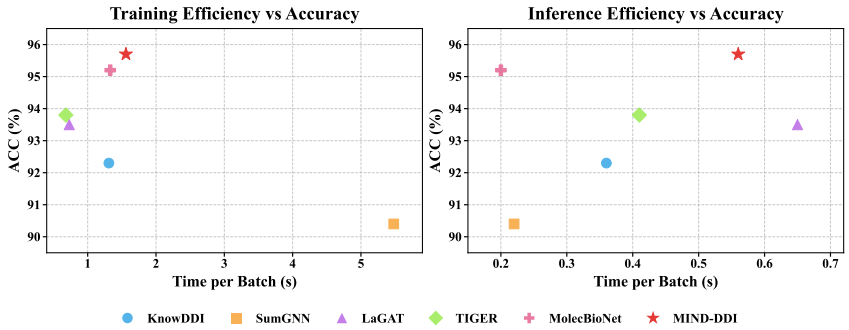

Figure 5: Accuracy-time trade-off analysis. **MIND-DDI** achieves the highest accuracy while maintaining competitive efficiency.

D.4 More Baselines

We further include additional comparisons with three LLM-based methods, Galactica [51], MolTC [**?**] and DyNAS-DDI [37], under the same Ryu and DrugBank settings. These methods rely on a different modeling paradigm, typically involving pre-trained language models, instruction tuning, or external knowledge. For a fair comparison in our setting, we evaluate them without introducing additional instruction tuning or extra pretraining stages.

As shown in Table 6, **MIND-DDI** consistently outperforms these additional baselines across all metrics on both datasets. The improvement is particularly evident in PR-AUC and Cohen's κ, suggesting that our multi-omics neural-symbolic framework provides stronger predictive performance while preserving interpretability.

Table 6: Comparison with additional LLM-based baselines on Ryu and DrugBank under the standard setting. The best results are reported in **bold**.

Dataset	Method	ACC	F1	PR-AUC	Cohen's κ
Ryu	Galactica	90.5±0.6	87.1±0.8	72.3±1.5	82.9±1.1
	MolTC	91.8±0.5	89.6±0.7	83.9±1.2	85.4±0.9
	DyNAS-DDI	94.1±0.3	88.7±0.5	90.2±0.8	89.6±0.6
	MIND-DDI	**95.7**±0.1	**93.1**±0.4	**96.3**±0.7	**94.7**±0.2
DrugBank	Galactica	87.2±0.7	78.4±0.9	60.2±2.8	75.6±2.3
	MolTC	89.3±0.6	80.5±0.8	76.8±1.4	83.1±1.1
	DyNAS-DDI	91.3±0.4	83.9±0.7	86.5±1.0	87.2±0.8
	MIND-DDI	**93.8**±0.1	**85.5**±0.2	**91.2**±0.1	**93.1**±0.1

E Complexity Analysis

According to Proposition 3.1, let $|V|$ and $|E|$ denote the number of nodes and edges in the neighborhood subgraph. For curriculum module, computing D_{dist} requires two shortest-path searches, incurring a cost of $O(2(|V| + |E|))$. D_{rel} is obtained via node-wise projections and attention, which costs $O(|V|d_h^2)$ with hidden dimension d_h. During sampling, thresholding and masking can be implemented by scanning nodes and edges, leading to cost of $O(|V| + |E|)$. We assume that each molecule is decomposed into an average of $\bar{M}$ motif nodes, and each target protein is represented by $\bar{L}_p$ semantic tokens. The interaction modeling module thus incurs an average computational cost of $O\big(\bar{M}^2 d_h^2 + \bar{M}\bar{L}_p d_h d_p\big)$, where d_p denotes the protein feature size. For a standard GNN each layer has a complexity of $O(|E|d_h)$, and for the multi-omics part the complexity becomes $O\big(|E|(\bar{M}^2 d_h^2 + \bar{M}\bar{L}_p d_h d_p)\big)$. For K normal operations and multi-omics operations over L layers, the complexity is $O\big(|E|L(\bar{M}^2 d_h^2 + \bar{M}\bar{L}_p d_h d_p + K d_h)\big)$. Let $\bar{L_S}$ be the average length of each skeleton. For the path reasoner, each symbolic skeleton results in a per-skeleton cost of $O(\bar{L_S} \cdot d_v^2)$. Aggregating all skeletons therefore incurs a total cost of $O(|\mathcal{SK}|\ \bar{L}\ d_v^2)$. Overall, the complexity of the whole process is $O\big(|E|L(\bar{M}^2 d_h^2 + \bar{M}\bar{L}_p d_h d_p + K d_h) + (|V| + |E|) + |V|d_h^2 + |\mathcal{SK}|\ \bar{L}\ d_v^2\big)$.

F Limitations

Despite its effectiveness, our framework still has several limitations. First, although the curriculum-guided subgraph sampler can progressively filter less informative regions and mitigate the influence of noisy biomedical relations, the quality of the learned explanations is still inherently constrained by the coverage and reliability of the underlying knowledge sources. Incomplete or biased multi-omics annotations may therefore affect both prediction and interpretation. Second, the neural-symbolic reasoning module introduces additional inference overhead, as it requires reasoning path retrieval, confidence-based path ranking, and mechanism-aware explanation generation. Third, our evaluation is mainly based on existing DDI benchmarks; further validation on larger-scale datasets, prospective experimental settings, and expert-curated biomedical assessments would strengthen the evidence for biological reliability and generalizability. Finally, the current framework focuses on static DDI prediction and does not explicitly model dose-dependent effects, temporal dynamics, or patient-specific factors, which are important for real-world clinical decision-making.

G Broader Impacts

This work has potential positive impacts on biomedical research by improving the prediction and interpretation of drug-drug interactions from heterogeneous multi-omics evidence. By providing mechanism-aware explanations, the proposed framework may help researchers better understand possible interaction mechanisms and reduce the cost of preliminary experimental screening.

At the same time, we acknowledge several potential risks. Computational predictions may be incorrect or incomplete due to limitations of the training data, knowledge sources, and model assumptions. Therefore, the predicted interactions and generated explanations should not be directly used for clinical decision-making without expert review and experimental or clinical validation. To mitigate these risks, we position our framework as a research tool for hypothesis generation rather than a clinical decision system, and emphasize that all predictions should be validated by domain experts before practical use.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the problem of multi-omics interpretable DDI prediction, the main technical challenges, and the corresponding components of our framework. The claims are supported by the theoretical discussion, predictive performance, ablation studies, and interpretability analyses reported in the paper, and are proven in the evaluated DDI datasets and experimental settings.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the main limitations of our framework, including dependence on the quality of biomedical knowledge sources, additional inference overhead, limited validation beyond existing DDI benchmarks, and the lack of explicit modeling of dose-dependent, temporal, and patient-specific factors in Appendix F.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Proposition 3.1 and proposition 3.2 explicitly state the assumptions required for the subgraph convergence and the stationarity result. The complete proof has been provided in Appendix A.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We disclose the main information needed for reproducibility by providing detailed model architecture, module designs, training objectives, and evaluation settings in the paper. Important hyperparameters and implementation details are further reported in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We provide runnable code and data in the supplementary material, together with scripts for quickly reproducing the main experimental results. The supplementary material also includes instructions for data preparation, training, and evaluation.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide the key experimental settings in the paper, especially in Appendix C, including dataset splits, training protocols, hyperparameter configurations, optimizer choices, and evaluation metrics. These details are intended to make the experimental results understandable and reproducible.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

Justification: All main experimental results are reported as mean ± standard deviation, providing information about the variability and statistical stability of the reported performance.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide sufficient information on the computational resources, including the hardware configuration and memory requirements, in Appendix C.5. We also include detailed training and inference runtime analyses in Appendix D.3.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the NeurIPS Code of Ethics and believe the work conforms to it.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.

- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss potential benefits as well as risks of automated DDI prediction in Appendix G. We emphasize that the method is intended for research use and requires expert and experimental validation before practical deployment.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release data or models with a high risk for misuse. The proposed framework is intended for biomedical research and hypothesis generation rather than direct clinical deployment.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We properly cite the original sources of all datasets, baselines, models, and software packages used in this work. We also describe the data sources, versions, and usage conditions where applicable, and ensure that all existing assets are used consistently with their licenses and terms of use.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: We release the implementation code of our proposed model as a new asset and provide accompanying documentation, including environment setup, data preparation, and training and evaluation commands. The released code is intended to facilitate reproducibility and responsible academic reuse.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing experiments, user studies, or research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing experiments, user studies, or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.