Acta Scientific MEDICAL SCIENCES (ISSN: 2582-0931)

Volume 10 Issue 7 July 2026

Research Article

Deducing Drug-Drug Interactions Through Graph Neural Networks

Dr. Anusha Sunder[1]*, Samyuktha Sunkara[2], Nikhilesh Anand[2] and Kirtikaa Chezhian[2]

[1]Doctorate in Life Science/Human Nutrition, Lead Scientist and Nutrigenetic Expert, Xcode Life Sciences, Pvt. Ltd. Chennai, India

[2]R&D Intern, Xcode Life Sciences, Pvt. Ltd. Chennai, India

***Corresponding Author:** Dr. Anusha Sunder, Doctorate in Life Science/Human Nutrition, Lead Scientist and Nutrigenetic Expert, Xcode Life Sciences, Pvt. Ltd. Chennai, India.

DOI: 10.31080/ASMS.2026.10.2254

Received: May 06, 2026

Published: June 24, 2026

© All rights are reserved by **Dr. Anusha Sunder., *et al.***

Abstract

Drug–drug interactions (DDIs) are a major contributor to adverse drug reactions, particularly in patients undergoing polypharmacy for chronic conditions such as diabetes. Although databases such as DrugBank provide validated interactions, their coverage remains incomplete, leaving many potential DDIs unreported. As a result, many computational approaches treat these unknown drug pairs as non-interacting, thereby introducing label noise and reducing prediction reliability. In this study, we propose a Graph Neural Network (GNN)-based framework that integrates heterogeneous biomedical data with Positive–Unlabeled (PU) learning to address these challenges. Known interactions are treated as positive samples, while unknown pairs are handled as unlabeled to identify reliable non-interactions through a data-driven approach. A heterogeneous graph incorporating drug–drug, drug–protein, protein–protein, and pathway relationships is constructed and further enriched with side effects and Gene Ontology annotations. The enriched model demonstrates improved predictive performance and generalization, and external validation on unseen diabetes drug combinations shows strong agreement with clinical literature. Overall, the proposed framework provides a clinically relevant and data-driven approach for DDI prediction, while also enabling extension to polypharmacy risk assessment.

Keywords: Drug–Drug Interactions (DDIs); DrugBank; Graph Neural Network (GNN)

Introduction

Drug–drug interactions occur when the pharmacological effect of one drug is altered by the presence of another, potentially leading to adverse outcomes such as toxicity or reduced therapeutic efficacy. This concern becomes more significant in patients with chronic diseases such as diabetes, where multiple medications are prescribed simultaneously to manage associated comorbidities, including hypertension, dyslipidemia, and cardiovascular disorders [1-3]. While such treatment strategies are clinically necessary, they substantially increase the likelihood of unintended interactions. However, despite the need for accurate identification, databases such as DrugBank provide experimentally validated interaction data with incomplete coverage, leaving many potential interactions uncharacterized [4]. As a result, computational approaches often assume that drug pairs without reported interactions are non-interacting. This closed-world assumption introduces incorrect

labels into the dataset and leads to biased model learning, as many unknown drug pairs may correspond to true but unreported interactions [5].

Graph Neural Networks (GNNs) have emerged as effective tools for modeling complex biological systems due to their ability to capture relationships between drugs, proteins, and pathways [6-8]. However, most existing approaches do not explicitly address the issue of unreliable negative samples arising from this assumption. To address this limitation, we integrate Positive–Unlabeled learning with heterogeneous graph modeling, enabling the identification of reliable negative samples in a data-driven manner [9]. This approach improves the robustness and clinical relevance of DDI prediction. In addition, the proposed framework supports extension to polypharmacy by enabling risk assessment across multi-drug combinations.

Materials and Methods

Two graph-based models were constructed to integrate functional data and physicochemical properties for predicting drug–drug interactions. The first model (Model 1) focuses on core biological relationships by incorporating drug–drug interactions, drug–protein associations, protein–protein interactions, and protein–pathway connections, thereby capturing how drugs interact through shared targets and biological pathways. Building on this foundation, the second model (Model 2) extends the framework by incorporating additional biological layers such as side effects and functional annotations, enabling the capture of interactions through phenotypic effects and deeper biological functions. For baseline comparison, a homogeneous model (Model 0) was constructed exclusively with drug nodes and drug–drug interaction edges to evaluate the contribution of biological context. To support these models, the overall methodology follows a structured pipeline as shown in Figure 1.

Dataset collection and heterogeneous graph construction

A set of 100 drugs related to diabetes and its comorbidities was selected from DrugBank, and known drug-drug interactions were used as positive samples. To add biological context, 216 protein nodes were included using drug protein interactions from DrugBank and protein protein interactions from STRING. In addition, 284 pathway nodes from KEGG were connected to proteins to represent biological processes.

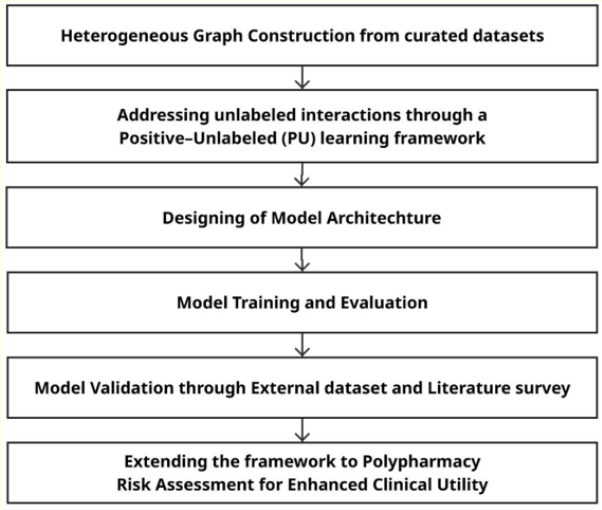


Figure 1: Methodology workflow.

Overview of the proposed pipeline, including heterogeneous graph construction, Positive–Unlabeled learning, model architecture, training and evaluation, external validation, and polypharmacy risk assessment.

In addition to the graph structure, each drug node was represented using structural and physicochemical features, including Morgan fingerprints and molecular descriptors such as molecular weight, LogP, hydrogen bond donors, and hydrogen bond acceptors. In the enriched model, more biological layers were added to improve representation. A total of 1307 side effect nodes from SIDER were connected to drugs, and 3113 functional annotation nodes from Gene Ontology were linked to proteins. This multi-layered structure helps the model understand drug interactions not only at the drug level but also through shared targets, pathways, and biological functions.

Positive–unlabeled learning

After constructing the heterogeneous graph, a major challenge in DDI prediction is the absence of confirmed negative samples. Therefore, instead of assuming that unknown drug pairs are non-interacting, they are treated as unlabeled data [9].

To address this, a preliminary neural network model is trained using known positive pairs and unlabeled pairs. This model assigns interaction probability scores to all unlabeled pairs. Based on these scores, a conservative threshold is applied to identify reliable non-

interaction pairs. Only those pairs with low predicted interaction probabilities are selected as negative samples, while uncertain pairs are excluded. This approach ensures that the training dataset contains only high-confidence samples, thereby reducing label noise and improving the model's ability to learn meaningful interaction patterns.

Model architecture

Two heterogeneous graph neural network models are developed in this study [6,7]. The first model integrates drug, protein, and pathway nodes, capturing core biological relationships. The second model extends this framework by incorporating side effects and Gene Ontology nodes, thereby adding phenotypic and functional context. Both models employ message-passing mechanisms to learn node representations and use a multilayer perceptron to predict interaction probabilities between drug pairs. The enriched model includes additional regularization to handle the increased complexity of the graph.

Training and evaluation

The models are trained on the curated dataset consisting of known positive interactions and reliable negative samples identified through the PU learning framework. The dataset is split into training, validation, and test sets using stratified sampling to maintain class balance. Training is performed using a binary cross-entropy loss function with appropriate class weighting, and model performance is evaluated using standard metrics including ROC-AUC, accuracy, precision, recall, and F1-score.

To further assess generalizability, the trained model is evaluated on unseen drug pairs generated from a subset of diabetes-related drugs. Interaction scores predicted by the model are categorized into low, moderate, and high-risk groups based on predefined thresholds, and selected predictions are validated against published clinical literature to ensure clinical relevance.

Polypharmacy risk assessment

The proposed framework is extended to evaluate drug interaction risk in polypharmacy scenarios. For each multi-drug combination, all possible pairwise drug–drug interactions are generated, and their interaction probabilities are predicted using the trained model. These pairwise scores are aggregated to compute an overall risk score for the drug combination. Additionally, drug-level contribution analysis is performed by examining the interaction scores associated with each drug across all pairs within the combination.

Results

Model performance

The performance of the proposed models is evaluated using standard metrics, including ROC-AUC, accuracy, precision, recall, and F1-score. The baseline heterogeneous model demonstrates a clear improvement over the homogeneous model with only drug nodes, highlighting the importance of incorporating biological relationships. The enriched model further improves performance across all metrics, particularly in recall and F1-score, indicating better identification of true interactions.

Metric	Model 0 (Homogeneous)	Model 1	Model 2
ROC-AUC	0.942	0.979	0.980
Accuracy	0.866	0.938	0.952
Precision	0.939	0.970	0.971
Recall	0.854	0.940	0.959
F1 score	0.895	0.955	0.965

Table 1: Performance comparison of the models.

ROC-AUC (Receiver-operating characteristic curve (ROC) Area under the curve) – ability of the model to distinguish between interacting and non-interacting drug pairs across different thresholds.

Accuracy – overall correctness of predicted drug–drug interactions.

Precision – proportion of predicted interactions that are true DDIs, indicating reliability of positive predictions.

Recall – ability to correctly identify known interacting drug pairs, important for minimizing missed harmful interactions.

F1 score (F-measure) – balance between precision and recall, reflecting overall effectiveness in detecting clinically relevant DDIs.

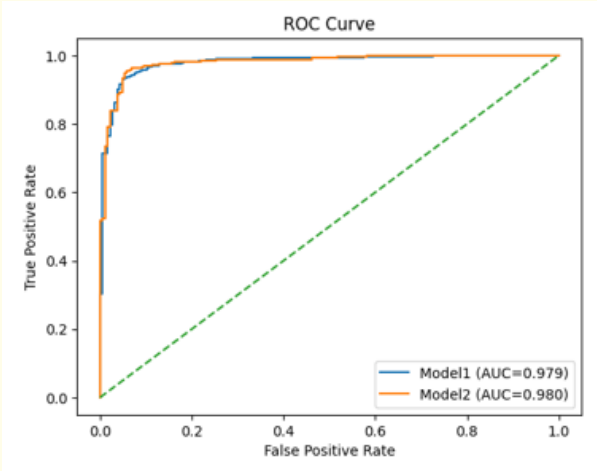
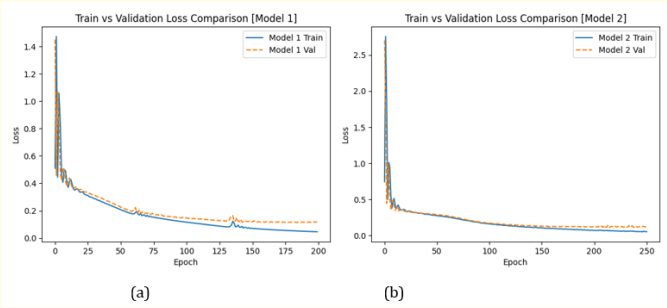


Figure 2: Receiver Operating Characteristic (ROC) curves for the models.

The enriched model shows a marginal improvement in AUC, indicating enhanced discriminative capability after incorporating additional biological information.



(a) (b)

Figure 3: Training and validation loss curves for Models 1 and 2.

(a) Model 1 shows steady convergence across epochs with a minimal gap between training and validation loss, indicating stable learning with limited overfitting. (b) Model 2 demonstrates faster convergence and lower validation loss compared to Model 1, suggesting improved generalization due to the incorporation of enriched biological features.

External validation and literature analysis

The model shows clear separation of predicted interaction scores across unseen drug pairs, with distinct low, moderate, and high-risk categories. High-risk predictions align with clinically reported adverse interactions, while lower-risk categories correspond to combinations with minimal or manageable interaction potential, indicating effective generalization and meaningful risk stratification.

S. No.	Drug 1	Drug 2	Predicted Risk	Clinical Interpretation
1	Biguanides (Metformin)	Sulfonylureas (Gliclazide, Glyburide, Glipizide, Glimepiride)	High	Patients newly prescribed a sulfonylurea in the setting of metformin had a higher risk of all-cause mortality (HR 1.44, 95% CI 1.12 to 1.84, p = 0.005) and major hypoglycemic episodes (HR 2.78, 95% CI 1.66 to 4.66, p < 0.001) than those prescribed an 'other' OH(Oral Hypoglycemic Agent) [13]. Combination therapy of metformin and sulfonylurea significantly increased the RR (Relative Risk) of cardiovascular hospitalization or mortality (fatal and nonfatal events) irrespective of the reference group [14].
2	SGLT2 Inhibitors (Ertugliflozin)	Diuretics (Spironolactone, Eplerenone)	High	Among 4,417,195 reports, 1,636 ARF (Acute Renal Failure) cases were associated with SGLT2i combinations, primarily involving diuretics, lipid-lowering agents, and anticoagulants [15].

3	ACE Inhibitors (Lisinopril, Fosinopril)	Sulfonylureas (Glyburide, Glipizide, Tolazamide)	High	The combination of ACE inhibitors with sulfonylureas in diabetic patients may increase the risk of hypoglycemia [16].
4	Statins (Lovastatin)	DPP-4 Inhibitors (Sitagliptin)	Moderate	Real-world FAERS (FDA Adverse Event Reporting System) data do not raise concern for muscular toxicity with DPP4-is in combination with statins, making a drug interaction very unlikely [17].
5	ARBs (Telmisartan)	SGLT2 Inhibitors (Ertugliflozin)	Moderate	The combination therapy with SGLT2 inhibitors and ACEIs/ARBs was well tolerated. Although SGLT2 inhibitors should not increase the risk of hypoglycaemia, as they do not activate insulin secretion, the results showed an increased risk of hypoglycaemia when treating with SGLT2 inhibitors in combination with ACEIs/ARBs. Treatment with SGLT2 inhibitors plus ACEIs/ARBs (249 patients) decreased eGFR (Estimated Glomerular Filtration Rate) [18].
6	Statins (Fluvastatin)	SGLT2 Inhibitors (Ertugliflozin)	Low	Clinically meaningful drug interactions via ertugliflozin-mediated inhibition or induction of CYP/UGT isozymes and transporters with concomitant medications that are substrates of OAT3 (sitagliptin), OCT2 (metformin), CYP isozymes (glimepiride, statins), and OATP1B1 (statins) are not anticipated [19].
7	SGLT2 Inhibitors (Dapagliflozin, Empagliflozin)	DPP-4 inhibitor (Sitagliptin), Biguanide (Metformin), Sulfonylurea (Glimepiride), Statin (Simvastatin)	Low	The lack of clinically meaningful pharmacokinetic interactions demonstrates that ertugliflozin an SGLT2 inhibitor can be coadministered safely with sitagliptin, metformin, glimepiride, or simvastatin without any need for dose adjustment [19].

Table 2: Literature validation of representative drug combinations.

ACE – Angiotensin-Converting Enzyme; ARB – Angiotensin Receptor Blocker; SGLT2 – Sodium–Glucose Cotransporter 2; DPP-4 – Dipeptidyl Peptidase-4.

Polypharmacy risk prediction for enhanced clinical utility

The model was applied to multi-drug scenarios to assess interaction risk across drug combinations, producing distinct risk profiles that enabled identification of high-risk regimens. The analysis further highlighted specific drugs that contribute disproportionately to the overall interaction risk within a combination. These insights support targeted optimization of prescriptions through substitution with lower-risk alternatives within the same therapeutic class. Overall, the results demonstrate the practical applicability of the model beyond pairwise prediction and highlight its potential as a clinical decision-support tool for safer polypharmacy management.

Discussion

The results demonstrate that integrating Positive–Unlabeled learning with heterogeneous graph modeling significantly improves drug–drug interaction prediction. By avoiding the incorrect assumption that unknown drug pairs are non-interacting, the PU learning framework reduces label noise and enables the model to learn from more reliable training data. This is particularly important in biomedical datasets, where many true interactions remain unreported. The heterogeneous graph structure further strengthens the model by capturing relationships across multiple biological levels, including drug–protein associations, protein interaction networks, and pathway-level mechanisms.

This layered representation helps the model capture indirect interactions that are not obvious from drug-level features alone. The improved performance of the enriched model highlights the value of incorporating broader biological context for accurate prediction. External validation also showed strong agreement with published clinical literature, supporting the clinical relevance of the model beyond the training dataset.

Importantly, extending the framework to polypharmacy increases its practical value in real-world prescribing. By evaluating multi-drug combinations, the model can identify high-risk regimens and highlight the specific drugs that contribute most to interaction risk. This bridges computational prediction and clinical decision-making, supporting safer prescription design and more informed polypharmacy management. Overall, the framework offers a scalable and clinically meaningful approach for DDI prediction and drug safety analysis.

Conclusion

The proposed framework integrates heterogeneous biological data with a Positive–Unlabeled learning strategy to improve drug–drug interaction prediction by addressing unreliable negative samples. By capturing multi-level relationships among drugs and their biological context, the model demonstrates improved predictive performance and strong agreement with clinically reported interactions. The framework further extends to polypharmacy by enabling risk assessment across multi-drug combinations. These findings highlight its potential as a clinically relevant tool for interaction analysis and safer decision-making in multi-drug treatment settings.

Bibliography

1. Alhumaidi RM., *et al*. "Risk of Polypharmacy and Its Outcome in Terms of Drug Interaction in an Elderly Population: A Retrospective Cross-Sectional Study". *Journal of Clinical Medicine* 12.12 (2023): 3960.

2. Mohamed M R., *et al*. "Association of polypharmacy and potential drug-drug interactions with adverse treatment outcomes in older adults with advanced cancer". *Cancer* 129.7 (2023): 1096-1104.

3. Rasool M F., *et al*. "Assessment of risk factors associated with potential drug-drug interactions among patients suffering from chronic disorders". *PloS One* 18.1 (2023): e0276277.

4. Peters LB., *et al*. "Evaluating drug-drug interaction information in NDF-RT and DrugBank". *Journal of Biomedical Semantics* 6 (2015): 19.

5. Yan Zhao., *et al*. "Drug-drug interaction prediction: databases, web servers and computational models". *Briefings in Bioinformatics* 25.1 (2024).

6. Ferdousi R., *et al*. "Computational prediction of drug-drug interactions based on drugs functional similarities". *Journal of Biomedical Informatics* 70 (2017): 54-64.

7. Mei S and Zhang K. "A machine learning framework for predicting drug-drug interactions". *Scientific Report* 11 (2021): 17619.

8. Lin Xuan., *et al*. "KGNN: Knowledge Graph Neural Network for Drug-Drug Interaction Prediction" (2020).

9. Abbas K., *et al*. "Graph neural network-based drug-drug interaction prediction". *Scientific Report* 15 (2025): 30340.

10. Hu EY., *et al*. "Enhancing link prediction in biomedical knowledge graphs with BioPathNet". *Nature Biomedical Engineering* (2026).

11. Zheng Y., *et al*. "DDI-PULearn: a positive-unlabeled learning method for large-scale prediction of drug-drug interactions". *BMC Bioinformatics* 20.19 (2019): 661.

12. Vilar S., *et al*. "Similarity-based modeling in large-scale prediction of drug-drug interactions". *Nature Protocols* 9.9 (2014): 2147-2163.

13. Rao A D., *et al*. "Is the Combination of Sulfonylureas and Metformin Associated With an Increased Risk of Cardiovascular Disease or All-Cause Mortality?" *Diabetes Care* 31.8 (2008): 1672-1678.

14. Hougen I., *et al*. "Safety of add-on sulfonylurea therapy in patients with type 2 diabetes using metformin: a population-based real-world study". *BMJ Open Diabetes Research and Care* 9.2 (2021): e002352.

15. Yu T., *et al*. "Risk of acute renal failure associated with combined use of SGLT2 inhibitors and potentially nephrotoxic drugs: an epidemiological surveillance study based on the FDA adverse event reporting system (FAERS)". *Expert Opinion on Drug Safety* (2025): 1-12.

16. Singh B., *et al*. "ACE inhibitors. StatPearls - NCBI Bookshelf (2025).

17. Antonazzo I C., *et al*. "Myopathy with DPP-4 inhibitors and statins in the real world: investigating the likelihood of drug-drug interactions through the FDA adverse event reporting system". *Acta Diabetologica* 57.1 (2019): 71-80.

18. Tian B., *et al*. "Efficacy and safety of combination therapy with sodium-glucose cotransporter 2 inhibitors and renin-angiotensin system blockers in patients with type 2 diabetes: a systematic review and meta-analysis". *Nephrology Dialysis Transplantation* 37.4 (2021): 720-729.

19. Dawra VK., *et al*. "Assessment of the drug interaction potential of ertugliflozin with sitagliptin, metformin, glimepiride, or simvastatin in healthy subjects". *Clinical Pharmacology in Drug Development* 8.3 (2018b): 314-325.