# LLM-Integrated Representative Path Selection for Context-Aware Drug Repurposing on Biomedical Knowledge Graphs

**Haerin Song**[1] **Dongmin Bang**[2,3] **Bonil Koo**[2] **Sun Kim**[1,2,3,5] **Sangseon Lee**[4*]
[1] Interdisciplinary Program in Artificial Intelligence, Seoul National University
[2] Interdisciplinary Program in Bioinformatics, Seoul National University
[3] AIGENDRUG Co., Ltd.
[4] Department of Artificial Intelligence, Inha University
[5] Department of Computer Science and Engineering, Seoul National University
{haerinsong, eugenomics, bikoo95, sunkim.bioinfo}@snu.ac.kr, ss.lee@inha.ac.kr

## Abstract

Drug repurposing, aiming to identify novel drug-disease associations by integrating biomedical knowledge, is hindered by challenges in modeling complex multi-hop relationships in knowledge graphs. We propose **DrugCORpath**, a novel approach integrating biomedical knowledge graphs with pretrained biomedical large language models (LLMs). Unlike traditional methods that learn isolated node representations, LLM integration allows capturing of biological contexts by embedding path sentences derived from multi-hop paths connecting drug and disease entities. Each path is converted into biological path sentences reflecting plausible mechanism of action and then embedded to capture rich semantic relationships among entities. We further employ a data-driven selective filtering using K-Means clustering and distance metric to identify meaningful paths, while eliminating redundancy and noise. Experiments on datasets show our method outperforming existing graph-based, LLM-based, and path-based baselines in drug repurposing, achieving up to 4.9% higher accuracy than the prior SOTA. Further analysis confirms that selective path filtering reduces noise and enhances biological diversity. Case studies validate the clinical relevance of the selected paths, ensuring improved interpretability. This collectively underscores the method's potential for interpretable, biologically plausible drug repurposing applications.

## 1 Introduction

Drug repurposing, a central challenge in systems pharmacology and drug discovery, is the task of identifying new uses for existing drugs to treat a given disease [1]. A key to solving this problem lies in understanding the underlying biological mechanisms that bridges diseases to potential treatments. Many diseases involve complex molecular disruptions, such as dysregulated gene expression or dysfunctional signaling pathways, which are in turn targeted by drugs through their mechanisms of action (MoA) [2]. Thus, accurately capturing these intermediate biological processes is crucial for generating plausible and interpretable drug repurposing predictions.

Biomedical knowledge graphs (BKGs) have emerged as a promising resource for modeling such mechanisms, as they structurally integrate diverse biomedical entities, such as genes, diseases, and drugs, and their relationships including gene–gene interactions, drug–gene targeting, and disease–gene

*Corresponding author.

The NeurIPS 2025 Workshop on AI Virtual Cells and Instruments: A New Era in Drug Discovery and Development (AI4D3 2025), San Diego, California, USA, 2025.

associations. These heterogeneous graphs allow for multi-hop reasoning that reflects how a disease may influence certain genes and their network modules, which are in turn modulated by drugs [3, 4].

However, many existing knowledge graph (KG)-based approaches rely on general-purpose graph embeddings or path-counting strategies that fail to reflect rich biological semantics [5]. Most notably, these methods focus on learning node-level embeddings for drugs and diseases and often overlook the gene-level context such as the MoAs of drugs or the molecular mechanisms of diseases, which are critical for accurate repurposing predictions [6]. To address this limitation, it is essential to incorporate gene context information in a biologically coherent manner during representation learning. This enables models to capture biologically meaningful interactions beyond structural proximity. Traditional embeddings such as transition-based or graph neural networks (GNNs)-based models emphasize connectivity but ignore the meaning, directionality, and function of intermediate biomedical relations [7]. As a result, these models often lack mechanistic interpretability and fail to explain why predicted drug–disease associations are biologically relevant [8]. Recent studies emphasize that integrating domain-specific and gene-level context is crucial for producing explainable and clinically meaningful drug repurposing predictions [6].

Second, BKGs are large, dense and heterogeneous, resulting in a combinatorial explosion of multi-hop paths, many of which are redundant or irrelevant [9]. For example, if multiple genes in the same pathway connect a drug-disease pair, the KG may yield multiple biologically equivalent paths [10]. This redundancy can bias analyses toward well-studied mechanisms, thereby overlooking biological diversity and potentially obscuring novel, meaningful insights. Studies such as Project Rephetio [11] further reported that only a small fraction of meta-paths were uniquely predictive, highlighting the need to prune extraneous paths through filtering for better performance and interpretability.

To address these challenges, we propose DrugCORpath: COntext-aware Reasoning path, a framework that combines KG traversal with a large language model (LLM)-based semantic encoding and selective path filtering. We first extract multi-hop disease–gene–drug paths likely to reflect plausible MoAs and convert them into natural language sentences. These sentences are embedded using a biomedical LLM [12] allowing us to capture rich semantic context that conventional graph embeddings miss.

The following step, a key innovation in DrugCORpath, is the data-driven filtering of the overly sampled paths using K-Means clustering in the embedding space, reducing redundancy by selecting diverse, biologically meaningful representatives. Specifically, instead of treating redundant paths equally, our model focuses on distinct mechanisms, improving both accuracy and interpretability. By addressing the two challenges through LLM integration and path selection, our framework is designed to support drug repurposing accompanied by interpretable, pathway-based explanations.

We evaluate our framework on drug repurposing tasks involving unseen diseases using the MSI and PrimeKG datasets to assess whether incorporating LLM-derived semantic context from the KG improves prediction performance. We conducted ablation studies and case analyses and saw that the path selection mechanism effectively filtered out noisy or redundant paths while highlighting clinically plausible mechanisms. We also observed that enforcing a fixed meta-path pattern reduced semantic drift and enabled compact yet informative path generation, leading to more efficient and interpretable predictions.

**Our key contributions include:**

1. DrugCORpath is a KG-LLM integrative framework that outperforms 16 baselines on multiple KGs, achieving up to 4.9% improvement in accuracy from the prior SOTA.

2. LLM-generated sentence-level path embeddings effectively capture biologically plausible and interpretable MoA contexts.

3. Our novel clustering-based path selection effectively filters out noise, yielding concise, diverse, and mechanism-specific path representations that support interpretable drug repurposing.

4. Case studies, including FDA-approved treatments, validate DrugCORpath's ability to identify clinically meaningful rationales beyond mere structural similarity.

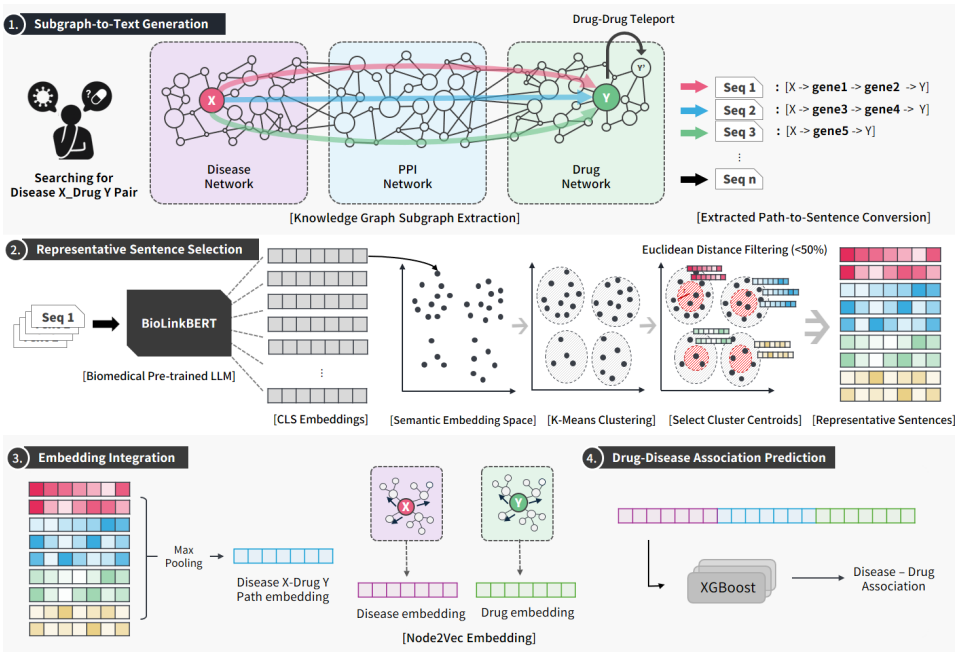



Figure 1: **Overview of DrugCORpath framework.** 1) Given a disease–drug pair, the model explores a biomedical knowledge graph to extract context paths, which are converted into sentences. 2) These sentences are encoded using a biomedical language model, then K-means clustering selects representative sentences to capture diverse mechanisms while reducing redundancy. 3) These embeddings are aggregated via max pooling to form a unified context embedding, which is then combined with Node2vec embeddings of the disease and drug. 4) The final representation is fed into an XGBoost classifier to predict the association.

## 2 Related Work

**Knowledge Graph-based Drug Repurposing** As BKG have become foundational in drug repurposing systems, traditional graph learning methods such as GNNs [13, 14, 15] or triplet-based methods [16, 17] are widely used for learning a structured representation of biomedical entities and their interactions. However, these methods predominantly rely on structural proximity and ignore characteristics of different biomedical entities and relations [18].

**Path-based Reasoning** Path-based models enhance explainability in KG reasoning by modeling chains of biomedical relations. DrugRep-KG [19] improves performance under label imbalance by using informative negative links. DREAMwalk [6] generates drug–gene–disease paths via semantic random walks. Select and Augment [20] boosts semantic path relevance by aligning KG embeddings with external textual descriptions. However, many methods still rely on shallow or random path sampling, which often yields noisy or redundant paths. Without adequate filtering, such noise can obscure mechanisms and hurt generalization to unseen diseases [21].

**LLM-based Biomedical Reasoning** LLMs pretrained on biomedical corpora, such as BioBERT [22], PubMedBERT [23], and SapBERT [24], have recently shown promise in clinical and molecular reasoning tasks. These models capture semantic dependencies beyond graph structure and support natural language inference [25]. While these approaches enable flexible reasoning across biomedical contexts, the lack of structured knowledge-based constraints can lead to hallucinated associations [26], resulting in predictions often lacking interpretability [27].

## 3 Methodology

In this section, we introduce DrugCORpath, a disease–drug association prediction framework. As shown in Figure 1, for a given disease–drug pair, the model explores a BKG to extract *disease-gene-*

*drug* meta-paths, converts them into sentences, and encodes them using a biomedical language model. Representative sentences are selected via K-Means clustering to capture diverse mechanisms and reduce redundancy. These are aggregated through max pooling, combined with network embeddings of the disease and drug, and used as input to a classifier model for prediction.

**Problem Formulation** We predict associations between diseases and drugs on a heterogeneous KG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set is defined as $\mathcal{V} = \mathcal{V}_{\text{drug}} \cup \mathcal{V}_{\text{disease}} \cup \mathcal{V}_{\text{gene}}$, and $\mathcal{E}$ represents biological relations such as gene–gene interactions, drug–gene associations, and disease–gene associations. Notably, $\mathcal{G}$ contains no direct edges between disease and drug nodes. Given a disease node $s \in \mathcal{V}_{\text{disease}}$ and a drug node $d \in \mathcal{V}_{\text{drug}}$, the task is to predict whether an association exists between them based on their context in $\mathcal{G}$. Formally, we aim to learn a function $f : (\mathcal{G}, s, d) \mapsto \hat{y} \in \{0, 1\}$, where $\hat{y}$ indicates the predicted presence or absence of an association. The model is trained to minimize the discrepancy between $\hat{y}$ and the true label $y$.

### 3.1 Disease-Drug Context Generation via Knowledge Graph Exploration

To capture the mechanistic context between a disease and a drug, we perform path extraction from a BKG $\mathcal{G}$ using a constrained meta-path pattern. Specifically, we focus on paths where a disease is connected to a drug through a sequence of gene nodes, reflecting plausible MoAs such as "a disease is associated with dysregulated genes, which are often targeted by drugs as part of their therapeutic mechanism." Formally, for each disease $s \in \mathcal{V}_{\text{disease}}$ and drug $d \in \mathcal{V}_{\text{drug}}$, we define the set of candidate paths as all possible cases of:

$$P_{s,d} = \{p \mid p = (s \to g_1 \to \cdots \to g_m \to d), \\ g_i \in \mathcal{V}_{\text{gene}},\ 1 \le m \le k_{\max}\}.$$

To ensure interpretability and mitigate noise from over-sampled paths, we limit the maximum number of gene nodes $k_{\max}$. Although adding gene nodes enriches context, the dense connectivity of genes can introduce redundant or irrelevant paths that obscure meaningful associations.

However, this constraint may result in the absence of valid paths between certain disease–drug pairs. To address this, inspired by DREAMwalk [6], we adopt a drug teleportation strategy based on the Anatomical Therapeutic Chemical (ATC) classification system, which groups drugs by their therapeutic function and MoA. Specifically, when no path exists between a given disease $s$ and drug $d$, we first sample a pharmacologically similar substitute drug $d'$ from the ATC similarity network of $d$, and construct an alternative meta-path of the form $s \to g_1 \to \cdots \to g_k \to d'$. Lastly, the path is linked to the original target drug $d$.

The probability of teleporting from $d$ to a candidate $d'$ is defined as:

$$p(d' \mid d, \mathbf{A}) = \frac{\mathbf{A}_{d,d'}}{\sum_{k \in \mathcal{N}_d} \mathbf{A}_{d,k}}. \tag{1}$$

Here, $\mathbf{A}$ is the ATC similarity matrix, where $\mathbf{A}_{d,d'}$ denotes the ATC similarity between the original drug $d$ and the substitute candidate $d'$ (detailed in Appendix A.3). $\mathcal{N}_d$ is the set of ATC-neighboring drugs of $d$, and the similarity scores $\mathbf{A}_{d,k}$ are precomputed based on their ATC codes. This teleportation mechanism is applied only once when no direct path exists between the disease $s$ and drug $d$, and is used solely to select a substitute drug $d'$ based on similarity.

Once the final path set $P_{s,d}$ is determined through direct extraction or ATC-based substitution, each path $p \in P_{s,d}$ is transformed into a biomedical sentence using directional prompts (e.g., "Disease A -> Gene B -> Gene C -> Drug D"). These textual representations are designed to capture the biological flow among entities and serve as context-rich proxies for potential MoAs.

### 3.2 Contextual Embedding and Selection of Disease-Drug Paths

The disease–drug contexts extracted from the KG take the form of sequential paths that pass through intermediate gene nodes. To effectively capture the semantic and biological information embedded in these sequences, we leverage LLMs, which are well-suited for modeling complex dependencies in textual input. To encode each path into a dense embedding, we utilize domain-specific pretrained LLM, BioLinkBERT [12], which is well-suited for biomedical data. We hypothesize that such models can better capture the mechanistic relationships among diseases, genes, and drugs. Importantly, we

Table 1: Performance comparison of DrugCORpath and 16 baselines across four model categories (Graph-based, Transition-based, LLM-based, and Path-based) on the MSI and PrimeKG datasets. Mean and standard deviation of 5 fold CV are provided.

| Methods | | MSI | | | PrimeKG | | |
|---|---|---|---|---|---|---|---|
| | | AUROC | AUPRC | Accuracy | AUROC | AUPRC | Accuracy |
| Graph-based | GCN | 0.619±0.039 | 0.591±0.028 | 0.598±0.037 | 0.922±0.006 | 0.893±0.009 | 0.844±0.012 |
| | GraphSAGE | 0.580±0.019 | 0.574±0.016 | 0.566±0.013 | 0.901±0.013 | 0.853±0.023 | 0.813±0.024 |
| | GIN | 0.598±0.040 | 0.651±0.017 | 0.654±0.022 | 0.835±0.048 | 0.789±0.074 | 0.776±0.042 |
| | GAT | 0.567±0.031 | 0.588±0.032 | 0.558±0.031 | 0.940±0.005 | 0.908±0.006 | 0.852±0.010 |
| Transition-based | TransE | 0.539±0.024 | 0.534±0.013 | 0.522±0.027 | 0.707±0.021 | 0.649±0.045 | 0.610±0.010 |
| | TransR | 0.571±0.017 | 0.579±0.019 | 0.547±0.015 | 0.616±0.026 | 0.547±0.047 | 0.591±0.034 |
| | RotatE | 0.546±0.010 | 0.559±0.017 | 0.521±0.010 | 0.787±0.016 | 0.742±0.013 | 0.660±0.017 |
| | ComplEx | 0.555±0.018 | 0.566±0.018 | 0.507±0.024 | 0.793±0.023 | 0.782±0.021 | 0.604±0.017 |
| | RESCAL | 0.500±0.020 | 0.501±0.030 | 0.504±0.018 | 0.625±0.035 | 0.558±0.036 | 0.573±0.021 |
| LLM-based | BioBERT | 0.744±0.020 | 0.753±0.008 | 0.664±0.030 | 0.931±0.011 | 0.902±0.017 | 0.785±0.029 |
| | BioLinkBERT | 0.696±0.018 | 0.705±0.009 | 0.630±0.024 | 0.921±0.007 | 0.888±0.018 | 0.764±0.022 |
| | PubMedBERT | 0.818±0.016 | 0.823±0.011 | 0.720±0.026 | 0.939±0.008 | 0.915±0.015 | 0.802±0.020 |
| | SapBERT | 0.798±0.019 | 0.812±0.014 | 0.711±0.029 | 0.926±0.010 | 0.902±0.019 | 0.774±0.019 |
| Path-based | Node2vec | 0.752±0.023 | 0.766±0.018 | 0.664±0.024 | 0.942±0.014 | 0.931±0.013 | 0.832±0.020 |
| | Drugrep-KG | 0.733±0.027 | 0.740±0.013 | 0.643±0.029 | 0.682±0.036 | 0.778±0.020 | 0.563±0.049 |
| | DREAMwalk | 0.763±0.008 | 0.769±0.007 | 0.663±0.112 | 0.941±0.011 | 0.929±0.014 | 0.838±0.024 |
| | DrugCORpath | **0.842±0.005** | **0.845±0.009** | **0.755±0.007** | **0.957±0.005** | **0.946±0.007** | **0.867±0.016** |

used a frozen version of the encoder without any task-specific fine-tuning, relying solely on the [CLS] token to extract contextual representations. The [CLS] token is a special classification token pre-pended to the input sequence in BERT, whose final hidden state is typically used to represent the entire sequence for classification tasks.

For each sentence $x_i$ derived from a path $p_i \in P_{s,d}$, we obtain its embedding by extracting the [CLS] token embedding from the final layer of the encoder: $\mathbf{h}_i = \phi(x_i) \in \mathbb{R}^{\text{dim}}$ where $\phi(\cdot)$ denotes the pretrained sentence encoder. The set of embeddings for all paths associated with a disease–drug pair $(s, d)$ is denoted as $\mathbf{H}_{s,d} = \{\mathbf{h}_1, \ldots, \mathbf{h}_{|P_{s,d}|}\}$.

Since a single disease–drug relationship can involve multiple distinct MoAs, and the KG exploration process often generates semantically similar or duplicate paths for the same pair, semantic redundancy is likely to arise. To address this, we apply clustering to identify representative and non-redundant contexts. Specifically, we perform K-Means on $\mathbf{H}_{s,d}$ to obtain $K$ cluster centroids $\mathbf{c}_1, \ldots, \mathbf{c}_K$. For each cluster $k$, we select embeddings that lie within the top $\rho\%$ closest to the centroid:

$$\mathbf{h}_{k,j}^{(\text{rep})} \in \{\mathbf{h}_i \in \text{cluster}_k \mid \|\mathbf{h}_i - \mathbf{c}_k\|_2 \leq \tau_k\}, \tag{2}$$

$$\tau_k = \text{Percentile}_\rho(\|\mathbf{h}_i - \mathbf{c}_k\|_2). \tag{3}$$

The representative set is then defined as

$$\mathbf{H}_{s,d}^{(\text{rep})} = \bigcup_{k=1}^{K} \left\{\mathbf{h}_{k,j}^{(\text{rep})}\right\}. \tag{4}$$

To construct a unified vector representation of the disease–drug context, we apply a max pooling operation over the representative set:

$$\mathbf{h}_{\text{Rep}} = \text{MaxPool}(\mathbf{H}_{s,d}^{(\text{rep})}) \in \mathbb{R}^{\text{dim}}. \tag{5}$$

Although more elaborate aggregation strategies could be considered, we found max pooling to be empirically effective and computationally efficient for capturing salient semantic features across diverse contexts.

### 3.3 Disease-Drug Association Prediction

To incorporate the topological characteristics of diseases and drugs within the BKG, we use node2vec [28] to encode each node based on the global structure of the KG $\mathcal{G}$. Node2vec performs biased

random walks to capture both local and global structural proximities. We denote the resulting embeddings as $\mathbf{z}_s = \text{Node2vec}(s)$ and $\mathbf{z}_d = \text{Node2vec}(d)$ for disease node $s$ and drug node $d$, respectively. To predict the likelihood of an association between $s$ and $d$, we concatenate the node embeddings and the representative context embedding $\mathbf{h}_{\text{Rep}}$ obtained from Equation 5: $\mathbf{z}_{\text{concat}} = [\mathbf{z}_s; \mathbf{h}_{\text{Rep}}; \mathbf{z}_d]$.

We adopt a stacking ensemble model for the association prediction task, leveraging the diversity of multiple gradient-boosted decision trees. We use three independently parameterized XGBoost classifiers as base learners and a Logistic Regression model as the meta-learner. This ensemble strategy enhances robustness and generalization by combining distinct decision boundaries learned from different base models. The final input feature vector $\mathbf{z}_{\text{concat}}$ is constructed by concatenating contextual and topological embeddings, and is passed through the ensemble model to output the predicted probability: $\hat{y}_{s,d} = f_{\text{stack}}(\mathbf{z}_{\text{concat}}) \in [0, 1]$, where $f_{\text{stack}}$ denotes the stacking classifier.

We train the model using the binary cross-entropy loss:

$$\mathcal{L} = -\left[y_{s,d} \log \hat{y}_{s,d} + (1 - y_{s,d}) \log(1 - \hat{y}_{s,d})\right], \tag{6}$$

where $y_{s,d}$ denotes the ground-truth label. To address the inherent class imbalance in disease–drug associations, we employ negative sampling by pairing diseases with randomly selected drugs that are not known to be associated. This strategy balances the number of positive and negative instances during training and improves model robustness.

# 4 Results

## 4.1 Dataset and Evaluation Protocol

We conducted experiments on two BKGs: MSI [29] and PrimeKG [30]. To evaluate the model's capability of predicting the drug-disease associations from zero-shot representations, all drug–disease edges were removed from each KG prior to training (Appendix A.2).

The task was formulated as binary classification, using known disease-drug associations as positives and randomly sampled unconnected pairs as negatives (1:1 ratio for MSI; 1.2:1 for PrimeKG). To evaluate generalization to unseen diseases, we split the data so that each disease appears exclusively in either the training or test set with no overlap. We performed a 5-fold cross-validation and evaluated performance using AUROC, AUPRC and Accuracy.

### 4.1.1 Baselines

To benchmark our proposed method, we compared it against four categories of baseline models: three among GNN-based models, five among KG embedding models, four pretrained biomedical language models and three path-based models (detailed in Appendix A.1).

## 4.2 Performance Comparison

Table 1 summarizes the prediction performance of DrugCORpath and all baselines on the MSI and PrimeKG datasets. DrugCORpath consistently achieves the best results across AUROC, AUPRC and Accuracy on both datasets, with low deviations indicating stable performance.

On the MSI dataset, it improves AUROC by 2.9%, AUPRC by 2.7%, and Accuracy by 4.9% over the strongest LLM-based baseline, PubMedBERT. On the PrimeKG dataset, DrugCORpath outperforms path-based models such as Node2vec and DREAMwalk, with gains of up to 1.7% in AUROC, 1.8% in AUPRC, and 4.2% in Accuracy.

Interestingly, DrugCORpath outperforms the baselines relying solely on structural embeddings or pretrained language models. These results demonstrate that leveraging informative biomedical contexts extracted from the KG is important for robust drug–disease association prediction. In particular, the consistently low standard deviations observed across folds suggest that the clustering-based selection of representative paths plays a key role in filtering noisy or redundant contexts, thereby stabilizing model performance while preserving mechanistic relevance.

### 4.3 Ablation Study

To demonstrate the effectiveness of DrugCORpath's path selection, we conduct ablation studies on: (1) meta-path pattern used for KG exploration and (2) clustering-based sentence filtering mechanism.

#### 4.3.1 Effect of Meta-Path Patterns

We evaluate the impact of the meta-path pattern of DrugCORpath through: (1) increasing the maximum number of intermediate gene nodes $k_{\max}$ from 2 to 3, (2) performing an additional round of ATC-based drug teleportation, and (3) randomly sampling intermediate gene nodes without following the KG topology.

Figure 2 shows that DrugCORpath consistently performs better than all ablation variants. Increasing $k_{\max}$ results in a 2.7% AUROC drop, suggesting that including more gene nodes may introduce irrelevant or redundant information. Similarly, increasing the number of ATC-based teleportation leads to a 3.2% drop in AUROC. While this may increase exposure to pharmacologically similar drugs, it reduces the reliability of the generated paths. Randomly sampling intermediate genes causes a 31.1% drop in AUROC, highlighting the importance of incorporating MoA-relevant gene paths when modeling disease–drug relationships. Complete performance results are reported in Appendix A.6.

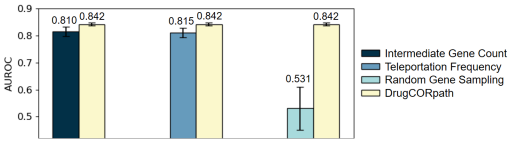



Figure 2: AUROC of meta-path ablation strategies on MSI. Each group compares DrugCORpath with a single-component ablation; values with error bars are shown above bars.

These results indicate that concise, MoA-consistent paths are more effective than overly complex or loosely structured alternatives for drug–disease reasoning.

#### 4.3.2 Impact of Clustering-based Sentence Selection

We examined how different sentence filtering thresholds ($\rho$ = 20%, 50%, 80%, 100%) affect DrugCORpath's performance by selecting varying proportions of sentences closest to each cluster center. As shown in Table 2, retaining 50% of sentences yields the best overall results, suggesting a strong balance between informative content and redundancy reduction. While 80% performs similarly, it introduces more variability, likely due to added noise. In contrast, keeping too few (20%) or too many (100%) sentences hurts performance—either by discarding useful data or allowing excess noise. These results highlight the value of clustering-based filtering for denoising while preserving key information.

Table 2: Model performance across sentence filtering thresholds ($\rho$). At 0%, no filtering is applied. Best results are bolded.

| $\rho$ | AUROC | AUPRC | Accuracy |
|---|---|---|---|
| 80% | 0.839±0.002 | 0.842±0.008 | 0.750±0.004 |
| 50% | **0.842±0.005** | **0.845±0.009** | **0.755±0.007** |
| 20% | 0.841±0.009 | 0.844±0.014 | 0.751±0.009 |
| 0% | 0.838±0.010 | 0.840±0.011 | 0.749±0.010 |

### 4.4 Mechanism-of-Action Interpretability

#### 4.4.1 GO-based Interpretation of MoA-Representative Path Clusters

To assess whether clustering captures mechanistically distinct biological processes, we analyzed representative gene ontology (GO) terms derived from each cluster for every disease-drug pair. We first performed a systematic analysis on the diversity of the enriched GO terms among each drug-disease pair, measured through GO term similarity (described in Appendix A.4). As shown in Figure 4a, the results showed that most disease–drug pairs were divided into clusters with low GO term similarity, suggesting that DrugCORpath's clustering-based path filtering captures a broad spectrum of distinct MoA for each drug-disease pair.

As a detailed case study, we extracted the representative enriched paths for B-Cell Lymphomas-fostamatinib pair. As illustrated in Figure 3, the resulting clusters revealed GO terms associated with protein modification, intracellular signaling regulation, and canonical NF-$\kappa$B signaling, reflecting fostamatinib's diverse influence on disease-relevant pathways. Further mechanistic interpretation

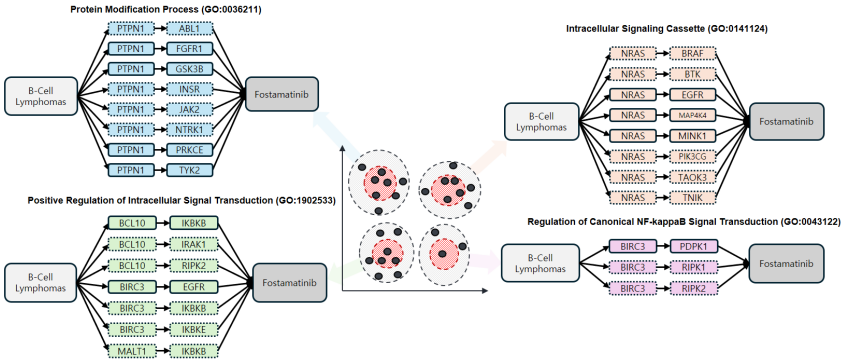
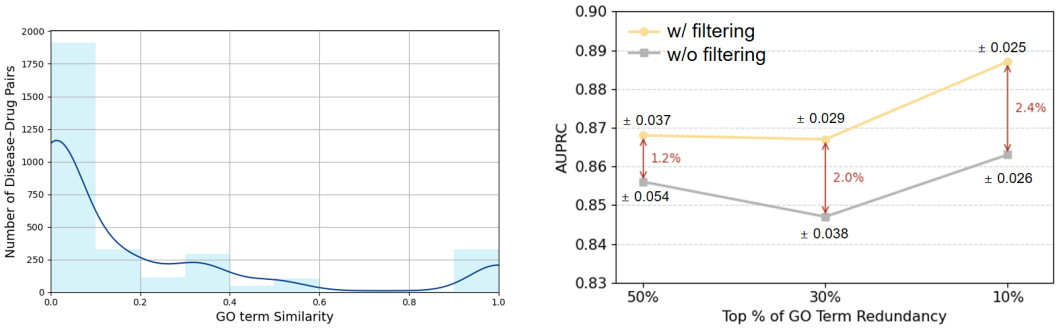



Figure 3: Case study illustrating path clusters identified by DrugCORpath for the B-cell lymphomas–Fostamatinib pair, with each cluster annotated by its most frequently observed GO term representing a mechanism of action (MoA). Shown clusters are enriched for distinct biological processes.




(a) (b)

Figure 4: Comparison of GO-term–based evaluation. (a) Distribution of similarity scores between GO terms assigned to different clusters within each disease–drug pair. (b) AUPRC comparison across GO term redundancy levels using the top disease–drug pairs. Standard deviations are shown in each plot.

highlighted a representative pathway: B-cell lymphomas → BCL10 → IKBKB → Fostamatinib, tracing fostamatinib's upstream interference in NF-$\kappa$B activation via BCR signaling blockade [31] (further detailed in Appendix A.4.2).

Collectively, these findings underscore DrugCORpath's power in capturing compact yet diverse and relevant mechanistic insights for drug-disease relationships.

### 4.4.2 Effect of GO-Based Redundant Path Filtering

To evaluate the impact of redundant paths, we focused on the top 10%, 30%, and 50% of drug-disease pairs exhibiting the highest functional redundancy, as measured by GO term overlap among path genes using the GO hierarchy. Figure 4b shows that the filtering strategy (yellow line) consistently outperforms the unfiltered setting (grey line) across all levels of redundancy. Notably, the performance gap widens as GO term redundancy increases, indicating that filtering becomes increasingly important for retaining biologically meaningful paths under high semantic redundancy.

### 4.4.3 Case studies: Blood Coagulation and Atopic Dermatitis

To evaluate the biological interpretability of DrugCORpath's predictions, we analyzed specific MoA paths identified for selected disease–drug pairs. In blood coagulation disorders case, DrugCORpath recommended Benzbromarone, a gout medication, as a candidate treatment via identified paths involving NRAS/KRAS and ABCC1 (Figure 5). Literature-supported roles for these genes in prothrombotic signaling and metabolite transport suggest a plausible route for modulating coagulation

Table 3: Predicted probabilities of drug–disease associations for newly approved indications (2024–2025) using DrugCORpath, BioLinkBERT, and Node2vec. Highest values for each case are marked in bold.

| New Approved Indication | Drug | DrugCORpath | BioLinkBERT | Node2vec |
|---|---|---|---|---|
| Pediatric Hemophilia B | Coagulation Factor IX | **0.903** | 0.411 | 0.799 |
| Diabetic Retinopathy | Ranibizumab | **0.903** | 0.246 | 0.178 |
| Urothelial Carcinoma | Nivolumab | **0.894** | 0.705 | 0.506 |
| Plaque Psoriasis | Apremilast | **0.860** | 0.649 | 0.762 |
| Bipolar I Disorder | Iloperidone | 0.820 | 0.715 | **0.885** |
| Pediatric HeFH | Alirocumab | **0.812** | 0.712 | 0.772 |
| Food Allergy | Omalizumab | 0.750 | **0.769** | 0.742 |
| Psoriasis | Usteknumab | 0.690 | 0.692 | **0.739** |
| Ph+ ALL | Ponatinib | **0.622** | 0.199 | 0.423 |
| Tonic-Clonic Seizures | Lacosamide | 0.236 | 0.684 | **0.749** |
| Atopic Dermatitis | Roflumilast | 0.194 | **0.657** | 0.259 |
| Severe Asthma | Benralizumab | 0.159 | 0.556 | **0.768** |

through ABCC1 inhibition, underscoring the model's capacity to focus on novel and biologically grounded hypotheses.

In another example, DrugCORpath correctly identified the clinically validated treatment of atopic dermatitis with Prednisolone. The utilized path captured known immunological mechanisms involving IL6-driven activation of NR3C1, a glucocorticoid receptor targeted by Prednisolone. This case illustrates the model's ability to recover established therapeutic mechanisms alongside uncovering novel associations. Full pathway details and supporting evidence for both cases are provided in Appendix A.5.

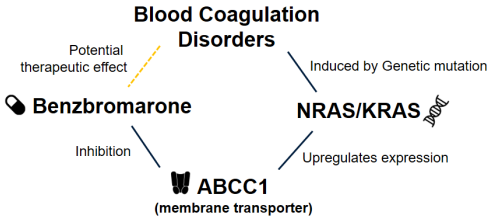



Figure 5: Case study of an interpretable drug–disease path. Edges show biological relations; the yellow dashed line marks the predicted association.

### 4.5 FDA Drug Repurposing Prediction

To assess the model's ability to be applied for drug repurposing, we evaluated recently approved drug–disease pairs from the FDA in 2024–2025[2], focusing on new indications assigned to existing drugs. As shown in Table 3, DrugCORpath not only assigned consistently high probabilities, but also ranked first in the majority of these novel drug–indication associations, outperforming its components, BioLinkBERT and Node2vec. In particular, for Ranibizumab, DrugCORpath predicted a strong association with its newly approved indication (Diabetic Retinopathy), which is quite distinct from its original indication age-related macular degeneration [32], while other models assigned considerably lower probabilities.

These results demonstrate that DrugCORpath can effectively recommend drugs that are later validated with new indications, highlighting its potential to identify clinically relevant drug–disease associations.

## 5 Conclusion

In this paper, we propose DrugCORpath, a novel KG-based drug repurposing method by integrating meta-path-based KG exploration and LLM-based biological context embedding. By leveraging cluster-based path selection, DrugCORpath effectively captures diverse MoA semantics while preserving biologically meaningful paths. Extensive evaluations demonstrate state-of-the-art performance on MSI and PrimeKG datasets under unseen disease split setting, suggesting that focusing on important paths from KG is a promising direction for improving drug repurposing tasks.

[2]Retrieved from https://www.drugs.com/new-indications.html.

## References


[1] Kexin Huang, Payal Chandak, Qianwen Wang, Shreyas Havaldar, Akhil Vaid, Jure Leskovec, Girish N Nadkarni, Benjamin S Glicksberg, Nils Gehlenborg, and Marinka Zitnik. A foundation model for clinician-centered drug repurposing. *Nature Medicine*, 30(12):3601–3613, 2024.

[2] Xiaoqin Pan, Xuan Lin, Dongsheng Cao, Xiangxiang Zeng, Philip S Yu, Lifang He, Ruth Nussinov, and Feixiong Cheng. Deep learning for drug repurposing: Methods, databases, and applications. *Wiley interdisciplinary reviews: Computational molecular science*, 12(4):e1597, 2022.

[3] Daeun Kong, Yoojin Ha, HaEun Yoo, Dongmin Bang, and Sun Kim. Survey on ai-drug discovery with knowledge graphs: Data, algorithm, and application. *Journal of Computing Science and Engineering*, 19(1):1–15, 2025.

[4] Yuxing Lu, Sin Yee Goi, Xukai Zhao, and Jinzhuo Wang. Biomedical knowledge graph: A survey of domains, tasks, and real-world applications. *arXiv preprint arXiv:2501.11632*, 2025.

[5] Daniel N Sosa, Georgiana Neculae, Julien Fauqueur, and Russ B Altman. Elucidating the semantics-topology trade-off for knowledge inference-based pharmacological discovery. *Journal of Biomedical Semantics*, 15(1):5, 2024.

[6] Dongmin Bang, Sangsoo Lim, Sangseon Lee, and Sun Kim. Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. *Nature Communications*, 14(1):3570, 2023.

[7] Zheng Gao, Gang Fu, Chunping Ouyang, Satoshi Tsutsui, Xiaozhong Liu, Jeremy Yang, Christopher Gessner, Brian Foote, David Wild, Ying Ding, et al. edge2vec: Representation learning using edge semantics for biomedical knowledge discovery. *BMC bioinformatics*, 20(1):306, 2019.

[8] Aryo Pradipta Gema, Dominik Grabarczyk, Wolf De Wulf, Piyush Borole, Javier Antonio Alfaro, Pasquale Minervini, Antonio Vergari, and Ajitha Rajan. Knowledge graph embeddings in the biomedical domain: Are they useful? a look at link prediction, rule learning, and downstream polypharmacy tasks. *Bioinformatics Advances*, 4(1):vbae097, 2024.

[9] Yue Hu, Svitlana Oleshko, Samuele Firmani, Zhaocheng Zhu, Hui Cheng, Maria Ulmer, Matthias Arnold, Maria Colomé-Tatché, Jian Tang, Sophie Xhonneux, et al. Path-based reasoning for biomedical knowledge graphs with biopathnet. *bioRxiv*, 2024.

[10] Shunian Xiang, Patrick J Lawrence, Bo Peng, ChienWei Chiang, Dokyoon Kim, Li Shen, and Xia Ning. Modeling path importance for effective alzheimer's disease drug repurposing. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, volume 29, page 306, 2024.

[11] Daniel Scott Himmelstein, Antoine Lizee, Christine Hessler, Leo Brueggeman, Sabrina L Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, and Sergio E Baranzini. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *elife*, 6:e26726, 2017.

[12] Michihiro Yasunaga, Jure Leskovec, and Percy Liang. Linkbert: Pretraining language models with document links. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8003–8016, 2022.

[13] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

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

[15] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.

[16] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.

[17] Maximilian Nickel, Volker Tresp, Hans-Peter Kriegel, et al. A three-way model for collective learning on multi-relational data. In *Icml*, volume 11, pages 3104482–3104584, 2011.

[18] Vikas Garg, Stefanie Jegelka, and Tommi Jaakkola. Generalization and representational limits of graph neural networks. In *International conference on machine learning*, pages 3419–3430. PMLR, 2020.

[19] Zahra Ghorbanali, Fatemeh Zare-Mirakabad, Mohammad Akbari, Najmeh Salehi, and Ali Masoudi-Nejad. Drugrep-kg: Toward learning a unified latent space for drug repurposing using knowledge graphs. *Journal of chemical information and modeling*, 63(8):2532–2545, 2023.

[20] Micheal Abaho and Yousef H Alfaifi. Select and augment: enhanced dense retrieval knowledge graph augmentation. *Journal of Artificial Intelligence Research*, 78:269–285, 2023.

[21] Tassallah Abdullahi, Ioanna Gemou, Nihal V Nayak, Ghulam Murtaza, Stephen H Bach, Carsten Eickhoff, and Ritambhara Singh. K-paths: Reasoning over graph paths for drug repurposing and drug interaction prediction. *arXiv preprint arXiv:2502.13344*, 2025.

[22] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[23] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1):1–23, 2021.

[24] Fangyu Liu, Ehsan Shareghi, Zaiqiao Meng, Marco Basaldella, and Nigel Collier. Self-alignment pretraining for biomedical entity representations. *arXiv preprint arXiv:2010.11784*, 2020.

[25] Karan Singhal, Shekoofeh Azizi, Tao Tu, and S. Sara *et al.*** Mahdavi. Large language models encode clinical knowledge. *Nature*, 620:172–180, 2023.

[26] Robin Wagner, Emanuel Kitzelmann, and Ingo Boersch. Mitigating hallucination by integrating knowledge graphs into llm inference–a systematic literature review. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 795–805, 2025.

[27] Garima Agrawal, Tharindu Kumarage, Zeyad Alghamdi, and Huan Liu. Can knowledge graphs reduce hallucinations in llms?: A survey. *arXiv preprint arXiv:2311.07914*, 2023.

[28] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

[29] Camilo Ruiz, Marinka Zitnik, and Jure Leskovec. Identification of disease treatment mechanisms through the multiscale interactome. *Nature communications*, 12(1):1796, 2021.

[30] Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1):67, 2023.

[31] Min Xia, Liron David, Matthew Teater, Ozlem Onder, Kojo SJ Elenitoba-Johnson, Lorena Fontan, Hao Wu, and Ari Melnick. Bcl10 gain-of-function mutations aberrantly induce canonical and non-canonical nf-kb activation and resistance to ibrutinib in abc-dlbcl. *Blood*, 136:2–3, 2020.

[32] Shahrokh Ramin, Masoud Soheilian, Gholamreza Habibi, Roghayeh Ghazavi, Reza Gharebaghi, and Fatemeh Heidary. Age-related macular degeneration: a scientometric analysis. *Medical Hypothesis, Discovery and Innovation in Ophthalmology*, 4(2):39, 2015.

[33] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

[34] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29, 2015.

[35] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.

[36] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR, 2016.

[37] Catia Pesquita, Daniel Faria, Andre O Falcao, Phillip Lord, and Francisco M Couto. Semantic similarity in biomedical ontologies. *PLoS computational biology*, 5(7):e1000443, 2009.

[38] Jay J Jiang and David W Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of the 10th Research on Computational Linguistics International Conference*, pages 19–33, 1997.

[39] Jonathan W Friedberg, Jeff Sharman, John Sweetenham, Patrick B Johnston, Julie M Vose, Ann LaCasce, Julia Schaefer-Cutillo, Sven De Vos, Rajni Sinha, John P Leonard, et al. Inhibition of syk with fostamatinib disodium has significant clinical activity in non-hodgkin lymphoma and chronic lymphocytic leukemia. *Blood, The Journal of the American Society of Hematology*, 115(13):2578–2585, 2010.

[40] SEM Herman, PM Barr, EM McAuley, D Liu, A Wiestner, and JW Friedberg. Fostamatinib inhibits b-cell receptor signaling, cellular activation and tumor proliferation in patients with relapsed and refractory chronic lymphocytic leukemia. *Leukemia*, 27(8):1769–1773, 2013.

[41] Ulf Klein and Sankar Ghosh. The two faces of nf-$\kappa$b signaling in cancer development and therapy. *Cancer cell*, 20(5):556–558, 2011.

[42] Sarah EM Herman, Paul M Barr, Erin M McAuley, Delong Liu, Jonathan W Friedberg, and Adrian Wiestner. Fostamatinib inhibits bcr signaling, and reduces tumor cell activation and proliferation in patients with relapsed refractory chronic lymphocytic leukemia. *Blood*, 120(21):2882, 2012.

[43] Kelly A Pike and Michel L Tremblay. Tc-ptp and ptp1b: regulating jak–stat signaling, controlling lymphoid malignancies. *Cytokine*, 82:52–57, 2016.

[44] Khine S Shan, Shivani Dalal, Nyein Nyein Thaw Dar, Omani McLish, Matthew Salzberg, and Brian A Pico. Molecular targeting of the fibroblast growth factor receptor pathway across various cancers. *International Journal of Molecular Sciences*, 25(2):849, 2024.

[45] Isabel Heidrich, Charlotte Rautmann, Cedric Ly, Robin Khatri, Julian Kött, Glenn Geidel, Alessandra Rünger, Antje Andreas, Inga Hansen-Abeck, Finn Abeck, et al. In-depth assessment of braf, nras, kras, egfr, and pik3ca mutations on cell-free dna in the blood of melanoma patients receiving immune checkpoint inhibition. *Journal of Experimental & Clinical Cancer Research*, 44(1):202, 2025.

[46] Davide Rossi, Silvia Deaglio, David Dominguez-Sola, Silvia Rasi, Tiziana Vaisitti, Claudio Agostinelli, Valeria Spina, Alessio Bruscaggin, Sara Monti, Michaela Cerri, et al. Alteration of birc3 and multiple other nf-$\kappa$b pathway genes in splenic marginal zone lymphoma. *Blood, The Journal of the American Society of Hematology*, 118(18):4930–4934, 2011.

[47] Beini Sun, Hongce Chen, Xiaoping Wang, and Tongsheng Chen. Regorafenib induces bim-mediated intrinsic apoptosis by blocking akt-mediated foxo3a nuclear export. *Cell Death Discovery*, 9(1):37, 2023.

[48] Meghna Mehta, James Griffith, Janani Panneerselvam, Anish Babu, Jonathan Mani, Terence Herman, Rajagopal Ramesh, and Anupama Munshi. Regorafenib sensitizes human breast cancer cells to radiation by inhibiting multiple kinases and inducing dna damage. *International journal of radiation biology*, 97(8):1109–1120, 2021.

[49] Yongjun Cha, Hwang-Phill Kim, Yoojoo Lim, Sae-Won Han, Sang-Hyun Song, and Tae-You Kim. Fgfr 2 amplification is predictive of sensitivity to regorafenib in gastric and colorectal cancers in vitro. *Molecular oncology*, 12(7):993–1003, 2018.

[50] SM Wilhelm, J Dumas, L Adnane, M Lynch, CA Carter, G Schütz, KH Thierauch, and D Regorafenib Zopf. A new oral multikinase inhibitor of angiogenic, stromal and oncogenic receptor tyrosine kinases with potent preclinical antitumor activity. *Int. J. Cancer*, 129(1):245–255, 2011.

[51] Zhikun Liang, Xiaoyan Huang, Jieling Mao, Jingwen Xie, Xiaoyan Li, and Li Qin. The impact of kras mutations on risk of venous thromboembolism recurrence in patients with metastatic colorectal cancer. *BMC gastroenterology*, 25(1):240, 2025.

[52] Iris C Salaroglio, Eleonora Mungo, Elena Gazzano, Joanna Kopecka, and Chiara Riganti. Erk is a pivotal player of chemo-immune-resistance in cancer. *International journal of molecular sciences*, 20(10):2505, 2019.

[53] Susan PC Cole. Multidrug resistance protein 1 (mrp1, abcc1), a "multitasking" atp-binding cassette (abc) transporter. *Journal of Biological Chemistry*, 289(45):30880–30888, 2014.

[54] Kalam Sirisha, Maddela Chandra Shekhar, Kulandaivelu Umasankar, Porika Mahendar, Abbagani Sadanandam, Garlapati Achaiah, and Vanga Malla Reddy. Molecular docking studies and in vitro screening of new dihydropyridine derivatives as human mrp1 inhibitors. *Bioorganic & medicinal chemistry*, 19(10):3249–3254, 2011.

[55] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631, 2019.

# A APPENDIX

## A.1 Baselines

We evaluated our approach against a diverse set of baseline models to benchmark its performance. These models are categorized into four groups based on their underlying principles.

- **GNN-based models**: These models learn node representations by aggregating information from a node's local neighborhood.
  - **GCN (Graph Convolutional Networks)**: A foundational GNN model that learns node embeddings by iteratively averaging feature vectors of a node and its neighbors. This convolution operation on the graph smooths features across the network [13].
  - **GraphSAGE (Graph SAmple and aggreGatE)**: An inductive GNN model designed to scale to large graphs. Instead of using all neighbors, it samples a fixed number of neighbors and aggregates their features using functions like mean or max-pooling. This allows it to generalize to unseen nodes [33].
  - **GIN (Graph Isomorphism Network)**: A powerful GNN architecture that is theoretically as expressive as the Weisfeiler-Lehman graph isomorphism test. GIN focuses on learning a strong aggregation function that can distinguish between different graph structures, making it highly effective for graph classification tasks [14].
  - **GAT (Graph Attention Networks)**: This model introduces an attention mechanism into the GNN framework. Instead of treating all neighbors equally, GAT learns a weight for each neighbor's contribution to the central node's representation. This allows the model to selectively focus on the most relevant neighbors, which is particularly useful for heterogeneous graphs [15].
- **Triplet-based models**: These models, also known as Knowledge Graph Embedding (KGE) models, learn to represent entities (like drugs and diseases) and relations as vectors in a continuous vector space.
  - **TransE (Translating Embeddings)**: A simple but influential model that represents relations as a translation vector. It assumes that for a valid triplet $(h, r, t)$, the embedding of the head entity $h$ plus the relation vector $r$ should be close to the tail entity $t$ (i.e., $h + r \approx t$) [16].
  - **TransR (Translating on Hyperplanes)**: An extension of TransE that addresses its limitations by projecting entities into a relation-specific vector space before performing the translation. This allows the model to better handle complex and varied relations [34].
  - **RotatE (Rotation-based Embeddings)**: This model represents relations as a rotation from the head entity to the tail entity in a complex vector space. This rotational mechanism is highly effective at capturing various relation patterns, including symmetry, anti-symmetry, and composition [35].
  - **ComplEx (Complex Embeddings)**: An embedding model that represents entities and relations using complex-valued vectors. It uses a bilinear scoring function to capture asymmetric relationships, which are common in knowledge graphs [36].
  - **RESCAL (Relation Factorization)**: This is a tensor factorization-based model that represents each relation as a full matrix. It learns a bilinear interaction between entity embeddings to predict the plausibility of a triplet, allowing it to model rich, complex relationships but with higher computational cost [17].
- **Pretrained biomedical language models**: These models are a different class of baselines that leverage the vast amount of biomedical text data available to learn rich, contextualized representations of biomedical entities.
  - **BioBERT, PubMedBERT, BioLinkBERT**: These are BERT-based models that have been pre-trained on large-scale biomedical corpora, such as PubMed abstracts and full-text articles. By training on this domain-specific text, they learn to understand the unique vocabulary and semantic relationships of biomedical concepts, which allows them to produce highly effective embeddings for tasks like entity recognition and relation extraction [22, 23, 12].

  - **SapBERT (Self-Alignment Pretraining for Biomedical Entity Representations)**: This model builds upon a biomedical language model (like PubMedBERT) but is further fine-tuned using a metric learning framework on massive biomedical ontologies (e.g., UMLS). Its primary goal is to align the representation space of synonymous or related entities, making it particularly powerful for entity linking tasks [24].

- **Path-based models**: These models capture information by exploring multi-hop paths and relational contexts within the knowledge graph, often through random walks.

  - **Node2vec**: A seminal model for learning node embeddings from random walks. It generates a series of random walks from each node and then uses an algorithm similar to Word2Vec to learn embeddings that capture both local neighborhood structure and global network position [28].
  - **DrugRep-KG**: A framework that constructs a drug-disease knowledge graph and uses a Word2Vec-like approach on the paths to create numerical vectors for entities and relationships. It specifically focuses on simplifying the relationships between drugs and diseases to better capture potential associations for repurposing [19].
  - **DREAMwalk (Drug Repurposing through Exploring Associations using Multilayer random walk)**: This model extends the concept of random walks by incorporating semantic information, like ATC codes, to guide the walk. When a walker lands on a drug or disease node, it can teleport to other semantically similar nodes. This process generates paths that are both biologically and semantically meaningful, leading to more informative embeddings [6].

### A.2 Statistics of the Knowledge Graphs

We evaluate DrugCORpath using two biomedical knowledge graphs: MSI and PrimeKG. To ensure that our evaluation focuses on mechanistically meaningful disease–drug reasoning, we curated subgraphs from both datasets.

The MSI knowledge graph was constructed from various supplementary biomedical datasets. For our experiments, we retained only nodes and edges related to drugs, proteins, diseases, and biological functions by incorporating the following types of relationships: drug–protein, disease–protein, protein–protein, protein–biological function, and biological function–biological function interactions. All drug–disease edges were removed prior to training, resulting in a filtered graph with 29,959 nodes and 478,728 edges.

PrimeKG provides a holistic and multimodal view of diseases by integrating 20 high-quality biomedical sources, encompassing 17,080 diseases and over four million relationships across ten biological levels. For our study, we excluded the drug effect relation, which connects drugs to their indications, to avoid data leakage. Among the wide array of relation types available in PrimeKG, we selected those most relevant to drug mechanisms: off-label use, drug–protein interactions, protein–protein interactions, disease–protein associations, and links between pathways or between biological processes and proteins. This filtering yielded a subgraph comprising 57,557 nodes and 670,013 edges.

For both KGs, we treated known drug–disease pairs as positive samples. Specifically, MSI included 5,926 such associations, while PrimeKG had 7,461. To generate negative samples, we randomly sampled drug–disease pairs that exist as nodes in the graph but are not connected by any edge. We adopted a 1:1 positive-to-negative ratio for MSI (5,926 positive and 5,926 negative samples), and a 1:1.2 ratio for PrimeKG (7,461 positive and 10,000 negative samples), reflecting the broader disease coverage in the latter.

### A.3 Computation of ATC similarities for Drug Teleport

To determine the teleport probability between drug-drug nodes, we use a semantic similarity measure based on the Anatomical Therapeutic Chemical (ATC) codes, a drug classification system defined by the WHO. We adapted an information content-based semantic similarity measure for drug-drug pairs by viewing the ATC hierarchy as a directed acyclic graph.

### A.3.1 Information content

In an ontology hierarchy, the information content (IC) of a term reflects its specificity. A term that appears less frequently or has fewer child terms is considered more specific and thus has a higher IC [37]. We compute the IC for a term $c$ directly from the hierarchy using this formula:

$$IC(c) = 1 - \log\left(\frac{N_{\text{child}}(c) + 1}{N_{\text{child}}(root) + 1}\right)$$

Here, $N_{\text{child}}(c)$ represents the number of direct child nodes of $c$. The denominator normalizes the IC values to be within the range of $[0, 1]$, with the root node having an IC of 0.

### A.3.2 Semantic similarity

We then compute the semantic similarity between two drugs, $c_1$ and $c_2$, based on their IC values. We adopt the distance measure from **(author?)** [38], which is defined as:

$$dist(c_1, c_2) = IC(c_1) + IC(c_2) - 2 \times IC\left(MICA(c_1, c_2)\right)$$

The term $MICA(c_1, c_2)$ refers to the Most Informative Common Ancestor (MICA), which is the most specific common parent of $c_1$ and $c_2$. To convert this distance measure into a similarity score that falls between $[0, 1)$, we use the following equation:

$$\mathbf{A}_{c_1,c_2} = sim(c_1, c_2) = 1 - \left(\frac{dist(c_1, c_2)}{2}\right)$$

This process is performed for all possible drug pairs within the knowledge graphs, resulting in a dense similarity matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, where $n$ is the total number of drugs.

## A.4 Detailed Results for MoA-Representative Path Clusters

To investigate how diverse biological mechanisms are captured by clustering in DrugCORpath, we analyze representative Gene Ontology (GO) of each cluster.

### A.4.1 Case 1: B-Cell lymphomas–Fostamatinib

For each path in a cluster, its constituent genes were tested against GO biological process terms, using all 17,660 genes in the KG as the background set. GO terms were considered significantly enriched if the adjusted p-value was below 0.05 and all genes in the path were included in the GO term. After identifying enriched GO terms at the path level, we summarized these terms at the cluster level by selecting terms that most frequently appeared across multiple paths within the cluster.

This case study demonstrates the diverse mechanisms of action (MoA) predicted for the association between Fostamatinib and B-cell lymphomas. Fostamatinib was selected for analysis due to its established therapeutic potential in targeting B-cell receptor (BCR) signaling, a crucial driver of B-cell malignancies, including lymphomas [39]. Each identified cluster represents distinct representative GO terms: Protein Modification Process (GO:0036211), Positive Regulation of Intracellular Signal Transduction (GO:1902533), Intracellular Signaling Cassette (GO:0141124), and Regulation of Canonical NF-kappaB Signal Transduction (GO:0043122). These clusters collectively illustrate how fostamatinib modulates disease pathways through distinct yet complementary biological mechanisms. For instance, some clusters highlight fostamatinib's involvement in regulating protein phosphorylation and modification, crucial for fine-tuning intracellular signaling dynamics [39]. Others emphasize its role in orchestrating complex intracellular signaling cascades that influence cellular proliferation and survival [40]. Additionally, fostamatinib distinctly modulates canonical NF-$\kappa$B signaling, a central pathway implicated in inflammation and cancer progression [41]. This diverse mechanistic landscape underscores the comprehensive biological relevance of fostamatinib's therapeutic action in B-cell malignancies.

Importantly, one representative and biologically well-supported path involves B-Cell Lymphomas $\rightarrow$ BCL10 $\rightarrow$ IKBKB $\rightarrow$ Fostamatinib. Here, BCL10 participates in the CARD11–BCL10–MALT1

(CBM) complex downstream of BCR signaling to activate IKK$\beta$ (encoded by IKBKB), thereby triggering NF-$\kappa$B transcriptional activity [31]. Fostamatinib, a SYK inhibitor, blocks BCR signaling upstream, disrupting CBM-mediated NF-$\kappa$B activation and showing clinical activity in diffuse large B-cell lymphoma [39, 42]. Additionally, other identified MoA clusters highlight mechanisms involving protein modification via key signaling proteins such as PTPN1 [43] and FGFR1 [44], intracellular signaling pathways via NRAS and EGFR [45], and further modulation of NF-$\kappa$B signaling through molecule like BIRC3 [46].

These findings demonstrate DrugCORpath's strength in producing interpretable and biologically grounded predictions by capturing a compact, diverse, and clinically relevant set of MoAs for a given disease–drug pair.

### A.4.2 Case 2: Breast Carcinoma–Regorafenib

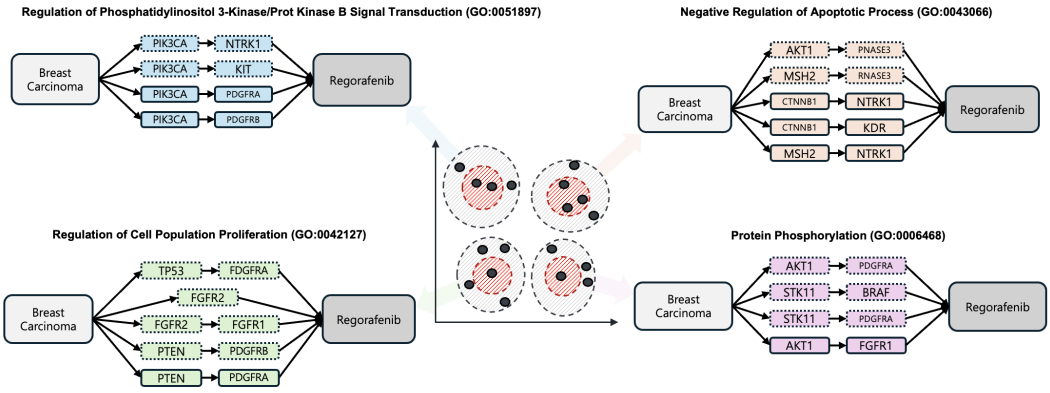



Figure 6: Case study on Breast Carcinoma–Regorafenib

To further demonstrate the interpretability of our method, we provide an additional case study in which clustered paths between Regorafenib and Breast Carcinoma exhibit distinct modes of action (MoA). As shown in Figure 6, enriched GO clusters such as Regulation of Phosphatidylinositol 3-Kinase/Protein Kinase B Signal Transduction (GO:0051897), negative regulation of apoptosis (GO:0043066), cell population proliferation (GO:0042127), and protein phosphorylation (GO:0006468) highlight different biological processes relevant to Regorafenib's mechanism.

Notably, each cluster emphasizes distinct gene sets and pathways.

For instance, PIK3CA, KIT, and PDGFRA/B in the Regulation of Phosphatidylinositol 3-Kinase/Protein Kinase B Signal Transduction (GO:0051897) cluster activate a central pro-survival axis in breast cancer, frequently deregulated by mutations and upstream receptor hyperactivation. Regorafenib inhibits these receptor tyrosine kinases (RTKs), suppressing AKT phosphorylation and thereby reactivating apoptotic regulators such as FOXO3a and Bim [47].

In the apoptosis regulation cluster, genes like CTNNB1, AKT1, and KDR act as negative regulators of cell death, maintaining survival through Wnt, PI3K, and VEGF signaling pathways. Regorafenib's inhibition of VEGFR2 (KDR) and AKT1 leads to increased apoptotic priming via mitochondrial depolarization and impaired endothelial support [48].

The cell population proliferation cluster features growth-promoting RTKs including FGFR1/2 and PDGFRA/B, which drive mitogenic signaling via MAPK and PI3K pathways. Regorafenib directly inhibits FGFR1 and PDGFRs, blocking mitogenic input and reducing tumor cell division and stromal interaction [49].

Lastly, the protein phosphorylation cluster, anchored by kinases like BRAF, STK11, and AKT1, encompasses core signaling nodes involved in oncogenic cascades. Regorafenib's potent inhibition of BRAF and multiple upstream RTKs disables phosphorylation-dependent signal propagation (e.g., MAPK and PI3K axes), disrupting cancer cell growth and survival [50].

These heterogeneous but biologically plausible MoAs demonstrate how our clustered embedding framework successfully captures diverse mechanistic perspectives within a single drug–disease pair, enhancing explainability at the subpathway level. This type of multi-view representation is particularly valuable for multi-kinase drugs like Regorafenib, which exert effects via multiple parallel mechanisms, each corresponding to distinct biological processes enriched in our subgraph clusters.

### A.5 Detailed Results For MoA Interpretability Cases

We further conducted case studies to evaluate the mechanistic validity of the generated paths. Specifically, our model was applied for recommending approved drugs for Blood Coagulation and Atopic Dermatitis.

#### A.5.1 Case 1: [Blood Coagulation Disorders → NRAS/KRAS → ABCC1 → Benzbromarone]

When predicting treatments for blood coagulation disorders, DrugCORpath recommended Benzbromarone, identifying a mechanistic path involving NRAS/KRAS and ABCC1. Although primarily used for gout and not FDA-approved for coagulation disorders, we explored the biological plausibility of this prediction.Mutations in NRAS/KRAS [51], key components of the RAS/MAPK pathway, have been linked to endothelial dysfunction and pro-thrombotic states, and are known to upregulate ABCC1 [52], a membrane transporter regulating drug and metabolite efflux. Overexpression of ABCC1 can disturb intracellular balance by excessively exporting endogenous bioactive molecules, thereby impacting inflammation and coagulation regulation [53]. Recent studies suggest that Benzbromarone inhibits ABCC1 [54], implying that it could counteract coagulation abnormalities arising from NRAS/KRAS-driven dysregulation. All intermediate steps in the predicted path were supported by existing biomedical evidence, highlighting DrugCORpath's ability to uncover clinically meaningful, biologically grounded associations, even beyond current FDA-approved indications.

#### A.5.2 Case 2: [Dermatitis, Atopic → IL6 → NR3C1 → Prednisolone]

The prediction of Prednisolone for atopic dermatitis reflects a well-established anti-inflammatory mechanism, demonstrating DrugCORpath's ability to recover clinically validated treatments. Prednisolone is a widely used therapy for atopic dermatitis, a chronic inflammatory condition characterized by excessive secretion of pro-inflammatory cytokines, such as IL6, in immune cells. IL6 is frequently observed in lesional skin of atopic patients and is known to modulate the activity of NR3C1, the glucocorticoid receptor. Prednisolone binds to NR3C1 and suppresses the expression of pro-inflammatory cytokines, including IL6, thereby reducing inflammation. This example illustrates how DrugCORpath captures biologically grounded MoAs by identifying meaningful molecular relationships among entities.

### A.6 Meta-path ablation Study Performance

Table 4: Performances of ablation models of DrugCORpath on MSI.

| | **AUROC** | **AUPRC** | **Accuracy** |
|---|---|---|---|
| Increased Intermediate Gene | 0.815 (0.017) | 0.827 (0.012) | 0.735 (0.016) |
| Increased Teleporatation Frequency | 0.810 (0.017) | 0.862 (0.009) | 0.735 (0.020) |
| Random Gene Sampling | 0.531 (0.080) | 0.588 (0.097) | 0.471 (0.077) |
| DrugCORpath | 0.842 (0.005) | 0.845 (0.009) | 0.775 (0.007) |

To investigate the impact of meta-path design choices on DrugCORpath's effectiveness, we conducted ablation studies by altering key aspects of the path generation process. Table 4 summarizes the performance changes on the MSI dataset when the following modifications were applied:

- **Increased Intermediate Gene:** We increased the maximum number of intermediate gene nodes ($k_{\max}$) from 2 to 3. While this allowed longer paths, it resulted in a 2.7% drop in AUROC (0.842 → 0.815), likely due to the inclusion of redundant or irrelevant genes that dilute the core mechanism of action (MoA).

- **Increased Teleportation Frequency:** We performed one additional round of ATC-guided teleportation, which slightly improved AUPRC but caused a 3.2% drop in AUROC (0.842 $\rightarrow$ 0.810). This suggests that while pharmacologically similar drugs may be included, the resulting paths become less biologically specific and introduce noise.
- **Random Gene Sampling:** Instead of following KG topology, we randomly selected intermediate genes. This led to a substantial drop in all metrics, with AUROC decreasing to 0.531 and accuracy falling to 0.471. These results highlight the importance of topologically and biologically meaningful path construction.

### A.7 Details Regarding Reproducibility

#### A.7.1 Data Availability

All datasets used in this study, namely MSI (Multi-Scale Interactome)[3] and PrimeKG[4], were retrieved directly from their original public repositories. Both MSI and PrimeKG are comprehensive biomedical knowledge graphs that integrate information from numerous databases, offering extensive annotations, including Anatomical Therapeutic Chemical (ATC) codes.

#### A.7.2 Hyperparameter Search Space

To optimize the hyperparameters of our stacking ensemble, we employed Optuna [55], a state-of-the-art Bayesian optimization framework. We first split the dataset into training, validation, and test sets, where the validation set was used solely for model selection and not exposed during final testing. During each trial, three base XGBoost models were trained with different sampled configurations, and their outputs were combined using logistic regression as a meta-classifier. The best configuration was selected by maximizing AUROC on the validation set.

The hyperparameter search space explored for our models is detailed in Table 5.

Table 5: Hyperparameter Search Space

| Hyperparameter | Search Space/Values | Selected Value |
|---|---|---|
| Learning Rate | [0.01 - 0.2] | 0.151 (XGB1), 0.142 (XGB2), 0.089 (XGB3) |
| Number of Estimators | [500 - 800] | 689 (XGB1), 800 (XGB2), 763 (XGB3) |
| Max Depth | [3 - 9] | 5 (XGB1/2), 9 (XGB3) |
| Subsample | [0.7 - 1.0] | 0.995, 0.812, 0.992 |
| Colsample by Tree | [0.6 - 1.0] | 0.707, 0.999, 0.858 |
| Min Child Weight | [1, 3, 6, 10] | 6 (XGB3 only) |
| Final Estimator | {Logistic Regression, XGB, MLP} | Logistic Regression |
| Hidden Dimensions | [128, 256, 512] | 128 (Node2Vec) |
| CV for Stacking | [3, 5, 10] | 5 |

#### A.7.3 Other Experimental Details

All experiments were conducted on a single NVIDIA RTX 3090 GPU with 24GB of memory, running on an Ubuntu 18.04 Linux operating system. Benchmark experiments were performed using 5-fold cross-validation. For each reported result, the mean and standard deviation are provided, as presented in the respective tables throughout the main body of our paper.

[3] https://github.com/snap-stanford/multiscale-interactome

[4] https://github.com/mims-harvard/PrimeKG