DyNAS-DDI: Dynamic Pairwise Architecture Search for Generalizable Drug-Drug Interaction LLM

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

Zeyang Zhang
DCST, Tsinghua University
Beijing, China
zhangzey16@tsinghua.org.cn

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

Wenwu Zhu[†]
DCST, BNRist, Tsinghua University
Beijing, China
wwzhu@tsinghua.edu.cn

Abstract

Drug-drug interaction (DDI) prediction is a pivotal task in biomedical research. Emerging multimodal approaches that integrate graph neural networks (GNNs) and large language models (LLMs) have gained traction, as GNNs capture molecular structures while LLMs provide a rich biomedical context. However, real-world DDI data often exhibit distribution shifts across structural and textual dimensions, stemming from variations in molecular scaffolds, drug sizes, and assay conditions. Existing methods assume an independent and identically distributed (I.I.D.) setting, failing to handle such shifts primarily due to there key limitations: (i) the entanglement of core interaction motifs with incidental structural features; (ii) inflexible message-passing GNN architectures ill-suited for diverse drug pairs; and (iii) underutilized biomedical knowledge in LLMs for capturing pairwise interaction semantics. These limitations highlight the need for a disentangled, dynamic, and pairwise-aware modeling strategy to achieve out-of-distribution generalized DDI prediction. To solve this problem, we propose **DyN**amic Pairwise **A**rchitecture **S**earch for Generalizable **D**rug-**D**rug **I**nteraction LLM (**DyNAS-DDI**), a novel framework that dynamically adapts network architectures for each molecular pair and integrates biomedical knowledge from LLMs to improve generalization under distribution shifts. Specifically, we propose three modules: (i) Motif-driven disentangled molecule encoding, which disentangles molecular representations into distinct motif-based features while preserving key structural signals through a self-supervised graph encoder; (ii) Attention-based pairwise neural architecture search, where multi-head attention enriches molecular features to guide a dynamic search mechanism that adaptively optimizes message passing for diverse interaction types; and (iii) retrieval-augmented molecular instruction tuning, where external biomedical knowledge is incorporated to improve interpretability and enable reasoning for unseen drug interactions. Extensive experiments on four datasets for DDI with out-of-distribution (OOD) splits demonstrate our method's superior generalization abilities under distribution shifts. Our code can be available at **https://github.com/EkkoXiao/DyNAS-DDI**.

CCS Concepts

• **Computer systems organization → Neural networks**; • **Applied computing → Bioinformatics**.

Keywords

Multimodal Fusion, Drug-Drug Interaction, Neural Architecture Search, Instruction Tuning

ACM Reference Format:
Linxin Xiao, Xin Wang, Zeyang Zhang, Yang Yao, and Wenwu Zhu. 2025. **DyNAS-DDI**: Dynamic Pairwise Architecture Search for Generalizable Drug-Drug Interaction LLM. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3746027.3755791

1 Introduction

Drug-drug interaction (DDI) is a fundamental problem in biomedical research. Predicting molecular interactions is crucial for drug screening, combination drug design, and disease treatment[21, 23, 54] due to their impact on patient safety and treatment efficacy[36, 42, 48]. As a quintessential multimodal problem, DDI prediction requires an understanding of both biomedical text and molecular graph structures. Recently, multimodal approaches that integrate graph neural networks (GNNs) and large language models (LLMs) have gained increasing attention. GNNs can capture molecular topology, while LLMs extract rich semantic features from biomedical literature[41, 47]. By combining structural and textual information, this fusion enhances the understanding of molecular interactions, leading to more accurate and comprehensive predictions.

In real-world DDI prediction tasks, distributional shifts between training and deployment data often arise due to the introduction of novel drugs and evolving experimental conditions. As a result, DDI datasets are frequently subject to out-of-distribution (OOD) scenarios, where differences in assay conditions, chemical scaffolds, and molecular sizes lead to substantial inconsistencies across datasets [18]. For instance, newly discovered drugs frequently exhibit new structural properties or are tested under conditions that

[*] DCST is the abbreviation for Department of Computer Science and Technology.
[†] Corresponding author. BNRist is the abbreviation for Beijing National Research Center for Information Science and Technology.

MM '25, Dublin, Ireland

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3755791

differ greatly from historical data[50, 59]. These distribution shifts pose fundamental challenges in aligning new data with previously observed distributions, making it difficult to maintain consistency across different settings.

However, most existing DDI learning models assume an independent and identically distributed (I.I.D.) setting. This oversimplification hinders their ability to distinguish core interaction motifs from incidental molecular features, adapt GNN architectures to diverse drug pairs, and incorporate broader biomedical context beyond the training interaction data—ultimately leading to poor generalization in OOD scenarios.

To bridge this gap, we aim to improve *out-of-distribution drug-drug interaction (OOD-DDI) prediction* in this work from three key perspectives. First, we seek to identify core interaction patterns that are invariant across distributions. Second, to enhance the adaptability of GNNs to diverse drug pairs, we aim to design architectures that can dynamically adjust to molecular variation. Third, to enrich the model's biomedical understanding beyond the training data, we explore the integration of external domain knowledge through LLMs. However, achieving these goals is highly non-trivial and presents the following key challenges:

- How to disentangle core interaction motifs from overall molecular representations, as motifs often co-occur with certain scaffolds, leading to biased learning where models may conflate general structural properties with motif-driven interactions?
- How to implement a flexible GNN architecture for DDI prediction and tailor a mechanism that can dynamically modify propagation rules without overfitting to specific interaction types, thus achieving careful balancing between adaptability and generalization across unseen drug pairs?
- How to effectively integrate the extensive biomedical knowledge in LLMs into DDI prediction while ensuring that the model remains robust to OOD drug interactions?

To address these challenges, we propose **DyN**amic Pairwise **A**rchitecture **S**earch for Generalizable **D**rug-**D**rug **I**nteraction LLM (**DyNAS-DDI**), a unified framework that dynamically adapts network architectures for each molecular pair and integrates biomedical knowledge from LLMs to improve generalization under distribution shifts. By automatically adjusting the model's information propagation rules based on disentangled interaction contexts, **DyNAS-DDI** enhances its ability to generalize effectively to unseen data distributions.

Specifically, we introduce *Motif-driven disentangled molecule encoding*, which disentangles a molecule's representation into distinct motif-based features rather than learning an entangled global structure. A self-supervised graph encoder trains the model to capture these motifs while preserving key structural influences. Using the obtained encoded representations, we propose *Attention-based pairwise neural architecture search*, where molecular features are enriched using multi-head attention[53]. These features then guide a dynamic search mechanism that optimizes message-passing strategies for different molecule pairs, allowing flexible adaptation to diverse interaction types. Furthermore, to improve molecular reasoning, we introduce *Retrieval-augmented molecular instruction tuning*. By retrieving and integrating external drug-related biomedical knowledge, the model enhances its interpretability and improves its reasoning for novel drug interactions. We evaluate our method on four DDI datasets under two OOD partitioning schemes: scaffold-based and size-based splits. Extensive results show that our approach generalizes well under these distribution shifts. Additionally, in-depth ablation studies demonstrate the contribution of each module to the overall performance. The main contributions of this paper are summarized as follows:

- We investigate the out-of-distribution phenomenon in drug-drug interaction prediction, an underexplored problem in literature. To the best of our knowledge, we are the first to address OOD DDI prediction within GNN-LLM multimodal integration.
- We propose a novel framework comprising three key modules: (i) Motif-driven disentangled molecule encoding, which enhances generalization by capturing motif-level features; (ii) Attention-based pairwise neural architecture search, which dynamically adapts message-passing strategies for diverse molecular interactions; and (iii) Retrieval-augmented generalizable molecular instruction tuning, which leverages external biomedical knowledge to improve reasoning and interpretability.
- We conduct experiments on four DDI datasets with two different distribution shift-based splits. Extensive results demonstrate that our method performs effectively in handling OOD drug interaction pairs.

2 Related Work

2.1 Drug-Drug Interaction Prediction

Traditional DDI prediction methods began with shallow models like logistic regression[15], which rely on handcrafted features. As deep learning emerged, models such as DeepDDI[49] and MatchMaker[20] introduced neural architectures to capture biological and sequential drug features. More advanced approaches like MHCADDI[5] and MDF-SA-DDI[31] further improved prediction by incorporating substructure-level representations, attention mechanisms, and multi-source feature fusion, highlighting a shift toward richer and more adaptive modeling frameworks. Recently, GNNs have gained prominence in DDI prediction by leveraging molecular structures directly. Methods like SSI-DDI[37] and DSN-DDI[29] utilize graph attention networks to capture substructure-level interactions within and between drug pairs. Others, such as CGIB[22] and CIGIN[40], incorporate dynamic substructure adaptation and message-passing frameworks to model chemical behavior and reactivity. Techniques like GCNNs with attention-based pooling[67] further enhance interpretability and predictive accuracy.

With the rapid advancement of LLMs, there has been a growing emphasis on integrating textual modalities with structured drug information. Early efforts primarily focused on single-molecule tasks, such as molecule–text retrieval [10] and molecule captioning[8], where LLMs demonstrated strong capabilities in aligning chemical structures with natural language descriptions[33, 39, 51, 66]. Building on this progress, emerging studies[11] have begun to extend such approaches to more complex settings like drug–drug interaction prediction. However, the aforementioned approaches are rarely designed to handle out-of-distribution scenarios in DDI prediction and lack effective mechanisms to jointly leverage LLMs and GNNs for robust generalization.

2.2 Graph Neural Architecture Search

Graph Neural Architecture Search (GNAS) extends the principles of Neural Architecture Search (NAS) to the graph domain, aiming to automatically discover effective GNN architectures tailored to specific tasks [13, 25, 26, 46, 60, 64]. Early GNAS methods explored discrete search spaces using reinforcement learning [69] or evolutionary algorithms [58], which often incurred high computational costs. To address this, differentiable NAS techniques such as DARTS [30, 32], GRACES [45], and others [3, 4, 16, 43, 65, 68] have been proposed, enabling a more efficient search through over-parameterized supernets and continuous relaxation. Recent advances in GNAS span multiple directions, including robustness under distribution shifts [14, 28], multi-modal architecture design [57, 62], and scalable search for large graphs [44, 63].

In the context of DDI prediction, GNAS has shown promise in automatically learning task-specific molecular graph encoders. For example, AutoDDI [12] leverages NAS to design specialized GNN architectures that capture drug structural information more effectively, and CSSE-DDI[7] demonstrates the potential of applying NAS to optimize both subgraph selection and encoding strategies for improved interpretability and performance. However, existing methods still lack the ability to dynamically adapt architecture search results to specific drug pairs, and no current approaches integrate NAS with LLM-based modality fusion to address the challenges of OOD DDI scenarios.

3 Method

In this section, we present a thorough description of our **DyNAS-DDI**. Section 3.1 introduces the problem formalization, laying the foundation for our approach. Sections 3.2 and 3.3 elaborate on motif-driven disentangled molecule encoding and attention-based pairwise neural architecture search, respectively. Section 3.4 discusses retrieval-augmented generalizable molecular instruction tuning, and finally, Section 3.5 provides a comprehensive overview of the entire process.

3.1 Problem Formalization

Drug-drug interaction (DDI) refers to the effect that occurs when two drugs interact, potentially altering their efficacy or causing adverse effects. The problem can be formulated as follows:

Given a set of drug molecules $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$ and their interaction labels Y, the goal is to learn a predictive function $f : (d_i, d_j) \rightarrow y$, where $d_i, d_j \in D$ represent two drug molecules, and $y \in Y$ denotes their interaction type. The function f aims to model whether a given drug pair interacts (binary classification) and, if so, to predict the nature of their interaction (multi-class classification). In the context of OOD DDI, we partition the distribution space of all drug molecules based on factors such as molecular scaffolds, drug sizes, or assay conditions. Let $\mathcal{D}_{ID}$ be the subset of drug molecules that belong to a specific in-distribution (ID) denoted by $P_{ID}(d)$, while the OOD drug molecules form the set $\mathcal{D}_{OOD}$, which consists of molecules that do not belong to $P_{ID}(d)$. These two sets are disjoint, i.e.,$\mathcal{D}_{ID} \cap \mathcal{D}_{OOD} = \emptyset$, and their union covers the entire drug space: $\mathcal{D}_{ID} \cup \mathcal{D}_{OOD} = \mathcal{D}$. We define the training set as consisting of molecular pairs sampled from an $\mathcal{D}_{ID}$, while the validation and test sets contain $\mathcal{D}_{OOD}$ samples:

$$\begin{aligned} \mathcal{T}_{\text{train}} &= \{(d_i, d_j) \mid d_i, d_j \in \mathcal{D}_{\text{ID}}\} \\ \mathcal{T}_{\text{valid}} \cup \mathcal{T}_{\text{test}} &= \{(d_i, d_j) \mid d_i \in \mathcal{D}_{\text{OOD}} \vee d_j \in \mathcal{D}_{\text{OOD}}\} \end{aligned} \tag{1}$$

3.2 Motif-driven Disentangled Molecule Encoding

To capture interaction-invariant features and mitigate spurious correlations, we introduce *Motif-driven disentangled molecule encoding*, which separates core interaction motifs from global molecular context. This module consists of two components: *Disentangled motif embedding* and *Self-supervised molecule encoding.*

Disentangled Motif Embedding. Extracting informative molecular substructures is essential for robust drug-drug interaction modeling[24]. Existing graph-based methods often rely on subgraph selection techniques that use heuristic or learning-based methods to select critical nodes and edges. However, these methods may yield structurally disconnected or arbitrary subgraphs, especially in molecular contexts where functional groups play a crucial role. To generate more interpretable and chemically valid motifs, we adopt the BRICS algorithm[6], which partitions molecules into meaningful fragments guided by established reaction mechanisms. A molecular graph is represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ denote the set of atoms and chemical bonds, respectively. After applying the BRICS decomposition, we obtain a motif set: $\mathcal{M}_{\mathcal{G}} = \{M_{\mathcal{G},1}, M_{\mathcal{G},2}, \ldots, M_{\mathcal{G},m}\}$, where each $M_{\mathcal{G},i}$ is a subgraph of $\mathcal{G}$. We then select the top-k motifs based on molecular weight: $\mathcal{M}_{\mathcal{G}}^{(k)} = \{M_{\mathcal{G},i} \mid M_{\mathcal{G},i} \in \mathcal{M}_{\mathcal{G}}, \text{rank}(w(M_{\mathcal{G},i})) \le k\}$.

For each $M_i^{\mathcal{G}} \in \mathcal{M}_{\mathcal{G}}^{(k)}$, we compute its PageRank[38] importance within the molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The PageRank score of node v is as follows:

$$\text{PR}(v) = \alpha \sum_{u \in \mathcal{N}(v)} \frac{\text{PR}(u)}{|\mathcal{N}(u)|} + (1 - \alpha) p(v) \tag{2}$$

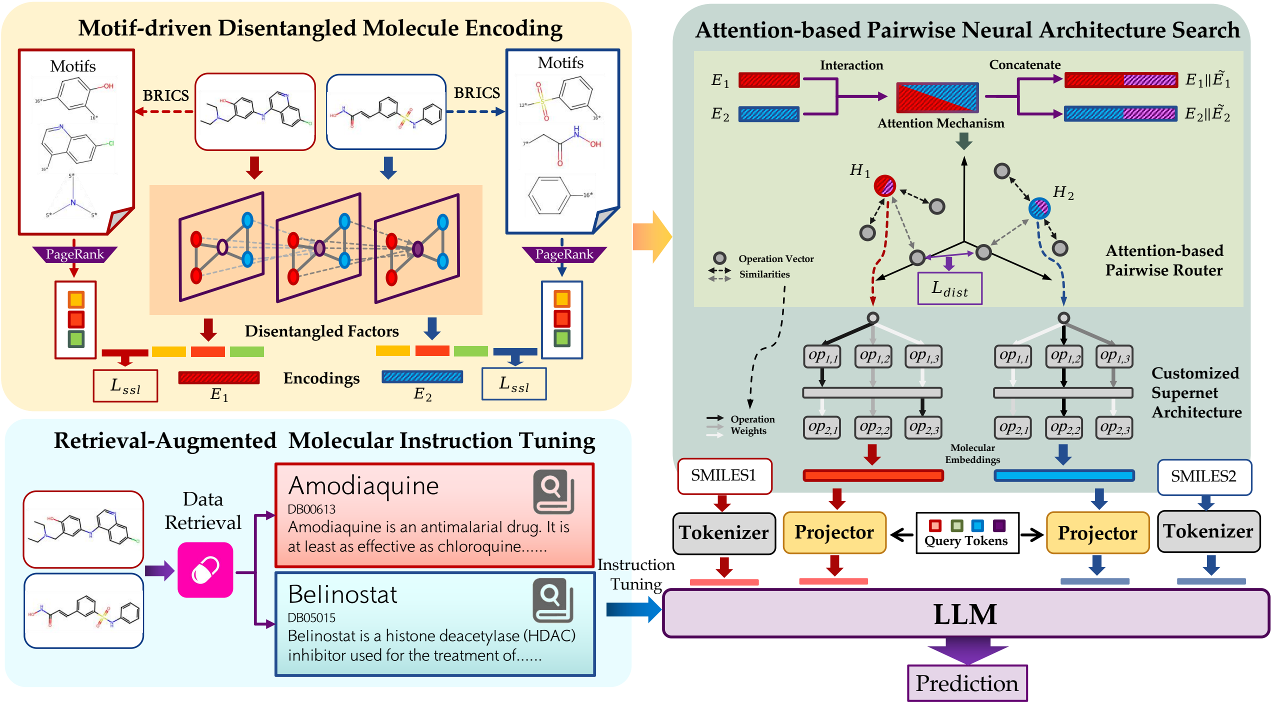
where α is the damping factor, $\mathcal{N}(v)$ denotes the set of neighbor nodes, and $p(v)$ is the prior probability distribution. The disentangled motif embedding is then aggregated as:

$$\mathcal{R}_{\mathcal{G}}^{(k)} = \{R(M_{\mathcal{G},i}) \mid M_{\mathcal{G},i} \in \mathcal{M}_{\mathcal{G}}^{(k)}\}, \quad R(M_{\mathcal{G},i}) = \sum_{v \in M_{\mathcal{G},i}} \text{PR}(v) \tag{3}$$

Through this PageRank method, motif importance is dynamically assigned based on their connectivity and influence in the molecular graph, allowing the model to capture essential interaction patterns and improve adaptability to out-of-distribution drug combinations.

Self-supervised Molecule Encoding. We propose a multi-view graph representation framework that integrates heterogeneous GNNs and readout functions to encode molecular structures. Formally, given the molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the encoding process is defined as:

$$\mathbf{z}_{\mathcal{G}} = \text{Readout}\left(\left\{ \Big\|_{k=1}^{K} \text{GNN}_k(v; \mathcal{G}) \,\middle|\, v \in \mathcal{V} \right\}\right) \tag{4}$$



Amodiaquine
DB00613
Amodiaquine is an antimalarial drug. It is at least as effective as chloroquine......

Belinostat
DB05015
Belinostat is a histone deacetylase (HDAC) inhibitor used for the treatment of......

Figure 1: The overall framework of DyNAS-DDI. It processes molecular graphs through two stages: *Motif-Driven disentangled molecule encoding* extracts structural patterns using self-supervised learning, and *attention-based pairwise neural architecture search* dynamically constructs task-specific graph neural networks. Then, projectors align the obtained graph features with large language model requirements. SMILES sequences are independently tokenized and combined with projected embeddings through the LLM backbone. Moreover, we design *retrieval-augmented molecular instruction tuning* to enhance molecular reasoning in out-of-distribution scenarios.

where $\|_{k=1}^{K}$ denotes the concatenation of node features from K independent GNNs. To align the molecular encoding with the disentangled motif embedding, we disentangle $\mathbf{z}_{\mathcal{G}}$ into multiple components, one representing the global structural information, and k components each corresponding to a motif in $\mathcal{M}_{\mathcal{G}}^{(k)}$. Formally, let $\mathbf{z}_{\mathcal{G}} \in \mathbb{R}^d$ be decomposed as: $\mathbf{z}_{\mathcal{G}} = \left[\mathbf{s}_{\mathcal{G},0}\|\mathbf{s}_{\mathcal{G},1}\|\mathbf{s}_{\mathcal{G},2}\|\ldots\|\mathbf{s}_{\mathcal{G},k}\right]$, $\mathbf{s}_{\mathcal{G},i} \in \mathbb{R}^{d/(k+1)}$. Each component $\mathbf{s}_i$ is trained to match the normalized PageRank score of motif $M_{\mathcal{G},i}$. We compute the self-supervised loss $\mathcal{L}_{\text{ssl},\mathcal{G}}$ as:

$$\mathcal{L}_{\text{ssl},\mathcal{G}} = \sum_{i=1}^{k} \left\|\phi(\mathbf{s}_{\mathcal{G},i}) - R(M_{\mathcal{G},i})\right\| \tag{5}$$

where $\phi : \mathbb{R}^{d/(k+1)} \to \mathbb{R}$ is a learnable projection head mapping component embeddings to scalar scores. In drug-drug interaction scenarios, the final $\mathcal{L}_{\text{ssl}}$ is implemented as the aggregation of $\mathcal{L}_{\text{ssl},\mathcal{G}}$ for each drug in the pair. Unlike methods relying on holistic graph embeddings, our motif-disentangled representation captures transferable interaction mechanisms and enables robust prediction when molecular topologies deviate from training distributions.

3.3 Attention-based Pairwise Neural Architecture Search

To dynamically design architectures that adapt to pairwise molecular interactions, we propose *Attention-based pairwise neural architecture search*. This approach integrates three key components: *Attention-based interaction modeling*, *Operation embedding space construction*, and *Pairwise adaptive architecture search*.

Attention-based Interaction Modeling. Given drug pair encodings $\mathbf{z}_{\mathcal{G}}, \mathbf{z}_{\mathcal{G}'} \in \mathbb{R}^d$, we model their interaction via multi-head attention[53] that explicitly attends to motif-disentangled components:

$$\mathbf{z}^*_{\mathcal{G}} = \mathbf{z}_{\mathcal{G}} + \text{FFN}\left(\mathcal{A}_{\text{multi}}(\mathbf{Q}, \mathbf{K}, \mathbf{V})\right) \tag{6}$$

where query vector $\mathbf{Q} = \mathbf{W}_Q[\mathbf{s}_{\mathcal{G},0}\|\mathbf{s}_{\mathcal{G},1}\ldots\|\mathbf{s}_{\mathcal{G},k}]$, key and value vector $\mathbf{K}, \mathbf{V} = \mathbf{W}_K[\mathbf{s}_{\mathcal{G}',0}\|\mathbf{s}_{\mathcal{G}',1}\|\ldots\|\mathbf{s}_{\mathcal{G}',k}]$. The mechanism is the same with $\mathbf{z}_{\mathcal{G}'}$ to produce $\mathbf{z}^*_{\mathcal{G}'}$. This design emphasizes motif-level compatibility between drugs, where attention weights reflect pharmacophore complementarity — a key factor in generalizing out-of-distribution DDIs with novel motif combinations.

Operation Embedding Space Construction. To dynamically adapt the GNN architecture to molecular substructure patterns, we propose an embedding-guided NAS algorithm. The core idea is to project the interaction-aware drug encoding $\mathbf{z}_{\mathcal{G}}^*$ onto a learnable operation space, determining the optimal GNN layer operations based on motif-driven compatibility.

For each GNN layer l, we define a candidate operation set $O^{(l)} = \{\mathrm{op}_1^l, \mathrm{op}_2^l, \ldots, \mathrm{op}_m^l\}$, where each operation op_i^l represents a specific message-passing mechanism. Each operation is associated with a trainable prototype vector $\mathbf{e}_i^{(l)} \in \mathbb{R}^{d_{\mathrm{op}}}$ for all $\mathrm{op}_i^l \in O^{(l)}$. These vectors form an *operation embedding space* $\mathcal{E}^{(l)} = \{\mathbf{e}_1^{(l)}, \ldots, \mathbf{e}_m^{(l)}\}$, initialized randomly and optimized during training. Moreover, to prevent these operation vectors from collapsing into a narrow region of the embedding space, we introduce a repulsive loss term that encourages uniform dispersion of $\mathbf{e}_i^{(l)}$ vectors. For each layer l, we define the dispersion loss as:

$$\mathcal{L}_{\mathrm{disp}}^{(l)} = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \frac{\mathbf{e}_i^{(l)} \cdot \mathbf{e}_j^{(l)}}{\|\mathbf{e}_i^{(l)}\|_2 \|\mathbf{e}_j^{(l)}\|_2} \tag{7}$$

The final loss $\mathcal{L}_{\mathrm{disp}}$ is thus computed by averaging $\mathcal{L}_{\mathrm{disp}}^{(l)}$ from each layer.

Pairwise Adaptive Architecture Search. At layer l, the probability of selecting operation op_i is determined by the similarity between $\mathbf{z}_{\mathcal{G}}^*$ and $\mathbf{e}_i^{(l)}$:

$$p^{(l)}(\mathrm{op}_i) = \frac{\exp\langle \mathbf{z}_{\mathcal{G}}^*, \mathbf{e}_i^{(l)} \rangle}{\sum_{j=1}^{m} \exp\langle \mathbf{z}_{\mathcal{G}}^*, \mathbf{e}_j^{(l)} \rangle} \tag{8}$$

where $\langle \cdot, \cdot \rangle$ denotes dot product measuring embedding-operation compatibility. The final operation at layer l is implemented as a weighted combination:

$$\mathbf{h}_{\mathcal{G}}^{(l+1)} = \sum_{i=1}^{m} p^{(l)}(\mathrm{op}_i) \cdot \mathrm{op}_i\left(\mathbf{h}_{\mathcal{G}}^{(l)}\right) \tag{9}$$

in which $\mathbf{h}_{\mathcal{G}}^{(l)}$ denotes the node features at layer l. This soft-selection strategy enables end-to-end differentiable optimization of both architecture weights and GNN parameters. This projection $\langle \mathbf{z}_{\mathcal{G}}^*, \mathbf{e}_i^{(l)} \rangle$ encodes motif-specific interaction patterns, guiding the NAS to prefer operations that can adapt to OOD drug pairs by emphasizing transferable message-passing schemes and suppressing noisy connections through operation dropout implicitly via low selection probabilities.

During architecture customization, operation selection probabilities are first computed via attention weighting based on the pair's interaction features, thus constructing unique context-aware computational paths for each molecule. The resulting $\mathbf{h}_{\mathcal{G}}$ and $\mathbf{h}_{\mathcal{G}'}$ encapsulate both intrinsic molecular characteristics and structural adaptation information relative to their interaction partner.

3.4 Retrieval-Augmented Molecular Instruction Tuning

To strengthen molecular reasoning under real-world knowledge constraints, we implement a targeted retrieval mechanism that selectively acquires partial but critical drug property profiles. Specifically, our system queries domain-specific databases (for example, PubChem[19], DrugBank[56]) to retrieve these sparse yet biochemically pivotal attributes. Upon retrieving these key drug attributes, we design structured instruction prompts that integrate molecular structures with their corresponding property profiles. During instruction tuning, the model then learns to predict missing properties given the structural and textual prompts, thereby enriching its understanding and reasoning of drug attributes. This phase employs a standard causal language modeling objective:

$$\mathcal{L}_{\mathrm{inst}} = \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[-\sum_{t=1}^{T} y_t \log p_\theta(y_t | x_{\leq t})\right] \tag{10}$$

where x represents the structured instruction prompts, while y denotes target property values. The model processes this combined input through the transformer's self-attention layers to predict target sequences autoregressively. For DDI classification training, $\mathcal{L}_{\mathrm{target}}$ follows the same formulation, where x utilizes a different prompt format and y represents interaction text labels. This consistency preserves the model's pre-trained biochemical reasoning capabilities while adapting to the new task. Full prompt designs appear in Section 3.5.

For this process, the composite loss function combines three key objectives:

$$\mathcal{L} = \mathbb{I}_{\mathrm{stage}} \cdot \mathcal{L}_{\mathrm{inst}} + (1 - \mathbb{I}_{\mathrm{stage}}) \cdot \mathcal{L}_{\mathrm{target}} + \alpha\mathcal{L}_{\mathrm{disp}} + \beta\mathcal{L}_{\mathrm{ssl}} \tag{11}$$

where $\mathbb{I}_{\mathrm{stage}}$ is an indicator function (1 for instruction tuning, 0 for DDI task training). This unified process strengthens the model's ability to generalize to out-of-distribution scenarios, enabling it to apply its biochemical knowledge more effectively.

3.5 Overall Framework

In this section, we will provide a detailed description of the overall model architecture and training process. The overall architecture of our proposed framework is illustrated in Figure 1.

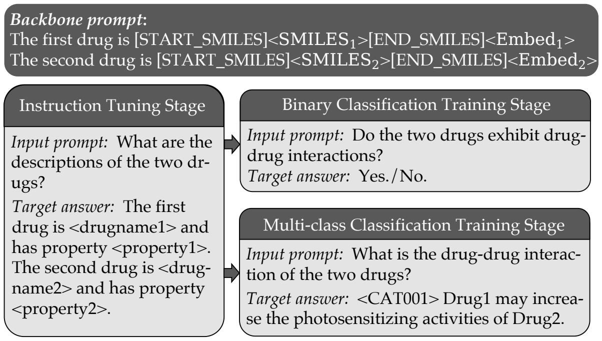
Data Propagation. For a drug pair denoted as (d_1, d_2), the Simplified Molecular Input Line Entry System (SMILES) representations are given as $\mathtt{SMILES_1}$ and $\mathtt{SMILES_2}$ respectively, and the molecular graph structures are formally represented as $\mathcal{G}_1$ and $\mathcal{G}_2$. The molecular graphs are processed through motif-driven disentangled molecule encoding to derive self-supervised disentangled representations $\mathbf{z}_{\mathcal{G}_1}$ and $\mathbf{z}_{\mathcal{G}_2}$. These representations subsequently guide attention-based pairwise neural architecture search to dynamically determine architectural weights $\{p(\mathrm{op}_i)\}$ for candidate graph operations $\{\mathrm{op}_i\}$. The resultant GNN architecture then hierarchically aggregates structure patterns from $\mathcal{G}_1$ and $\mathcal{G}_2$, ultimately yielding adaptive graph-level embeddings $\mathbf{h}_{\mathcal{G}_1}$ and $\mathbf{h}_{\mathcal{G}_2}$ through layer-wise attentive message passing and dynamic operation fusion.

To project the dynamic graph representations into the language model's latent space, we employ BERT-based projectors [27] (f_{pro1}, f_{pro2}) enhanced with cross-attention mechanisms, transforming them into LLM-compatible embeddings $\mathbf{E}_1$ and $\mathbf{E}_2$ that preserve graph-structured semantics while aligning with the LLM's hidden dimension. To incorporate sequential molecular information, the SMILES sequences are tokenized through the LLM's native encoder into discrete tokens $\mathbf{S}_1$ and $\mathbf{S}_2$. The integrated multimodal sequence $[\mathbf{E}_1, \mathbf{S}_1, \mathbf{E}_2, \mathbf{S}_2]$ formatted with prompt $\mathbf{P}$ is processed through an

Table 1: OOD Dataset Splitting Statistics based on Scaffold and Size.

Dataset	Threshold		SCAFFOLD Splitting					SIZE Splitting				
	Scaffold(Mol)	Size(Da)	$\mathcal{D}_{\text{ID}}$	$\mathcal{D}_{\text{OOD}}$	$\mathcal{D}_{\text{train}}$	$\mathcal{D}_{\text{valid}}$	$\mathcal{D}_{\text{test}}$	$\mathcal{D}_{\text{ID}}$	$\mathcal{D}_{\text{OOD}}$	$\mathcal{D}_{\text{train}}$	$\mathcal{D}_{\text{valid}}$	$\mathcal{D}_{\text{test}}$
ChChMiner	10	260	802	157	23019	5325	5325	764	195	22698	5485	5485
DeepDDI	8	250	1319	385	207941	54327	54327	1323	381	202077	57259	57259
ZhangDDI	7	245	442	102	75389	19291	19291	443	101	75876	19048	19048
Drugbank	9	245	1282	422	127525	32172	32173	1323	381	129431	31219	31220

LLM backbone galactica-1.3b[52] pretrained on scientific corpora following MolTC[11].



Backbone prompt:
The first drug is [START_SMILES]<SMILES$_1$>[END_SMILES]<Embed$_1$>
The second drug is [START_SMILES]<SMILES$_2$>[END_SMILES]<Embed$_2$>

Instruction Tuning Stage

Input prompt: What are the descriptions of the two drugs?

Target answer: The first drug is <drugname1> and has property <property1>. The second drug is <drugname2> and has property <property2>.

Binary Classification Training Stage

Input prompt: Do the two drugs exhibit drug-drug interactions?

Target answer: Yes./No.

Multi-class Classification Training Stage

Input prompt: What is the drug-drug interaction of the two drugs?

Target answer: <CAT001> Drug1 may increase the photosensitizing activities of Drug2.

Figure 2: Prompt design across different stages. The backbone prompt provides consistent molecular identifiers and base instruction templates. During instruction tuning, we incorporate drug property profiles retrieved via RAG as prediction targets. For binary classification tasks, single-token outputs (`Yes./No.`) are employed, while multi-class classification introduces specialized tokens (`<CATXXX>`) paired with detailed reaction type descriptions as targets.

Training Pipeline. Our training procedure consists of two synergistic phases. During the instruction tuning phase, we employ retrieval-augmented molecular instruction tuning to enrich the LLM's biochemical knowledge base with pre-designed structured prompts $\{P_i\}$ to integrate molecular graphs with drug property corpora retrieved from authoritative databases. Subsequently, the domain adaptation phase utilizes DDI datasets with task-oriented prompts $\{P_t\}$, where the complete data propagation pipeline guides the model to generate classification outputs (binary or multi-class). This phased approach ensures that the model first establishes fundamental biochemical comprehension before specializing in molecular interaction reasoning. Our prompt design is illustrated in Figure 2.

4 Experiments

To rigorously evaluate our model's capability in both ID and OOD scenarios, we conduct comprehensive experiments across four benchmark DDI datasets. We adopt both scaffold-based and size-stratified splitting to partition each dataset into ID and OOD sets, ensuring that structurally distinct molecules in the OOD set never appear in training. Extensive comparisons with various categories of state-of-the-art baselines demonstrate our method's superior performance across all evaluation settings.

4.1 Experimental Settings

Dataset Construction. Following the methodology defined in Section 3.1, we construct evaluation datasets using four established DDI benchmarks: ChChMiner[35], DeepDDI[49], ZhangDDI[61], and DrugBank[56]. For each dataset $\mathcal{D}$, we implement dual threshold-based partitioning: (i) *scaffold-based* splitting via Bemis-Murcko[2] frameworks with molecular structure thresholds, and (ii) *size-based* splitting using molecular weight thresholds. Molecules exceeding the thresholds are assigned to $\mathcal{D}_{\text{ID}}$, while those below form $\mathcal{D}_{\text{OOD}}$. This is done to ensure distinct structural and physicochemical distributions between the two sets. The data is then divided into train/valid/test subsets while maintaining an approximately 4:1:1 ratio through stratified sampling as specified in Section 3.1. Detailed dataset information is in Table 1.

Baselines. We conduct comprehensive benchmarking against three categories of baseline methods: (i) Conventional GNN models: CIGIN[40], SSI-DDI[37], CMRL[24], CGIB[22], DSN-DDI[29] and DeepDDS[55]; (ii) Non-GNN machine learning approaches: DeepDDI[49], MHCADDI[5] and MatchMaker[20] utilizing traditional feature engineering; (iii) State-of-the-art LLM-based methods: including Galactica[52], MolT5[9], and MolTC[11]. This multi-paradigm comparison ensures rigorous evaluation across different architectural philosophies and learning mechanisms.

Metrics. We employ three standard classification metrics: (i) *Accuracy*, measuring overall prediction correctness, (ii) *Area Under the Receiver Operating Characteristic Curve (AUC-ROC)* evaluating model discrimination ability across all classification thresholds, and (iii) *F1-score* balancing precision and recall for unbalanced interaction classes.

Details. Our optimization framework employs AdamW[34] optimizer with $\epsilon = 1\text{e}{-8}$ and weight decay regularization ($\lambda = 0.05$) to prevent overfitting. The learning rate follows a hybrid scheduling strategy: linearly warming up from 1e−6 to 1e−4, then decaying through cosine annealing to 1e−5. Parameter groups are assigned separate learning rates, but all follow the same global scheduling curve. For parameter-efficient adaptation, we integrate Low-Rank Adaptation (LoRA)[17], configuring trainable matrices ($r = 16$) on Galactica's query/key projections and fully connected layers, while freezing 99.8% of the base LLM parameters. The cross-modal projectors are initialized using SciBERT embeddings[1]. The majority of our experiments are performed using six NVIDIA GeForce RTX 3090 GPUs (24GB). More details are illustrated in the Appendix.

Table 2: Comparative performance on scaffold-based and size-based OOD DDI prediction tasks. The top-performing method is highlighted in bold, with second-best results underlined across all evaluation metrics. We also report the percentage improvement of these metrics compared to the second-best performing method.

Setting	Model	ChChMiner			ZhangDDI			DeepDDI			DrugBank		
		Acc ↑	AUC ↑	F1 ↑	Acc ↑	AUC ↑	F1 ↑	Acc ↑	AUC ↑	F1 ↑	Acc ↑	AUC ↑	F1 ↑
Scaffold-Based OOD DDI Prediction Tasks													
GNN-Based	CIGIN	63.94	69.29	67.70	63.81	58.77	16.72	48.88	74.10	32.10	47.82	80.78	43.90
	SSI-DDI	55.13	57.39	46.86	53.36	54.45	50.38	53.80	55.03	53.28	57.46	75.96	54.19
	DSN-DDI	51.93	52.93	52.40	43.12	39.13	21.37	54.42	57.06	50.15	51.00	73.36	58.38
	CMRL	67.29	72.23	70.64	66.72	68.75	42.37	66.83	74.27	66.06	52.07	77.41	47.75
	CGIB	66.82	71.82	70.32	68.05	68.21	46.01	66.15	72.55	66.50	51.95	78.35	48.32
	DeepDDS	65.69	68.27	67.70	63.32	53.46	49.11	65.77	71.06	65.15	36.46	47.23	19.49
ML-Based	DeepDDI	64.26	69.42	66.63	67.95	69.61	51.26	69.49	76.26	69.37	54.25	79.61	50.53
	MHCADDI	53.78	51.58	68.83	63.37	53.92	49.17	51.18	55.87	35.02	16.12	0.50	0.09
	MatchMaker	68.81	70.99	72.07	64.58	66.80	36.45	65.49	72.01	65.46	14.13	48.34	0.03
LLM-Based	Galactica	75.12	83.43	74.96	69.51	72.86	66.00	69.36	76.89	69.17	48.59	75.98	44.48
	MolT5	52.53	50.01	34.44	54.59	50.50	54.23	51.11	50.36	34.58	9.11	50.22	10.99
	MolTC	72.57	82.88	76.47	68.51	71.70	54.60	68.29	74.90	66.97	52.23	79.81	50.02
	DyNAS-DDI	**79.44**	**87.22**	**79.35**	**72.93**	**77.58**	**72.54**	**71.48**	**78.94**	**72.62**	**64.86**	**88.86**	**61.27**
	%↑	**+5.75%**	**+4.54%**	**+3.77%**	**+4.92%**	**+6.48%**	**+9.91%**	**+2.86%**	**+2.67%**	**+4.69%**	**+12.88%**	**+10.00%**	**+4.95%**
Size-Based OOD DDI Prediction Tasks													
GNN-Based	CIGIN	66.89	72.12	66.05	63.00	61.02	48.70	44.83	71.19	27.75	48.40	81.91	44.39
	SSI-DDI	58.95	63.20	53.94	55.04	56.84	53.30	53.76	54.91	53.46	58.72	71.98	65.20
	DSN-DDI	53.89	56.94	38.66	53.88	56.92	44.34	55.12	58.07	51.59	34.98	76.58	36.64
	CMRL	74.19	82.86	77.83	69.04	72.10	50.18	71.14	78.77	73.61	56.01	80.92	53.24
	CGIB	71.38	78.85	75.25	65.33	67.07	38.23	69.91	76.35	70.70	52.82	79.83	48.72
	DeepDDS	61.28	64.41	61.43	66.19	66.29	64.84	69.25	72.52	68.88	39.07	42.05	21.95
ML-Based	DeepDDI	64.56	69.79	64.28	66.85	70.55	66.90	71.81	78.17	70.58	59.46	84.69	55.86
	MHCADDI	59.15	54.82	52.61	62.75	54.01	50.44	59.95	65.24	53.53	17.25	47.04	17.72
	MatchMaker	63.12	65.61	61.52	67.36	68.99	63.60	66.53	73.74	66.26	18.68	47.66	11.95
LLM-Based	Galactica	79.48	88.50	79.23	69.88	74.33	69.31	65.40	70.77	65.32	53.23	80.37	49.67
	MolT5	56.58	49.76	36.14	47.96	50.00	48.45	55.17	50.51	39.23	37.97	48.66	22.29
	MolTC	77.73	86.82	80.45	67.57	72.37	67.66	71.06	78.02	71.00	56.34	80.33	53.97
	DyNAS-DDI	**81.74**	**90.62**	**83.16**	**76.62**	**83.61**	**75.95**	**73.56**	**80.81**	**79.76**	**67.97**	**90.23**	**65.64**
	%↑	**+2.84%**	**+2.40%**	**+3.37%**	**+9.65%**	**+12.48%**	**+9.58%**	**+2.44%**	**+2.59%**	**+8.35%**	**+14.31%**	**+6.54%**	**+0.67%**

Our training regimen consists of an initial 10-epoch instruction tuning followed by a 50-epoch task-specific fine-tuning for DDI classification. For performance measurement, we compute prediction scores by applying softmax normalization over the LLM's full vocabulary logits to derive the AUC-ROC metric. Categorical accuracy and F1 scores are determined through strict token-level matching between generated responses and target labels.

4.2 Scaffold-based Interaction Results

Our experimental results demonstrate **DyNAS-DDI**'s superior generalization capabilities under scaffold-based OOD splits. As shown in Table 2, **DyNAS-DDI** achieves state-of-the-art performance on all evaluation metrics, outperforming all baselines from three categories. Compared to GNN-based approaches and ML-based methods, **DyNAS-DDI** achieves an average 10.3% improvement in accuracy, validating that our dynamic architecture search effectively adapts to structural distribution shifts in scaffold-based OOD scenarios. Although accuracy reflects overall generalizability, our method also shows consistent advantages in other metrics: it outperforms all baselines by more than 4% in AUC, demonstrating superior ranking capability for novel scaffolds, and exceeds the second-best MolTC by an average of 8% in F1-score, indicating a better precision-recall balance in OOD settings. Even compared to LLM-based methods that inherently leverage chemical knowledge, **DyNAS-DDI**maintains an average improvement of 5% in all metrics, due to our retrieval-augmented instructions that inject domain-specific interaction patterns and dynamic GNN-LLM coupling that adaptively fuses structural and semantic features.

In particular, our results highlight the stronger advantages of our method in the more challenging DrugBank dataset, where traditional machine learning approaches sometimes fail to achieve meaningful performance in this complex 86-class prediction OOD scenario. However, our framework demonstrates significantly more robust performance, demonstrating its superior capability in handling difficult real-world DDI prediction scenarios.

4.3 Size-based Interaction Results

Unlike scaffold splits that create abrupt discontinuities in chemical space, size-based divisions maintain partial overlap as molecular size correlates slightly weaker with chemical reactivity patterns. This explains the universally observed performance elevation across all methods, as key pharmacophores may remain recognizable despite size variations. Our approach extends its inherent advantage across all metrics, allowing a more robust transfer of chemical knowledge across size boundaries. This advantage is more evident on the Drugbank dataset, highlighting the robustness of our model architecture in handling more complex and demanding OOD DDI scenarios. Specifically, compared to GNN- and ML-based methods, our **DyNAS-DDI** demonstrates superior performance across multiple evaluation metrics, achieving improvements ranging from approximately 6-8% (minimum) to over 20% (maximum). Moreover, compared to LLM-based approaches, **DyNAS-DDI** maintains a consistent competitive advantage. These results indicate that our method exhibits strong generalizability across different OOD settings through its adaptive architecture and knowledge-enhanced molecular reasoning.

4.4 Ablation Study

In this section, we conduct ablation studies to evaluate the impact of different components on model performance. The experiments are performed on scaffold-based OOD ChChMiner and Drugbank datasets using Accuracy and AUC-ROC as the primary evaluation metrics. Specifically, we evaluate the following variants:

- **w/o R-IT**: Without *Retrieval-augmented Instruction Tuning*, meaning the model is trained directly without the instruction fine-tuning stage.
- **w/o Att**: Without *Attention-based interaction modeling*, where molecule interactions are modeled using initial encodings instead of attention mechanisms.
- **w/o $\mathcal{L}_{ssl}$**: Without *Disentangled Motif Embedding*, meaning that self-supervised motif-based disentanglement is removed.
- **w/o $\mathcal{L}_{cos}$**: Without *Operation Embedding Space Construction*, meaning the operation vector constraints are removed.
- **Handcraft**: The entire molecule embedding pipeline is replaced with a handcrafted GNN-based molecular representation.

The ablation results are summarized in Table 3. According to the results, our full model achieves the best performance across both datasets, demonstrating the effectiveness of each component in handling distribution shifts. Notably, the exclusion of *Attention-based interaction modeling* (w/o Att) results in the most significant decline in accuracy, which indicates its crucial role in capturing fine-grained pairwise molecular interactions. Furthermore, the performance gap is especially pronounced on Drugbank when using a handcrafted GNN. This highlights the advantage of our dynamic pairwise architecture search, which enables better adaptation to diverse drug-pair patterns and improved generalization.

Table 3: Ablation study results.

Dataset	ChChMiner		Drugbank	
	Accuracy	AUC-ROC	Accuracy	AUC-ROC
Full Model	**79.44**	**87.22**	**64.86**	**88.86**
w/o R-IT	77.41	86.58	61.77	86.12
w/o Att	72.15	81.38	58.68	86.30
w/o $\mathcal{L}_{ssl}$	73.45	83.62	61.05	87.17
w/o $\mathcal{L}_{disp}$	72.68	85.55	61.37	85.30
handcraft	72.57	82.88	52.23	79.81

4.5 Complexity Analysis

For a molecular graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let d_z denote the dimensionality of the self-supervised molecular encoding $\mathbf{z}_{\mathcal{G}}$ in Section 3.2, and d_h denote the dimensionality of the molecule embedding $\mathbf{h}_{\mathcal{G}}$ obtained from the dynamically customized neural network in Section 3.3. We use O to represent the set of candidate operations in the architecture search space and $|\cdot|$ to indicate the cardinality of a set. The self-supervised disentangled molecule encoder incurs a cost of $O(|\mathcal{E}|d_z + |\mathcal{V}|d_z^2)$. The attention mechanism in interaction modeling introduces an overhead of $O(d_z^2)$. For the dynamic architecture search module, the primary computational burden lies in computing $\mathcal{L}_{\text{disp}}$, which requires $O(|O|^2 d_z)$. The customized super-network performs message passing based on the searched architecture, contributing a complexity of $O(|O|(|\mathcal{E}|d_h + |\mathcal{V}|d_h^2))$.

Combining all components, the total complexity of our framework concerning graph modeling becomes $O(|\mathcal{E}|(d_z + |O|d_h) + |\mathcal{V}|(d_z^2 + |O|d_h^2) + |O|^2 d_z + d_z^2)$, which remains linear in graph size $(\mathcal{V}, \mathcal{E})$. For the LLM part, the reasoning complexity is $O(N(Ld_l^2 + L^2 d_l))$, where N is the number of layers, L is the input token length, and d_l is the hidden size.

5 Conclusion

In our work, we address the challenge of distribution shifts in DDI prediction by proposing a novel framework that rethinks how molecular information is encoded, processed, and adapted. Rather than relying on static GNN architectures or assuming an I.I.D. setting, our attention-based pairwise architecture search dynamically adjusts the message-passing strategy for each molecular pair, enabling finer-grained modeling of pairwise interactions. The motif-driven disentangled molecule encoder ensures that core interaction features are preserved and decoupled from irrelevant structural noise, while the integration of external biomedical knowledge via retrieval-augmented molecular instruction tuning provides additional robustness when reasoning about unseen drug combinations. Through comprehensive evaluations on multiple benchmark datasets and under realistic OOD splits, we demonstrate consistent performance gains over existing methods. These findings highlight the importance of incorporating dynamic, pairwise-aware, and knowledge-informed mechanisms for building reliable and generalizable DDI models.

6 Acknowledgments

This work was supported by the National Key Research and Development Program of China No.2023YFF1205001, National Natural Science Foundation of China No. 62222209, Beijing National Research Center for Information Science and Technology under Grant No. BNR2023TD03006, and Beijing Key Lab of Networked Multimedia.

References

[1] Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. SciBERT: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676* (2019).

[2] Guy W Bemis and Mark A Murcko. 1996. The properties of known drugs. 1. Molecular frameworks. *Journal of medicinal chemistry* 39, 15 (1996), 2887–2893.

[3] Jie Cai, Xin Wang, Chaoyu Guan, Yateng Tang, Jin Xu, Bin Zhong, and Wenwu Zhu. 2022. Multimodal continual graph learning with neural architecture search. In *Proceedings of the ACM Web Conference 2022*. 1292–1300.

[4] Jie Cai, Xin Wang, Haoyang Li, Ziwei Zhang, and Wenwu Zhu. 2024. Multimodal graph neural architecture search under distribution shifts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 8227–8235.

[5] Andreea Deac, Yu-Hsiang Huang, Petar Veličković, Pietro Liò, and Jian Tang. 2019. Drug-drug adverse effect prediction with graph co-attention. *arXiv preprint arXiv:1905.00534* (2019).

[6] Jorg Degen, Christof Wegscheid-Gerlach, Andrea Zaliani, and Matthias Rarey. 2008. On the art of compiling and using'drug-like'chemical fragment spaces. *ChemMedChem* 3, 10 (2008), 1503.

[7] Haotong Du, Quanming Yao, Juzheng Zhang, Yang Liu, and Zhen Wang. 2024. Customized subgraph selection and encoding for drug-drug interaction prediction. *Advances in Neural Information Processing Systems* 37 (2024), 109582–109608.

[8] Carl Edwards, Tuan Lai, Kevin Ros, Garrett Honke, Kyunghyun Cho, and Heng Ji. 2022. Translation between molecules and natural language. *arXiv preprint arXiv:2204.11817* (2022).

[9] Carl Edwards, Tuan Lai, Kevin Ros, Garrett Honke, Kyunghyun Cho, and Heng Ji. 2022. Translation between molecules and natural language. *arXiv preprint arXiv:2204.11817* (2022).

[10] Carl Edwards, ChengXiang Zhai, and Heng Ji. 2021. Text2mol: Cross-modal molecule retrieval with natural language queries. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 595–607.

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

[12] Jianliang Gao, Zhenpeng Wu, Raeed Al-Sabri, Babatounde Moctard Oloulade, and Jiamin Chen. 2024. Autoddi: drug–drug interaction prediction with automated graph neural network. *IEEE Journal of Biomedical and Health Informatics* 28, 3 (2024), 1773–1784.

[13] Yang Gao, Hong Yang, Peng Zhang, Chuan Zhou, and Yue Hu. 2020. Graph Neural Architecture Search. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. 1403–1409.

[14] Chendi Ge, Xin Wang, Ziwei Zhang, Yijian Qin, Hong Chen, Haiyang Wu, Yang Zhang, Yuekui Yang, and Wenwu Zhu. 2025. Behavior importance-aware graph neural architecture search for cross-domain recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 11708–11716.

[15] Assaf Gottlieb, Gideon Y Stein, Yoram Oron, Eytan Ruppin, and Roded Sharan. 2012. INDI: a computational framework for inferring drug interactions and their associated recommendations. *Molecular systems biology* 8, 1 (2012), 592.

[16] Chaoyu Guan, Xin Wang, Hong Chen, Ziwei Zhang, and Wenwu Zhu. 2022. Large-scale graph neural architecture search. In *International Conference on Machine Learning*. PMLR, 7968–7981.

[17] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. *ICLR* 1, 2 (2022), 3.

[18] Yuanfeng Ji, Lu Zhang, Jiaxiang Wu, Bingzhe Wu, Lanqing Li, Long-Kai Huang, Tingyang Xu, Yu Rong, Jie Ren, Ding Xue, et al. 2023. Drugood: Out-of-distribution dataset curator and benchmark for ai-aided drug discovery–a focus on affinity prediction problems with noise annotations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 8023–8031.

[19] Sunghwan Kim, Paul A Thiessen, Evan E Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A Shoemaker, et al. 2016. PubChem substance and compound databases. *Nucleic acids research* 44, D1 (2016), D1202–D1213.

[20] Halil Ibrahim Kuru, Oznur Tastan, and A Ercument Cicek. 2021. MatchMaker: a deep learning framework for drug synergy prediction. *IEEE/ACM transactions on computational biology and bioinformatics* 19, 4 (2021), 2334–2344.

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

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

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

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

[25] Haoyang Li, Xin Wang, Ziwei Zhang, and Wenwu Zhu. 2022. Out-of-distribution generalization on graphs: A survey. *arXiv preprint arXiv:2202.07987* (2022).

[26] Haoyang Li, Xin Wang, Xueling Zhu, Weigao Wen, and Wenwu Zhu. 2025. Disentangling invariant subgraph via variance contrastive estimation under distribution shifts. In *Forty-second International Conference on Machine Learning*.

[27] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*. PMLR, 19730–19742.

[28] Peiwen Li, Xin Wang, Zeyang Zhang, Yijian Qin, Ziwei Zhang, Jialong Wang, Yang Li, and Wenwu Zhu. 2024. Causal-aware graph neural architecture search under distribution shifts. *arXiv preprint arXiv:2405.16489* (2024).

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

[30] Hanwen Liang, Shifeng Zhang, Jiacheng Sun, Xingqiu He, Weiran Huang, Kechen Zhuang, and Zhenguo Li. 2019. Darts+: Improved differentiable architecture search with early stopping. *arXiv preprint arXiv:1909.06035* (2019).

[31] Shenggeng Lin, Yanjing Wang, Lingfeng Zhang, Yanyi Chu, Yatong Liu, Yitian Fang, Mingming Jiang, Qiankun Wang, Bowen Zhao, Yi Xiong, et al. 2022. MDF-SA-DDI: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Briefings in Bioinformatics* 23, 1 (2022), bbab421.

[32] Hanxiao Liu, Karen Simonyan, and Yiming Yang. 2019. Darts: Differentiable architecture search. In *Proceedings of the 7th International Conference on Learning Representations*.

[33] Zhiyuan Liu, Sihang Li, Yanchen Luo, Hao Fei, Yixin Cao, Kenji Kawaguchi, Xiang Wang, and Tat-Seng Chua. 2023. Molca: Molecular graph-language modeling with cross-modal projector and uni-modal adapter. *arXiv preprint arXiv:2310.12798* (2023).

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

[35] Sagar Maheshwari Marinka Zitnik, Rok Sosič and Jure Leskovec. 2018. BioSNAP Datasets: Stanford Biomedical Network Dataset Collection. http://snap.stanford.edu/biodata.

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

[37] Arnold K Nyamabo, Hui Yu, and Jian-Yu Shi. 2021. SSI–DDI: substructure–substructure interactions for drug–drug interaction prediction. *Briefings in Bioinformatics* 22, 6 (2021), bbab133.

[38] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford infolab.

[39] Jinyoung Park, Minseong Bae, Dohwan Ko, and Hyunwoo J Kim. 2024. LLaMo: Large Language Model-based Molecular Graph Assistant. *arXiv preprint arXiv:2411.00871* (2024).

[40] Yashaswi Pathak, Siddhartha Laghuvarapu, Sarvesh Mehta, and U Deva Priyakumar. 2020. Chemically interpretable graph interaction network for prediction of pharmacokinetic properties of drug-like molecules. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 873–880.

[41] Qizhi Pei, Lijun Wu, Kaiyuan Gao, Jinhua Zhu, Yue Wang, Zun Wang, Tao Qin, and Rui Yan. 2024. Leveraging biomolecule and natural language through multi-modal learning: A survey. *arXiv preprint arXiv:2403.01528* (2024).

[42] Mark N Prichard and Charles Shipman Jr. 1990. A three-dimensional model to analyze drug-drug interactions. *Antiviral research* 14, 4-5 (1990), 181–205.

[43] Yijian Qin, Xin Wang, Peng Cui, and Wenwu Zhu. 2021. Gqnas: Graph q network for neural architecture search. In *2021 IEEE International Conference on Data Mining (ICDM)*. IEEE, 1288–1293.

[44] Yijian Qin, Xin Wang, Ziwei Zhang, Hong Chen, and Wenwu Zhu. 2023. Multi-task graph neural architecture search with task-aware collaboration and curriculum. *Advances in neural information processing systems* 36 (2023), 24879–24891.

[45] Yijian Qin, Xin Wang, Ziwei Zhang, Pengtao Xie, and Wenwu Zhu. 2022. Graph neural architecture search under distribution shifts. In *International Conference on Machine Learning*. 18083–18095.

[46] Yijian Qin, Xin Wang, Zeyang Zhang, and Wenwu Zhu. 2021. Graph Differentiable Architecture Search with Structure Learning. In *Advances in neural information processing systems*.

[47] Mayk Caldas Ramos, Christopher J Collison, and Andrew D White. 2025. A review of large language models and autonomous agents in chemistry. *Chemical Science* (2025).

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

[49] Jae Yong Ryu, Hyun Uk Kim, and Sang Yup Lee. 2018. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the national*

academy of sciences 115, 18 (2018), E4304–E4311.

[50] Xu Shen, Yili Wang, Kaixiong Zhou, Shirui Pan, and Xin Wang. 2024. Optimizing ood detection in molecular graphs: A novel approach with diffusion models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2640–2650.

[51] Bing Su, Dazhao Du, Zhao Yang, Yujie Zhou, Jiangmeng Li, Anyi Rao, Hao Sun, Zhiwu Lu, and Ji-Rong Wen. 2022. A molecular multimodal foundation model associating molecule graphs with natural language. *arXiv preprint arXiv:2209.05481* (2022).

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

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

[54] Brian Walsh, Sameh K Mohamed, and Vít Nováček. 2020. Biokg: A knowledge graph for relational learning on biological data. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 3173–3180.

[55] Jinxian Wang, Xuejun Liu, Siyuan Shen, Lei Deng, and Hui Liu. 2022. DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Briefings in Bioinformatics* 23, 1 (2022), bbab390.

[56] David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, et al. 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* 46, D1 (2018), D1074–D1082.

[57] Beini Xie, Heng Chang, Ziwei Zhang, Zeyang Zhang, Simin Wu, Xin Wang, Yuan Meng, and Wenwu Zhu. 2024. Towards lightweight graph neural network search with curriculum graph sparsification. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3563–3573.

[58] Lingxi Xie and Alan Yuille. 2017. Genetic cnn. In *Proceedings of the IEEE international conference on computer vision*. 1379–1388.

[59] Nianzu Yang, Kaipeng Zeng, Qitian Wu, Xiaosong Jia, and Junchi Yan. 2022. Learning substructure invariance for out-of-distribution molecular representations. *Advances in Neural Information Processing Systems* 35 (2022), 12964–12978.

[60] Yang Yao, Xin Wang, Yijian Qin, Ziwei Zhang, Wenwu Zhu, and Hong Mei. 2024. Customized Cross-device Neural Architecture Search with Images. In *2024 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.

[61] Wen Zhang, Yanlin Chen, Feng Liu, Fei Luo, Gang Tian, and Xiaohong Li. 2017. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC bioinformatics* 18 (2017), 1–12.

[62] Zeyang Zhang, Xin Wang, Yijian Qin, Hong Chen, Ziwei Zhang, Xu Chu, and Wenwu Zhu. 2024. Disentangled continual graph neural architecture search with invariant modular supernet. In *Forty-first International Conference on Machine Learning*.

[63] Zeyang Zhang, Xin Wang, Ziwei Zhang, Guangyao Shen, Shiqi Shen, and Wenwu Zhu. 2023. Unsupervised graph neural architecture search with disentangled self-supervision. *Advances in Neural Information Processing Systems* 36 (2023), 73175–73190.

[64] Ziwei Zhang, Xin Wang, and Wenwu Zhu. 2021. Automated Machine Learning on Graphs: A Survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. Survey track.

[65] Zeyang Zhang, Ziwei Zhang, Xin Wang, Yijian Qin, Zhou Qin, and Wenwu Zhu. 2023. Dynamic heterogeneous graph attention neural architecture search. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. 11307–11315.

[66] Haiteng Zhao, Shengchao Liu, Ma Chang, Hannan Xu, Jie Fu, Zhihong Deng, Lingpeng Kong, and Qi Liu. 2023. Gimlet: A unified graph-text model for instruction-based molecule zero-shot learning. *Advances in neural information processing systems* 36 (2023), 5850–5887.

[67] Yi Zhong, Xueyu Chen, Yu Zhao, Xiaoming Chen, Tingfang Gao, and Zuquan Weng. 2019. Graph-augmented convolutional networks on drug-drug interactions prediction. *arXiv preprint arXiv:1912.03702* (2019).

[68] Yuwei Zhou, Xin Wang, Hong Chen, Xuguang Duan, Chaoyu Guan, and Wenwu Zhu. 2022. Curriculum-nas: Curriculum weight-sharing neural architecture search. In *Proceedings of the 30th ACM International Conference on Multimedia*. 6792–6801.

[69] Barret Zoph and Quoc V Le. 2017. Neural architecture search with reinforcement learning. In *Proceedings of the 5th International Conference on Learning Representations*.

A Experiment Details

A.1 Datasets

We analyzed the chemical space coverage under two distinct scenarios to better understand distribution shifts: (i) **Inter-dataset setting**: We selected the test sets from three binary classification datasets, each following the same splitting strategy, to compare the feature distributions across different datasets. (ii) **Intra-dataset setting**: We took ChChMiner as a representative case and compared the chemical feature distributions between its training and test splits. For both scenarios, we computed a series of physicochemical descriptors for each compound, as summarized in Table 4.

Table 4: Molecular descriptors used in this study and their definitions

Descriptor	Definition
MW	Molecular weight of the compound
LogP	Octanol-water partition coefficient (logP), indicating lipophilicity
HBD	Number of hydrogen bond donors
HBA	Number of hydrogen bond acceptors
TPSA	Topological polar surface area, related to drug absorption
RotBonds	Number of rotatable bonds, representing molecular flexibility
AromaticRings	Number of aromatic rings
HeavyAtoms	Number of heavy (non-hydrogen) atoms
FractionCSP3	Fraction of carbon atoms with sp^3 hybridization
RingCount	Total number of rings in the molecule

We applied t-SNE for dimensionality reduction on the full set of computed molecular descriptors. The resulting two-dimensional embeddings are illustrated in Figure 3. In the inter-dataset setting, the ChChMiner samples are distributed distinctly from those of ZhangDDI and DeepDDI, whereas the latter two exhibit a more similar distribution in the chemical space. In contrast, the intra-dataset setting clearly demonstrates the distributional shift between the training and test sets of ChChMiner. This validates the effectiveness of our out-of-distribution (OOD) dataset splitting strategy.

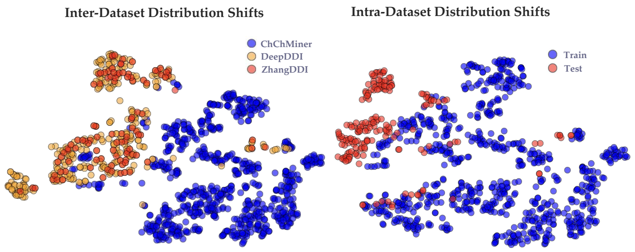
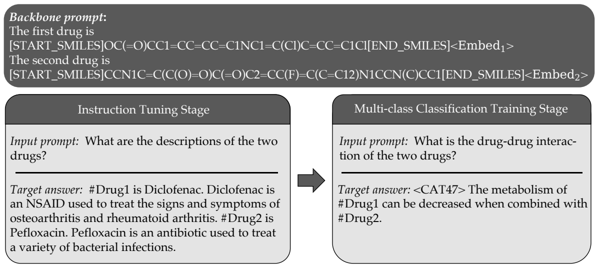


Figure 3: t-SNE visualization of molecular features derived from physicochemical descriptors across multiple datasets. This figure illustrates both inter-dataset and intra-dataset distribution shifts.

A.2 Prompt Example

To better illustrate the textual content involved in our training process, including prompts and instructions, and to highlight the potential clinical interpretability, we provide a representative example from a multi-class classification task used during training, as shown in Figure 4.



***Backbone prompt*:**
The first drug is
[START_SMILES]OC(=O)CC1=CC=CC=C1NC1=C(Cl)C=CC=C1Cl[END_SMILES]<Embed$_1$>
The second drug is
[START_SMILES]CCN1C=C(C(O)=O)C(=O)C2=CC(F)=C(C=C12)N1CCN(C)CC1[END_SMILES]<Embed$_2$>

Instruction Tuning Stage

Input prompt: What are the descriptions of the two drugs?

Target answer: #Drug1 is Diclofenac. Diclofenac is an NSAID used to treat the signs and symptoms of osteoarthritis and rheumatoid arthritis. #Drug2 is Pefloxacin. Pefloxacin is an antibiotic used to treat a variety of bacterial infections.

Multi-class Classification Training Stage

Input prompt: What is the drug-drug interaction of the two drugs?

Target answer: <CAT47> The metabolism of #Drug1 can be decreased when combined with #Drug2.

Figure 4: One qualitative prompt example.

A.3 Retrieval Implementation

In our implementation, we employ a curated retrieval strategy to ensure the quality and consistency of drug-related knowledge used during instruction tuning and inference. Specifically, we construct a drug-description database through the following steps:

- **Metadata Collection**: We crawl drug metadata from *DrugBank*, focusing on entries under the *IDENTIFICATION* section. The *Summary* field is selected as the primary textual description for each drug due to its concise and informative nature.
- **Manual Verification**: For a small subset of drugs without well-structured metadata, we manually retrieve and verify descriptions from reputable biomedical sources to maintain information quality.
- **Database Construction**: All processed drug descriptions are compiled into a local database, indexed by *DrugBank IDs*. This database is used during both training and inference for consistent retrieval of drug information.

Because our knowledge source is strictly confined to *DrugBank* and manually verified texts, the resulting retrieval process is highly reliable. Moreover, indexing by DrugBank ID ensures consistency across all mentions of the same drug. The final retrieval database occupies **25.4 MB** of disk space.

B More Results

B.1 Hyperparameter Sensitivity

We conducted a comprehensive analysis to investigate how several key hyperparameters influence the final performance of our dynamic NAS model. The first three groups of experiments were conducted on the *ChChMiner* dataset, while the last group was performed on the *DrugBank* dataset. The hyperparameters studied include:

- **Number of operational prototypes per layer**: values tested were 2, 4, and 6.
- **Dispersion loss factor** α (in $\mathcal{L}_{\text{disp}}$): values tested were 0.0, 0.001, 0.005, 0.01, and 0.1.
- **Self-supervised loss factor** β (in $\mathcal{L}_{\text{ssl}}$): values tested were 0.0, 0.001, 0.005, 0.01, and 0.1.
- **Length of retrieval prompt**: values tested were 9, 12, and 15 words.

For the retrieval prompt length, we also aimed to examine the effect of different prompt styles on model performance. Specifically, we designed the following variations:

- **P1**: *What is the drug-drug interaction of the two drugs?*
- **P2**: *Classify the interaction between the two drugs by its primary pharmacological mechanism.*
- **P3**: *As a pharmacovigilance officer, how would you classify the interaction between the two drugs?*

The results are shown in Figure 5. From the experimental results, we observed clear and consistent trends regarding the impact of each hyperparameter on the model's predictive performance. First, increasing the number of prototypes per layer from 2 to 6 led to a steady improvement in performance. This suggests that a richer prototype space enhances the model's representational capacity. The effect of the $\mathcal{L}_{disp}$ factor α demonstrated a sweet spot around $\alpha = 0.005$, where the model achieved its highest accuracy and near-peak AUC-ROC. Smaller values still provided substantial gains, while excessive regularization degraded performance. Meanwhile, tuning the $\mathcal{L}_{ssl}$ factor β revealed a similar trend: performance peaked at $\beta = 0.005$, beyond which both accuracy and AUC-ROC declined. This suggests that self-supervised contrastive regularization is effective in improving generalization when balanced appropriately but may introduce noise when overemphasized. In our main experiments, we adopted the following hyperparameter configuration: the number of prototypes per layer was set to 6, and both the α and β factors were set to 0.005.

Finally, we analyzed the effect of prompt length on model performance. Increasing the prompt length from 9 to 12 words led to a noticeable improvement in both accuracy and AUC-ROC, indicating that moderately enriched instructions provide more informative guidance to the model. However, extending the prompt further to 15 words resulted in only marginal gains in AUC-ROC and a slight decrease in accuracy, suggesting diminishing returns and potential semantic redundancy. Despite these variations, the model maintains consistently strong performance across different instruction lengths, demonstrating its robustness and ability to generalize well across stylistic variations in prompts.

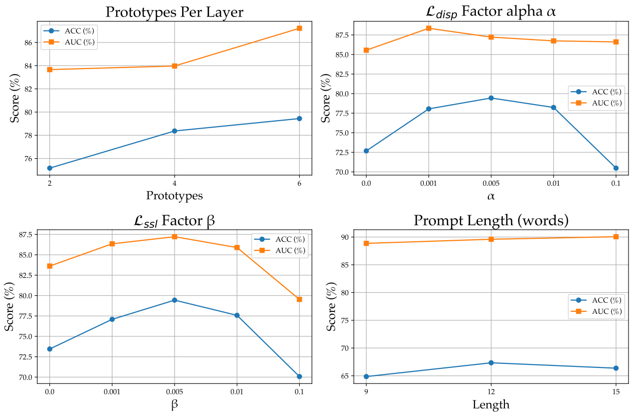


Figure 5: Hyperparameter sensitivity analysis.

B.2 Failure Cases

Despite the overall strong performance of our model, a detailed error analysis reveals several areas where prediction remains challenging. Understanding these failure cases is crucial for guiding future improvements and addressing model limitations under real-world constraints. We categorize the primary types of failure into two broad groups:

Extremely Rare Categories. Certain interaction types, such as `<CAT18>` (*#Drug1 can cause an increase in the absorption of #Drug2 resulting in an increased serum concentration and potentially a worsening of adverse effects*) and `<CAT02>` (*#Drug1 may increase the anticholinergic activities of #Drug2*), each occur only once in the test set, with frequencies below 0.1%. These ultra-rare categories pose significant challenges in a multi-class classification setting. In contrast, slightly more frequent types such as `<CAT22>` and `<CAT26>`, which appear twice in the test set, still achieve a 50% prediction accuracy.

Extreme Molecular Heterogeneity. In our out-of-distribution splitting scenario, predictions failed for drug pairs such as `[Cl-].[Cl-].[Ca++]` and `[Na+].OP(O)([O-])=O`. These compounds exhibit highly atypical molecular structures, lacking well-defined scaffolds and diverging significantly from the predominantly organic compounds seen during training.

These failure cases underscore the challenges of learning under extremely imbalanced class distributions and the limitations of motif-based representations in capturing highly heterogeneous molecular forms.

B.3 Time Cost

To analyze the time cost of our method compared with other LLM-based approaches, we conduct experiments on ChChMiner and DrugBank under the scaffold split setting. For training, we report the average runtime per epoch. For inference time, we measure the total time required to complete inference on the entire test set after training. As shown in Figure 6, our method exhibits approximately a 10% increase in training time cost. This additional cost is accompanied by a notable gain in accuracy, suggesting that the extra computational investment contributes meaningfully to model performance. In terms of inference, our method achieves comparable latency to mainstream baselines, demonstrating that the architectural enhancements do not introduce runtime overhead during deployment.

B.4 Memory Consumption

We analyzed the peak memory usage of our proposed method in comparison to several representative LLM-based approaches under a single-GPU setting. All experiments were conducted with a fixed batch size of 4 on the DrugBank dataset.

The memory consumption of each model during both fine-tuning and inference stages is shown in Table 5. The results show that **DyNAS-DDI** supports full training and inference on a single 24GB GPU with memory consumption comparable to other models. Specifically, during fine-tuning, **DyNAS-DDI** consumes around 22.9–23.4GB, similar to Galactica and MolTC, indicating efficient training even with the added NAS module. For inference, **DyNAS-DDI** uses 7.34GB, slightly higher than Galactica and MolTC but still well within practical limits. This demonstrates that **DyNAS-DDI** balances performance and efficiency, making it suitable for deployment on standard hardware.

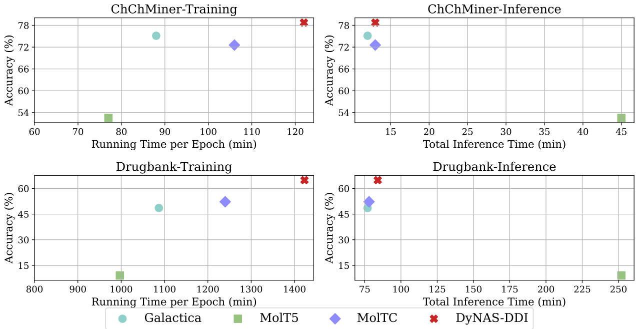


Figure 6: Experimental time cost.

Table 5: Peak memory usage (MB) of different models during finetuning and inference.

Model	Finetuning (MB)	Inference (MB)
Galactica	22828	6866
MolT5	13094	1418
MolTC	23824	6608
DyNAS-DDI (R-IT)	22992	
DyNAS-DDI (Train)	**23386**	**7340**
DyNAS-DDI (w/o NAS)	22926	

C Discussion

Our proposed disentangle–search–augment framework is inherently domain-agnostic and demonstrates strong potential for extension to multimodal retrieval tasks, such as image-text or video-text retrieval. Specifically, its three core components exhibit cross-domain adaptability: (i) The *Motif-driven disentangled molecule encoding mechanism* can be generalized to visual modalities, such as scene graphs or region proposals, where disentangling semantically meaningful patterns from noisy or entangled visual/textual features is equally crucial. This allows the model to isolate informative substructures in a way analogous to molecular motif extraction. (ii) The *Attention-based pairwise neural architecture search* serves as a flexible optimization paradigm for discovering dynamic reasoning pathways. In multimedia settings, it can be employed to adaptively search for optimal cross-modal fusion strategies, which is particularly advantageous in tasks such as image-text retrieval or video question answering, where static fusion mechanisms often fall short in performance. (iii) The *Retrieval-augmented molecular instruction tuning* component offers a general strategy for integrating external knowledge into the model's reasoning process. In multimodal contexts, retrieval prompts can be derived from structured metadata, such as image tags, scene descriptions, or video transcripts, to guide cross-modal alignment and enhance semantic understanding.