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

Abstract

Drug repurposing, which seeks novel drug–disease associations by integrating biomedical knowledge, is hindered by modeling complex multi-hop relationships in knowledge graphs. We propose **DrugCORpath**, which integrates biomedical knowledge graphs with pretrained biomedical large language models (LLMs). Unlike node-centric methods, it captures biological context by converting multi-hop drug–disease paths into sentences reflecting plausible mechanisms of action (MoAs) and embedding them. We then apply clustering–based selective filtering with a distance metric to retain meaningful paths while removing redundancy and noise. Experiments show DrugCORpath outperforms graph-based, LLM-based, and path-based baselines in drug repurposing, achieving up to 4.9% higher accuracy than the prior SOTA. Analyses confirm that filtering reduces noise and enhances biological diversity, and case studies validate clinically relevant rationales, improving interpretability. Collectively, these results underscore the method's potential for interpretable, biologically plausible drug repurposing.

1 Introduction

Drug repurposing, a central challenge in systems pharmacology and drug discovery, seeks new uses for existing drugs [1]. A key is understanding biological mechanisms that bridge diseases to treatments: many diseases involve dysregulated genes or dysfunctional signaling pathways, which drugs target via mechanisms of action (MoA) [2]. Accurately capturing these intermediate processes is crucial for plausible and interpretable predictions. Biomedical knowledge graphs (BKGs) integrate genes, diseases, drugs, and relations (gene–gene, drug–gene, disease–gene), enabling multi-hop reasoning that reflects how diseases influence genes and network modules, which are in turn modulated by drugs [3, 4]. However, many KG-based approaches rely on general-purpose embeddings or path counting that overlook gene-level MoA context and the meaning, directionality, and function of intermediate relations [5, 6], limiting mechanistic interpretability and accuracy [7, 8]. BKGs are also large and dense, yielding many redundant or irrelevant multi-hop paths [9]. When multiple genes in the same pathway connect a drug–disease pair, the graph can produce biologically equivalent paths [10]; prior work found only a small fraction of meta-paths uniquely predictive, underscoring the need for filtering to enhance performance and interpretability [11].

*Corresponding author.
[1]Emails: {haerinsong, eugenomics, bikoo95, sunkim.bioinfo}@snu.ac.kr, ss.lee@inha.ac.kr

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: 2nd Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences.

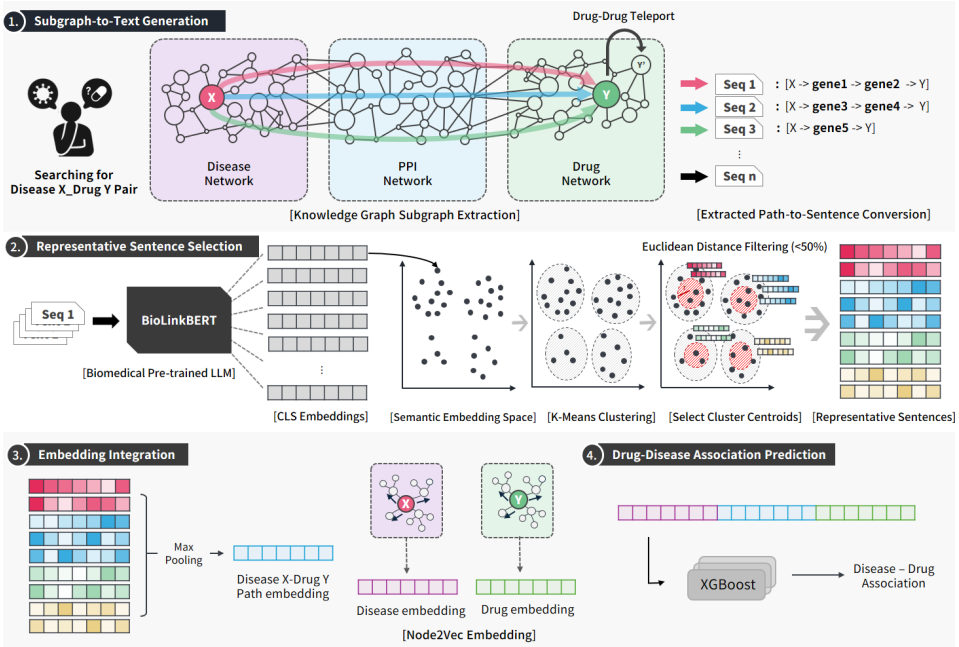

Figure 1: **Overview of DrugCORpath.** For a disease–drug pair, context paths from a biomedical KG are converted to sentences and encoded by a biomedical LLM. K-means selects representative paths; their embeddings are max-pooled, combined with Node2vec disease/drug embeddings, and fed to an XGBoost classifier to predict the association.

To address these issues, we propose **DrugCORpath** (COntext-aware Reasoning path), which couples KG traversal with LLM-based semantic encoding and selective path filtering. We extract disease–gene–drug paths likely to reflect plausible MoAs, convert them into natural-language sentences, and embed them using biomedical LLM [12]. We then apply K-Means clustering to reduce redundancy and retain diverse, biologically meaningful paths, improving both accuracy and interpretability. Evaluations on MSI and PrimeKG with unseen diseases, and case analyses show that incorporating LLM-derived context and clustering-based selection filters noise, reduces semantic drift via a fixed meta-path pattern, and yields compact yet informative, pathway-based explanations.

In summary, 1) we introduce DrugCORpath, a KG–LLM integrative framework that outperforms 14 baselines across multiple KGs, achieving up to 4.9% accuracy improvement over prior SOTA; 2) we show that LLM-generated sentence-level path embeddings capture biologically plausible and interpretable MoA contexts; 3) we propose clustering-based path selection that filters noise and preserves concise, diverse, mechanism-specific path representations for interpretable repurposing; and 4) through case studies (including FDA-approved treatments), we validate clinically meaningful rationales beyond mere structural similarity.

2 Methodology

DrugCORpath predicts disease–drug associations by (i) extracting *disease–gene–drug* meta-paths from a biomedical knowledge graph (BKG), (ii) converting paths to sentences and encoding them with a biomedical LLM, (iii) selecting representative sentences via K-Means to capture diverse mechanisms while reducing redundancy, (iv) aggregating their embeddings (max pooling), and (v) combining this context with disease/drug network embeddings for classification (Figure 1).

Problem Formulation We consider a heterogeneous KG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \mathcal{V}_{\text{drug}} \cup \mathcal{V}_{\text{disease}} \cup \mathcal{V}_{\text{gene}}$ (relations: gene–gene, drug–gene, disease–gene) and no direct disease–drug edges. Given disease s and drug d, we learn a binary predictor of association.

Disease–Drug Context Generation via KG Exploration We extract mechanistic context as paths in which a disease connects to a drug through a sequence of gene nodes under a fixed, MoA-consistent

Table 1: Performance of DrugCORpath vs. 14 baselines across four model categories on MSI and PrimeKG (mean±std over 5-fold CV).

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
Transition-based	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**

meta-path schema (disease→gene→ $\cdots$ →gene→drug). The search is constrained to relations that reflect plausible biology (e.g., disease–gene association, gene–gene interaction, drug–gene targeting), respects directionality by normalizing edges to disease→drug flow, and disallows node repetitions to avoid trivial cycles. We cap the number of intermediate genes at $k_{\max}$ to curb redundancy from dense gene subgraphs and to reduce semantic drift. For each (s, d), the candidate paths are

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

When no such path exists, we apply a *single* ATC-based substitution: replace d with a pharmacologically similar d' from the ATC similarity network to form $s \to g_1 \to \cdots \to g_k \to d'$, and finally link back to d (details in Appendix A.4). Each finalized path is verbalized with a relation-aware, directional template (e.g., "Disease A associated-with Gene B; Gene B interacts-with Gene C; Drug D targets Gene C"), yielding a sentence that preserves biological flow for identifying plausible MoAs.

Contextual Embedding and Selection of Disease–Drug Paths Each path sentence is encoded with BioLinkBERT [12] using a *frozen* encoder; we take the [CLS] representation as its contextual embedding, producing $\mathbf{H}_{s,d} = \{\mathbf{h}_i\}$. To address oversampling and emphasize distinct mechanisms, we cluster $\mathbf{H}_{s,d}$ with K-Means (Euclidean distance) and select representatives per cluster by proximity to the centroid. Concretely, for cluster k with centroid $\mathbf{c}_k$, we keep embeddings within the top $\rho\%$ closest to $\mathbf{c}_k$:

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

where τ_k is the ρ-th percentile of distances in cluster k. The resulting representatives form a compact, diverse set of MoA-consistent contexts. We aggregate their embeddings via max pooling to obtain a unified context vector, which is later fused with disease/drug topology for prediction. We set $k = 4$, which yielded stable performance across evaluation metrics. K-Means was chosen for its simplicity, interpretability, and efficiency in high-dimensional embedding space, allowing us to effectively filter redundant paths and retain diverse mechanism-consistent contexts.

Disease–Drug Association Prediction To incorporate topology, we compute Node2vec [13] embeddings for disease and drug nodes and concatenate them with the unified context. A stacking ensemble (three XGBoost base learners with a Logistic Regression meta-learner) outputs the association probability. We train with binary cross-entropy and mitigate class imbalance via negative sampling of non-associated pairs.

3 Results and Conclusion

We evaluate DrugCORpath on two biomedical KGs: MSI [14] and PrimeKG [15]. To enforce a zero-shot setting, all disease–drug edges are removed prior to training (Appendix A.3. The task is binary classification with positives from known associations and negatives from randomly sampled

unconnected pairs (MSI: 1:1; PrimeKG: 1.2:1). We adopt an unseen-disease split (no disease overlap between train/test), perform 5-fold cross-validation, and report AUROC, AUPRC, and Accuracy.

For comparison, we benchmark against 14 baselines across four categories—GNN-based, KG-embedding, pretrained biomedical LMs, and path-based—with details in Appendix A.2.

Table 1 summarizes results on MSI and PrimeKG. DrugCORpath attains the best AUROC, AUPRC, and Accuracy on both datasets with low variance.

On MSI, it improves over the strongest LLM baseline (PubMedBERT) by +2.9% AUROC, +2.7% AUPRC, and +4.9% Accuracy. On PrimeKG, it surpasses path-based models (Node2vec, DREAMwalk) by up to +1.7% AUROC, +1.8% AUPRC, and +4.2% Accuracy. These gains show that KG-derived context with LLM path semantics outperforms structural-only or text-only baselines. The consistently low standard deviations suggest the clustering-based selection effectively filters noisy/redundant paths, stabilizing performance while preserving mechanistic relevance.

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.

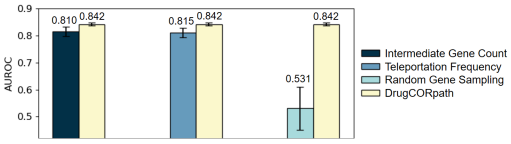

Figure 2: AUROC of meta-path ablation strategies on MSI. Each group compares DrugCORpath with a single-component ablation;

We probe the meta-path design with three changes: (i) raising $k_{\max}$ to 3 extends paths, adding redundant or weakly relevant genes that dilute MoA specificity; (ii) an extra ATC teleportation broadens exposure to similar drugs but weakens causal grounding; (iii) random gene sampling ignores KG topology, breaking MoA consistency. As shown in Figure 2, AUROC drops by 2.7%, 3.2%, and 31.1% for (i)–(iii), confirming that concise, topology-respecting paths are more effective for capturing MoA-relevant relationships.

For clustering-based selection, path embeddings are clustered (K-Means, Euclidean), and the top $\rho\%$ near each centroid are kept to form a compact, diverse set. As shown in Table 2, Sweeping $\rho \in 20, 50, 80, 100$ shows $\rho = 50$ yields the best AUROC/AUPRC/Accuracy; $\rho = 80$ is similar but noisier, while $\rho = 20$ discards signal and $\rho = 100$ admits redundancy.

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

ρ	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

Biological Case Study To assess the biological interpretability of DrugCORpath, we analyzed both systematic evaluations and case studies. First, representative GO terms from clustered paths were examined to determine whether the model captures mechanistically distinct processes.

As shown in Figure 3, most disease–drug pairs were divided into clusters with low GO term similarity, suggesting that clustering-based path filtering captures diverse mechanisms of action (MoAs). Figure 4 illustrates that, in the B-cell lymphomas–Fostamatinib case, clusters aligned with GO terms such as protein modification, intracellular signaling regulation, and canonical NF-κB signaling, with one representative path (B-cell lymphomas $\rightarrow$ BCL10 $\rightarrow$ IKBKB $\rightarrow$ Fostamatinib) tracing upstream interference in NF-κB activation. We also examined the effect of redundant path filtering using GO-term overlap, and observed that filtering consistently improved performance, with the gap widening as redundancy increased (Appendix A.6). Beyond systematic evaluation, DrugCORpath uncovered novel hypotheses and validated therapies. For blood coagulation disorders, it recommended Benzbromarone via NRAS/KRAS and ABCC1, supported by literature linking these genes to thrombotic

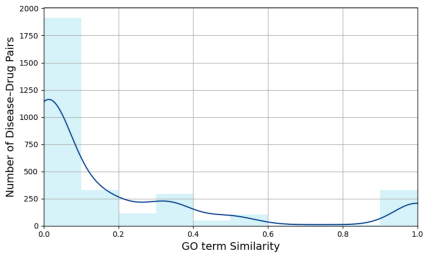

Figure 3: Distribution of similarity scores between GO terms assigned to different clusters within each disease–drug pair.

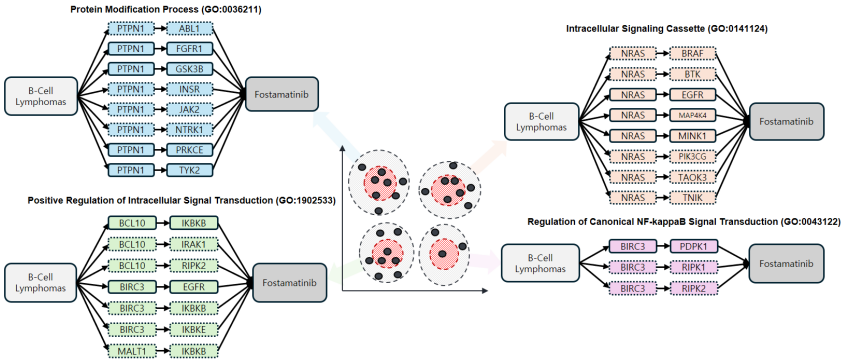

Figure 4: Case study illustrating path clusters identified by DrugCORpath for the B-cell lymphomas–Fostamatinib pair, with each cluster annotated by its most frequently observed GO term representing a mechanism of action (MoA). Shown clusters are enriched for distinct biological processes.

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**

signaling and metabolite transport. For atopic dermatitis, it correctly identified the established treatment Prednisolone through IL6-driven activation of NR3C1 (Appendix A.7). The Benzbromarone–coagulation link, though not FDA-approved, represents a biologically grounded repurposing case highlighting unexplored therapeutic opportunities. Collectively, these results demonstrate that DrugCORpath not only captures compact and diverse mechanistic insights but also provides clinically relevant predictions aligned with both known and emerging treatments.

FDA Drug Repurposing Prediction To further demonstrate the applicability of our framework to real-world drug repurposing, we evaluated recently approved drug–disease pairs from the FDA in 2024–2025[2], focusing on new indications assigned to existing drugs. In a prospective evaluation using these recent FDA-approved indications, DrugCORpath consistently assigned high probabilities to newly validated drug–disease associations and ranked first in most cases (Table 3). Notably, it correctly predicted the approval of Ranibizumab for diabetic retinopathy, distinct from its original indication in age-related macular degeneration, whereas baseline models assigned substantially lower scores.

Conclusion : We presented DrugCORpath, a KG-based drug repurposing framework that combines meta-path exploration with LLM-derived biological context. Using clustering-based path selection, it captures diverse MoA semantics while preserving biologically meaningful paths. On MSI and PrimeKG under unseen-disease splits, DrugCORpath achieves state-of-the-art performance, suggesting that prioritizing informative KG paths is a promising direction for improving drug repurposing.

[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] 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.

[7] 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.

[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] 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.

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

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

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

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

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

[26] 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.

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

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

[29] 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.

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

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

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

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

[40] 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.

[41] 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.

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

[43] 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.

[44] 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.

[45] 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-κb pathway genes in splenic marginal zone lymphoma. *Blood, The Journal of the American Society of Hematology*, 118(18):4930–4934, 2011.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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 Related Work

Knowledge Graph-based Drug Repurposing As BKG have become foundational in drug repurposing systems, traditional graph learning methods such as GNNs [16, 17, 18] or triplet-based methods [19, 20] 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 [21].

Path-based Reasoning Path-based models enhance explainability in KG reasoning by modeling chains of biomedical relations. DrugRep-KG [22] improves performance under label imbalance by using informative negative links. DREAMwalk [7] generates drug–gene–disease paths via semantic random walks. Select and Augment [23] 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 [24].

LLM-based Biomedical Reasoning LLMs pretrained on biomedical corpora, such as BioBERT [25], PubMedBERT [26], and SapBERT [27], have recently shown promise in clinical and molecular reasoning tasks. These models capture semantic dependencies beyond graph structure and support natural language inference [28]. While these approaches enable flexible reasoning across biomedical contexts, the lack of structured knowledge-based constraints can lead to hallucinated associations [29], resulting in predictions often lacking interpretability [30].

A.2 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 [16].
  - **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 [31].
  - **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 [17].
  - **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 [18].
- **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., $h + r \approx t$) [19].
  - **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 [32].

  - **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 [33].
  - **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 [34].
  - **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 [20].

- **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 [25, 26, 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 [27].

- **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 [13].
  - **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 [22].
  - **DREAMwalk (Drug Repurposing through Exploring Associations using Multi-layer 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 [7].

A.3 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.4 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.4.1 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 [35]. 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.4.2 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?)** [36], 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.5 Ablation Study Performance

A.5.1 Meta-path ablation

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:

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)

- **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 $\rightarrow$ 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.5.2 Effect of Clustering Percentile Thresholds

To evaluate the impact of clustering-based sentence filtering, we varied the proportion of path sentences (ρ) retained within each cluster (Table 2). At $\rho = 0\%$, no filtering was applied, while larger values retained sentences closest to cluster centroids. The results show that $\rho = 50\%$ achieves the best overall performance (AUROC=0.842, AUPRC=0.845, Accuracy=0.755), striking a balance between removing redundancy and preserving informative content. Retaining too many sentences ($\rho = 80\%, 100\%$) introduces noise, whereas overly aggressive filtering ($\rho = 20\%$) discards useful information. These findings confirm that clustering-based selection at moderate thresholds improves robustness and accuracy by denoising redundant paths while maintaining mechanistic diversity.

A.6 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) terms of each cluster. As shown in Figure 5, filtering redundant paths consistently improves performance across varying levels of GO term redundancy. Specifically, when evaluating the top 50%, 30%, and 10% of disease–drug pairs with the highest redundancy, the filtered setting (yellow) outperforms the unfiltered setting (gray) in AUPRC by margins of 1.2%, 2.0%, and 2.4%, respectively. These results highlight that redundancy-aware filtering is especially beneficial under conditions of high functional overlap, as it helps retain biologically meaningful and mechanism-specific paths while reducing noise.

A.6.1 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 [37]. As illustrated in Figure 4, 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

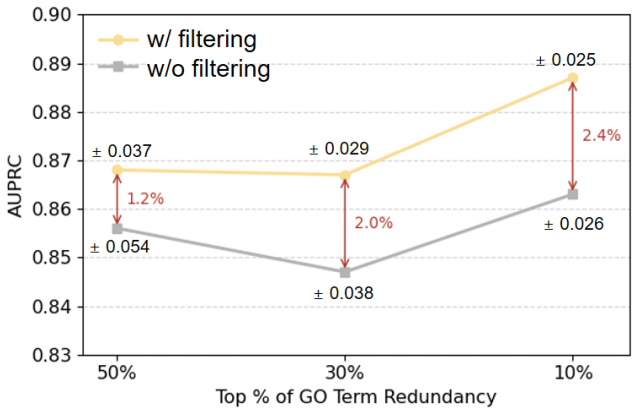

Figure 5: AUPRC comparison across GO term redundancy levels using the top disease–drug pairs. Standard deviations are shown.

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 [37]. Others emphasize its role in orchestrating complex intracellular signaling cascades that influence cellular proliferation and survival [38]. Additionally, fostamatinib distinctly modulates canonical NF-κB signaling, a central pathway implicated in inflammation and cancer progression [39]. 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 IKKβ (encoded by IKBKB), thereby triggering NF-κB transcriptional activity [40]. Fostamatinib, a SYK inhibitor, blocks BCR signaling upstream, disrupting CBM-mediated NF-κB activation and showing clinical activity in diffuse large B-cell lymphoma [37, 41]. Additionally, other identified MoA clusters highlight mechanisms involving protein modification via key signaling proteins such as PTPN1 [42] and FGFR1 [43], intracellular signaling pathways via NRAS and EGFR [44], and further modulation of NF-κB signaling through molecule like BIRC3 [45].

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.6.2 Case 2: 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 [46].

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

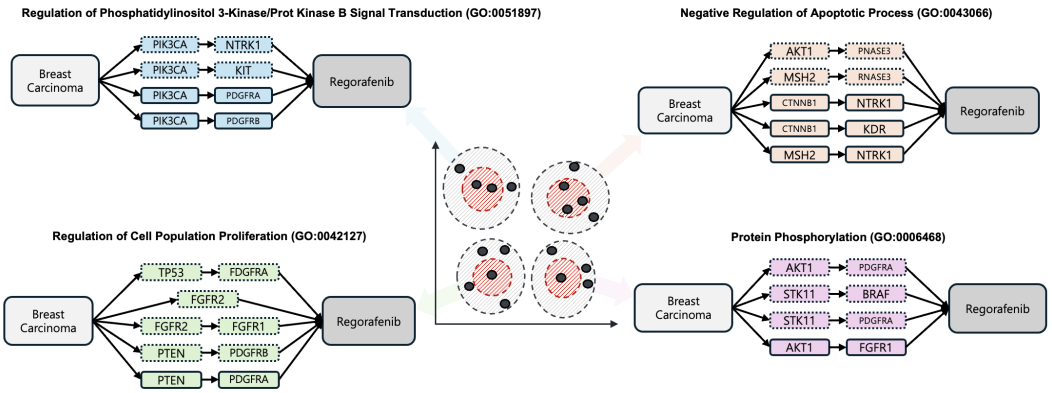

Figure 6: Case study on Breast Carcinoma–Regorafenib

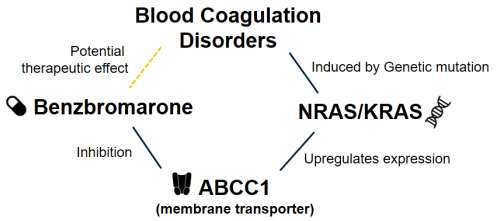

Figure 7: Case study of an interpretable drug–disease path. Edges show biological relations; the yellow dashed line marks the predicted association.

inhibition of VEGFR2 (KDR) and AKT1 leads to increased apoptotic priming via mitochondrial depolarization and impaired endothelial support [47].

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 [48].

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 [49].

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.7 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.7.1 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, as shown in Figure 7. Although primarily used for gout and not FDA-approved for coagulation disorders, we explored the biological plausibility of this prediction.Mutations in NRAS/KRAS [50], key components of

the RAS/MAPK pathway, have been linked to endothelial dysfunction and pro-thrombotic states, and are known to upregulate ABCC1 [51], 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 [52]. Recent studies suggest that Benzbromarone inhibits ABCC1 [53], 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.7.2 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.8 FDA Drug Repurposing Prediction

Table 3 reports predicted probabilities of drug–disease associations for FDA-approved new indications from 2024–2025, comparing DrugCORpath with BioLinkBERT and Node2vec. DrugCORpath assigns consistently higher probabilities for most cases, ranking first in 7 out of 11 new indications. For example, Ranibizumab–Diabetic Retinopathy and Nivolumab–Urothelial Carcinoma are strongly predicted by DrugCORpath with probabilities of 0.903 and 0.894, respectively, whereas baseline models assign much lower scores. Although BioLinkBERT and Node2vec occasionally achieve higher probabilities (e.g., Iloperidone–Bipolar I Disorder, Bernalizumab–Severe Asthma), these results demonstrate that DrugCORpath can effectively anticipate drug–disease associations later validated as new FDA indications, highlighting its potential utility in real-world drug repurposing scenarios.

A.9 Details Regarding Reproducibility

A.9.1 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.9.2 Hyperparameter Search Space

To optimize the hyperparameters of our stacking ensemble, we employed Optuna [54], 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.

[3]https://github.com/snap-stanford/multiscale-interactome
[4]https://github.com/mims-harvard/PrimeKG

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.9.3 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.