Graph Neural Networks and Molecular Data Integration for Drug Repurposing in Neuropsychiatric Disorders

B. Patel[1], P. Lam[2], A. Khan[3], D. Panwala[4], E. Kim[5], N. Liang[6], P. Kavandal[7], S. Shoaib Hasnain[8], S. Aul[9], S. Yoele[10], T. Narynov[11], Z. Afroz[12]

[1]Seattle University, Seattle, CA [2]Sunset High School, Portland, OR [3]University of Houston, Houston, TX [4]Purdue University, West Lafayette, IN [5]Centreville High School, Clifton, VA [6]Pennsylvania State University, Centre County, PA [7]Vista Del Lago High School, CA [8]University of California, Berkeley, CA [9]Columbia University, New York, NY [10]West Ranch High School, CA [11]High Bluff Academy, San Diego, CA [12]Northgate High School, Walnut Creek, CA

ePoster Presentation • *Cognitive & Computational Neuropharmacology*

Introduction

Bringing a new drug to market takes around 13 years and $1 billion, and the failure rate is especially high for neuropsychiatric disorders, where symptoms overlap and many genes and pathways are involved at once. Drug repurposing is a much faster route, but classical methods miss the more subtle links between drugs, genes, and disease. Graph neural networks (GNNs) operate directly on biological networks, so they fit this problem in a way that similarity- or matrix-based methods do not.

Objective

We reviewed how GNNs, paired with molecular data, are being used to predict drug repurposing candidates for neuropsychiatric disorders, and what their current limits are.

Methods

We ran a bibliometric review on the **Web of Science Core Collection**.

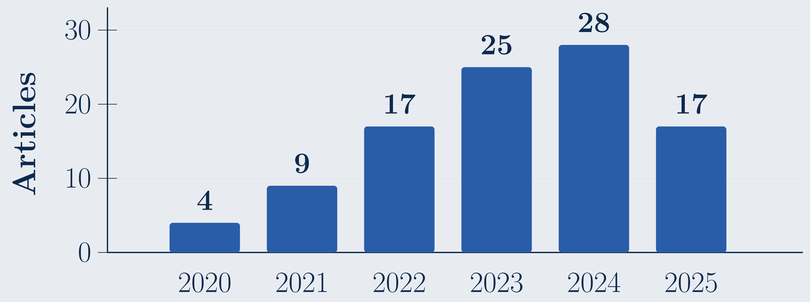
Search: "graph neural network," "drug repurposing," "neuropsychiatric disorders."

Pipeline:

- Initial query returned **123 records**
- Kept articles published **after 2020**
- Took the **top 100 by citation count**
- Coded each paper in Excel for model architecture, molecular inputs, and reported findings

Publication Trends (2020–2025)

Top-100 cited articles by year



Output in this area grew roughly **7×** from 2020 to 2024. The 2025 value only covers the first half of the year.

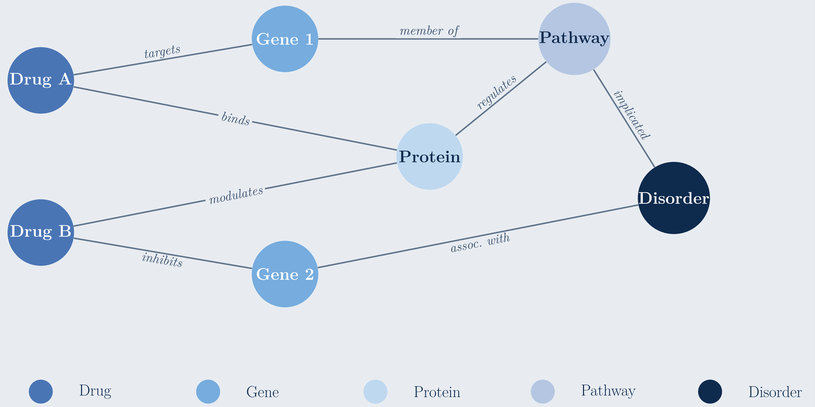
Key Models Identified

TxGNN (Huang et al., 2024) — zero-shot prediction over a multi-relational biomedical knowledge graph, with multi-hop pathway explanations.

GDRNet (Doshi & Chepuri, 2022) — combines gene expression and phenotype data; first benchmarked on COVID-19 repurposing.

Other recurring architectures: **GCN**, **GAT**, **R-GCN**, and heterogeneous message-passing variants applied to drug–target, drug–disease, and gene–disease prediction.

Heterogeneous Knowledge Graph



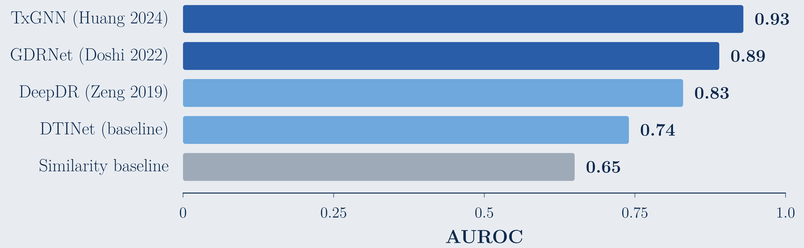
Each node type (drug, gene, protein, pathway, disorder) and each relation has its own embedding. The model learns these by passing messages between connected nodes, so a predicted drug–disease link reflects multi-hop evidence across the graph rather than a single similarity score.

Discussion

GNNs handle the heterogeneity of neuropsychiatric disorders better than earlier approaches, but the field still runs into the same three problems across studies: **interpretability** (why a prediction was made), **generalizability** (models don't transfer well between datasets), and **data sparsity** (many disorders have very few labelled examples). Better explainer methods and larger, multimodal datasets are the obvious next step.

Results: Benchmark Performance

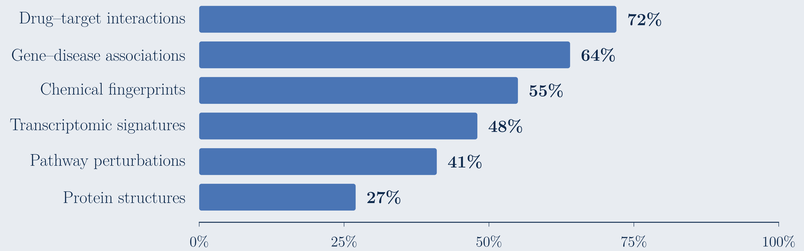
AUROC on drug–disease prediction



GNN-based models beat the older similarity and matrix-factorization baselines across the studies we reviewed. **TxGNN** (Huang et al., 2024) reported a **+49.2%** relative gain on indication prediction in a zero-shot setting.

Molecular Data Modalities

Frequency across reviewed studies (n = 100)



Drug–target and gene–disease links appear in most of the studies. Transcriptomic data and chemical fingerprints are showing up more often, especially in work focused on neuropsychiatric conditions.

Conclusion

Pairing GNNs with molecular data is a clear step up over previous repurposing methods: better accuracy on benchmarks, some level of explanation for each prediction, and useful candidates for disorders where there is not much else to work with. The main bottleneck going forward is data quality and interpretability, not the models themselves.

Selected References

Huang, K., et al. (2024). A foundation model for clinician-centered drug repurposing. *Nature Medicine.*

Doshi, S., & Chepuri, S. P. (2022). A computational approach to drug repurposing using graph neural networks. *Computers in Biology and Medicine.*

Bonner, S., et al. (2022). A review of biomedical knowledge graph reasoning for drug discovery. *Briefings in Bioinformatics.*

Gysi, D. M., et al. (2021). Network medicine framework for identifying drug repurposing opportunities. *PNAS.*