

# A Survey on AI-Powered Drug Repurposing using Graph Neural Networks

Alan Antony  
Department of Computer Science and  
Engineering  
FISAT, Angamaly, India

Alen Chacko  
Department of Computer Science and  
Engineering  
FISAT, Angamaly, India

Ms. Preethi NP  
Department of Computer Science and  
Engineering  
FISAT, Angamaly, India

Alen Bigi  
Department of Computer Science and Engineering  
FISAT, Angamaly, India

Anju Augustine  
Department of Computer Science and Engineering  
FISAT, Angamaly, India

**Abstract** - Drug repurposing has emerged as a cost-effective alternative to traditional drug discovery, particularly for rare and orphan diseases. With the rapid growth of biomedical data, artificial intelligence (AI) techniques have gained prominence in identifying novel drug - disease associations. Among these, Graph Neural Networks (GNNs) have shown remarkable potential due to their ability to model complex biological relationships. This survey presents a comprehensive review of AI-powered drug repurposing approaches, with a particular focus on graph- based and deep learning methods.

## General Terms

Drug Repurposing, Graph Neural Networks, Artificial Intelligence

**Keywords** - Biomedical Graphs, Deep Learning

## 1. INTRODUCTION

Drug discovery is a lengthy, expensive, and high-risk process, often requiring more than ten years and billions of dollars to develop a single successful drug. This challenge is amplified for rare and orphan diseases, where limited patient populations and low commercial incentives discourage pharmaceutical investment. As a result, a significant proportion of rare diseases lack approved therapeutic options.

Drug repurposing aims to identify new clinical indications for existing drugs, thereby reducing development time, cost, and safety risks. With the rapid growth of biomedical data, artificial intelligence (AI) has emerged as a powerful tool to automate and scale drug repurposing pipelines. Machine learning and deep learning models can analyze large volumes of molecular, genomic, and clinical data to uncover hidden drug-disease relationships.

Recently, Graph Neural Networks (GNNs) have gained prominence due to their ability to operate on graph-structured data. Biological systems are inherently graph-based, involving complex networks of

interactions between drugs, targets, genes, and diseases. This survey focuses on AI-powered drug repurposing methods with a particular emphasis on graph- based learning techniques.

## 2. BACKGROUND AND MOTIVATION

Biomedical data has grown exponentially due to advances in high-throughput sequencing, electronic health records, and large-scale molecular profiling initiatives. These heterogeneous data sources capture complex relationships between drugs, genes, proteins, pathways, and diseases. However, extracting meaningful insights from such interconnected data remains challenging using conventional machine learning techniques.

Graph-based representations naturally model these interactions, where nodes represent biological entities and edges encode relationships such as binding, regulation, or association. Graph Neural Networks (GNNs) extend traditional neural networks to graph-structured data, enabling effective learning from both node features and graph topology. This capability makes GNNs particularly suitable for drug repurposing tasks, where therapeutic effects often arise from indirect or multi-hop biological interactions.

The motivation for this survey is threefold: (i) to provide a structured overview of AI-driven drug repurposing methods, (ii) to analyze how GNNs outperform traditional approaches in relational learning, and (iii) to identify research gaps related to scalability, interpretability, and clinical translation.

The motivation for this survey is to consolidate existing research on AI-powered drug repurposing, highlight the role of GNNs, and identify research gaps.

## 3. DATA SOURCES AND GRAPH CONSTRUCTION

### 3.1 Biomedical Data Sources

AI-driven drug repurposing relies fundamentally on the availability of high-quality and well-curated biomedical datasets that

describe drugs, targets, diseases, and their interrelationships. Drug-related information is primarily obtained from publicly available chemical and pharmacological databases such as DrugBank and ChEMBL. These resources provide structured data on approved and experimental drugs, including canonical SMILES representations, molecular properties, bioactivity measurements, and known drug–target interactions. Such information is critical for learning meaningful chemical representations and for supervising drug–target interaction prediction models.

Disease–gene association data are sourced from repositories such as DisGeNET and Orphanet. DisGeNET aggregates gene–disease associations from curated databases, genome-wide association studies, and scientific literature, offering broad coverage across common and rare diseases. Orphanet, in particular, focuses on rare and orphan diseases and provides expert-curated disease definitions along with associated genes and external cross-references. These associations enable the mapping of disease phenotypes to their underlying molecular mechanisms.

Protein-centric information is obtained from databases such as UniProt and STRING. UniProt provides stable protein identifiers and functional annotations, which are essential for unifying data from multiple sources. STRING and BioGRID offer protein–protein interaction (PPI) networks that describe functional and physical interactions between proteins, allowing the modeling of biological pathways and molecular complexes involved in disease progression.

In advanced repurposing pipelines, transcriptomic datasets from GEO and LINCS L1000 are often incorporated to capture gene expression changes associated with disease states or drug perturbations. Although optional in this work, such datasets can further enhance model performance by integrating functional cellular responses into the learning framework.

### 3.1.1 Graph Construction Strategies

Graph-based drug repurposing approaches represent biomedical knowledge as interconnected graphs, where entities are modeled as nodes and their relationships as edges. This representation allows complex biological systems to be analyzed using graph neural networks (GNNs), which are capable of learning both local and global structural patterns.

At the molecular level, drugs are represented as molecular graphs, where atoms correspond to nodes and chemical bonds correspond to edges. Each atom node is associated with feature vectors encoding physicochemical properties such as atomic number, valence, aromaticity, and hybridization state. Bond features capture bond type, conjugation, and ring membership. These molecular graphs preserve the structural and chemical context of compounds, enabling GNNs to learn expressive drug embeddings directly from chemical structure.

At the biological level, heterogeneous graphs are constructed by integrating multiple entity types, including drugs, proteins, genes, and diseases.

Node features in the heterogeneous graph may include molecular embeddings for drugs, learnable embedding vectors for proteins, and identifier-based representations for diseases. Protein nodes are indexed

and embedded using trainable embedding layers, enabling the model to learn functional protein representations based on observed interaction patterns. Disease nodes are connected to proteins through curated gene–protein mappings derived from Orphanet, thereby linking disease phenotypes to molecular targets.

By combining molecular graphs with higher-level biomedical interaction networks, the resulting graph representation captures both fine-grained chemical structure and system-level biological context. This unified representation enables graph neural networks to model complex drug–target–disease relationships and supports the identification of potential repurposing candidates for orphan diseases.

**Table 1 : Advantages of Graph Neural Networks for Drug Repurposing**

Feature	Benefit
Graph structure	Natural representation of biomedical data
Message passing	Captures multi-hop biological interactions
Heterogeneous graphs	Integrates drugs, genes, and diseases
Explainability tools	Subgraph-level interpretation
Scalability	Suitable for large biomedical networks

## 4. COMPREHENSIVE SURVEY OF EXISTING LITERATURE

This section presents an in-depth survey of ten representative research works in AI-powered drug repurposing. Each study is analyzed based on its methodology, data sources, key contributions, and limitations, illustrating the evolution of computational approaches in this domain.

### 4.1.1 Singh (2024): AI for Drug Repurposing Against Infectious Diseases

Singh (2024) proposed a multi-stage artificial intelligence framework aimed at accelerating drug repurposing during infectious disease outbreaks such as COVID-19 and Ebola. The framework begins with large-scale biomedical literature mining using natural language processing techniques, including named entity recognition and relation extraction, to identify candidate drug–target–disease relationships from sources such as PubMed and COVID-19. These candidates are refined using deep learning models, primarily convolutional and recurrent neural networks, to predict drug–target interactions based on molecular fingerprints and protein sequence features.

In the final stage, shortlisted candidates undergo molecular docking and binding affinity estimation to validate structural

feasibility. The key contribution of this work lies in demonstrating how AI-driven pipelines can significantly reduce hypothesis generation time during public health emergencies. However, the approach relies heavily on curated literature and high computational resources, limiting scalability and explainability.

#### 4.1.2 Prasad and Kumar (2021): AI-driven Drug Repurposing for SARS-CoV-2

Prasad and Kumar (2021) developed a hybrid AI-based drug repurposing pipeline targeting SARS-CoV-2 viral proteins. The methodology integrates deep learning models trained on molecular descriptors with structure-based validation using molecular docking and molecular dynamics simulations. Convolutional neural networks extract spatial molecular features, while recurrent neural networks capture sequential patterns from SMILES representations.

The framework achieved biologically meaningful predictions and demonstrated improved confidence through structural validation. However, its dependence on high-quality protein structures and intensive computation limits generalizability and scalability beyond viral diseases.

#### 4.1.3 Jin and Wong (2014): Toward Better Drug Repositioning

Jin and Wong (2014) presented a foundational review categorizing drug repositioning strategies into phenotypic screening, target-based approaches, molecular docking, and network-based techniques. The study emphasized the advantages of integrative computational frameworks and discussed the limitations of isolated methodologies.

Although the work does not introduce a predictive model, it provides a conceptual foundation that influenced subsequent research, particularly the development of network-based and graph-driven repurposing approaches.

#### 4.1.4 Pan et al. (2023): AI-DrugNet

Pan et al. (2023) introduced AI-DrugNet, a graph neural network framework for predicting drug-disease associations, with a focus on neurological disorders. The model constructs a heterogeneous biomedical graph integrating drug-target, disease-gene, and protein-protein interaction data. Graph convolutional networks are used to learn node embeddings that capture both topological structure and biological context.

The results demonstrate superior performance over traditional machine learning methods. However, the framework requires extensive curated interaction data and faces scalability challenges when applied to large biomedical networks.

#### 4.1.5 Iorio et al. (2013): Connectivity Map

Iorio et al. (2013) pioneered transcriptomics-based drug repurposing through the Connectivity Map approach. The

methodology compares disease-induced gene expression signatures with drug-induced transcriptional profiles to identify compounds capable of reversing disease states. Drugs exhibiting strong negative correlation with disease signatures are prioritized as repurposing candidates.

The primary advantage of this approach is its interpretability and independence from prior drug-target knowledge. However, it does not incorporate molecular structure or protein interaction networks, limiting its predictive scope.

#### 4.1.6 Amiri et al. (2023): IDDI-DNN

Amiri et al. (2023) proposed IDDI-DNN, a hybrid framework combining Similarity Network Fusion with convolutional neural networks to predict drug-disease associations. Multiple similarity matrices derived from chemical and biological data are fused into a unified representation and processed by a CNN classifier.

While the model achieves high predictive accuracy, it treats similarity matrices as images, resulting in loss of explicit relational structure and reduced biological interpretability compared to graph-based approaches.

#### 4.1.7 Wen et al. (2021): Clinical Connectivity Map

Wen et al. (2021) developed a clinical connectivity map using real-world electronic health record laboratory data. Drug and disease signatures were constructed from longitudinal laboratory measurements, and complementarity-based scoring was applied to identify repurposable drugs.

This approach offers strong clinical relevance and interpretability; however, it is constrained by data sparsity, confounding clinical factors, and limited availability of high-quality EHR datasets.

#### 4.1.8 Kulkarni (2023): Review on Computational Drug Repurposing

Kulkarni (2023) presented a comprehensive survey of computational drug repurposing techniques, covering similarity-based methods, network-based approaches, and deep learning models. The review benchmarked representative studies and highlighted challenges such as data heterogeneity and lack of standardized evaluation protocols.

Although the study does not propose new predictive models, it provides valuable insights into methodological trends and open research problems.

#### 4.1.9 Wang (2022): DrugRepo

Wang (2022) introduced DrugRepo, a scoring-based framework that integrates chemical similarity, target overlap, and protein-protein interaction distance to rank drug-disease associations. The framework emphasizes scalability and

interpretability, enabling large-scale screening.

However, its reliance on linear scoring limits the ability to capture complex nonlinear biological interactions, motivating the adoption of learning-based graph models.

#### 4.1.10 Firoozbakht et al. (2021): Network-based Repurposing

Firoozbakht et al. (2021) proposed a network-based drug repurposing approach for breast cancer subtypes. The framework integrates gene expression data with protein interaction networks to identify drugs capable of reversing subtype-specific molecular signatures.

The study demonstrated biologically meaningful predictions but requires extensive omics data and lacks large-scale clinical validation, limiting broader applicability.

or phenotypic similarities. Chemical similarity was often computed using molecular fingerprints such as ECFP or MACCS keys, while biological similarity relied on target overlap or pathway co-membership. These similarity measures were then used to infer potential drug-disease associations through nearest-neighbor or scoring-based techniques.

Traditional machine learning models, including support vector machines (SVMs), random forests, and logistic regression, were later employed to improve predictive accuracy. These models treated drug repurposing as a binary classification or ranking problem using handcrafted features derived from similarity matrices. Although such approaches were computationally efficient and interpretable, they suffered from several limitations. In particular, they struggled to generalize to unseen drugs or diseases, failed to capture nonlinear biological interactions, and were highly sensitive to feature engineering and data imbalance.

**Table 2: Comparison of Drug Repurposing Approaches**

Approach Type	Strengths	Limitations	Representative Works
Similarity-Based	Simple, interpretable	Ignores complex relations	Wang (2022)
Transcriptomics-Based	Biologically meaningful	Limited to gene expression	Iorio et al. (2013)
Deep Learning (CNN/RNN)	Learns complex features	Black-box, no topology	Prasad (2021)
Network-Based	Captures indirect effects	Data-dependent	Firoozbakht (2021)
Graph Neural Networks	Relational learning, scalable	Computationally expensive	Pan et al. (2023)

## 5. AI AND GNN-BASED DRUG

### 5.1 Similarity-Based and Machine Learning Approaches

Early computational drug repurposing methods primarily relied on similarity-based inference, where drugs were compared based on chemical structure, shared targets, side-effect profiles,

### 5.2 Deep Learning Models

The advent of deep learning introduced more expressive models capable of automatically learning latent representations from raw biomedical data. Convolutional neural networks (CNNs) were applied to molecular graphs and descriptor matrices to capture spatial and structural features, while recurrent neural networks (RNNs) and long short-term memory (LSTM) networks were used to model sequential representations such as SMILES strings. Autoencoders and variational autoencoders further enabled unsupervised feature learning and dimensionality reduction.

Deep learning models demonstrated improved performance over traditional machine learning approaches, particularly in large-scale datasets. However, most deep models operated on vectorized or grid-like representations and ignored the explicit relational structure inherent in biological systems. As a result, these models lacked biological interpretability and were unable to reason over multi-hop interactions between drugs, genes, proteins, and diseases.

### 5.3 Graph Neural Networks

Graph Neural Networks (GNNs) represent a paradigm shift in AI-based drug repurposing by explicitly modeling biomedical systems as graphs. In such graphs, nodes correspond to entities such as drugs, proteins, genes, or diseases, while edges represent interactions including binding, regulation, or association. GNNs learn node embeddings through iterative message-passing mechanisms, allowing information to propagate across local and global neighborhoods.

Popular architectures such as Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), and heterogeneous GNNs have been successfully applied to drug repurposing tasks. These models are particularly effective in capturing multi-hop biological relationships, integrating heterogeneous data sources, and providing a unified framework

for relational learning. Recent studies consistently demonstrate that GNN-based approaches outperform similarity-based and deep learning models in predicting novel drug–disease associations. Moreover, the integration of explainability techniques, such as attention mechanisms and subgraph extraction, enhances interpretability and supports clinical trust.

## 6. EVALUATION METRICS AND BENCHMARKING

Evaluation of drug repurposing models is typically framed as a link prediction or ranking problem. Due to class imbalance and the absence of true negative samples, careful metric selection is critical. Widely used metrics include Area Under the Receiver Operating Characteristic Curve (AUC), Area Under the Precision–Recall Curve (AUPR), Precision, Recall and Hit Rate.

**Table 3 : Evaluation Metrics**

Metric	Description
Accuracy	Overall correctness
Precision	Correct positive predictions
Recall	Sensitivity to positives
AUC	Ranking capability

Benchmark datasets are commonly derived from curated repositories such as DrugBank and CTD. Cross-validation and temporal split strategies are employed to evaluate generalization. However, lack of standardized benchmarks remains a major limitation in the field.

## 7. CHALLENGES AND OPEN ISSUES

Despite strong progress in AI-powered drug repurposing, several open challenges remain.

Data quality and sparsity is a primary concern, as biomedical databases are often incomplete, noisy, and biased toward well-studied drugs and diseases. The lack of reliable negative drug–disease associations further complicates supervised learning and leads to overly optimistic performance estimates during evaluation.

Scalability poses another significant challenge, since large heterogeneous biomedical graphs may contain millions of nodes and edges. Training deep Graph Neural Networks on such large-scale graphs requires substantial computational resources, efficient sampling strategies, and memory optimization techniques. This limits the applicability of complex models in real-world clinical settings.

Interpretability and explainability are critical issues for the adoption of AI-based drug repurposing methods. Many deep learning and GNN-based models operate as black boxes, making it difficult for clinicians and researchers to understand the rationale behind predictions. Although recent explainable AI techniques such as attention mechanisms and subgraph-based explanations have been proposed, they are not yet standardized or widely validated.

Another important challenge is evaluation inconsistency. Existing studies often use different datasets, validation strategies, and performance metrics, making direct comparison across models difficult. The absence of standardized benchmarks and gold-standard datasets remains a major barrier to fair assessment.

Finally, clinical translation and validation remain limited. Most AI-driven drug repurposing studies rely on retrospective data and computational validation, with few predictions progressing to experimental studies or clinical trials. Bridging

the gap between computational predictions and real-world therapeutic applications requires closer collaboration between computational scientists, biologists, and clinicians.

## 8. FUTURE RESEARCH DIRECTIONS

Future research in AI-powered drug repurposing is expected to move toward multimodal and multi-scale learning, integrating molecular structure, gene expression, protein interaction networks, clinical records, and biomedical literature. Such holistic representations can improve robustness and biological relevance.

The integration of large language models (LLMs) with GNNs is an emerging direction, enabling automated knowledge extraction from scientific literature and enhanced reasoning over biomedical graphs. Explainable AI will play a key role in clinical adoption, necessitating transparent and faithful interpretation mechanisms.

Federated and privacy-preserving learning frameworks may enable cross-institutional collaboration without sharing sensitive patient data. Additionally, closer integration with experimental validation and clinical trials will be essential to translate computational predictions into real-world therapies.

## 9. CONCLUSION

This survey presented an extensive review of AI-powered drug repurposing approaches, with a strong emphasis on graph-based learning techniques. By analyzing ten representative studies, we highlighted methodological trends, strengths, and limitations across traditional machine learning, deep learning, and Graph Neural Network models. GNN-based approaches demonstrate superior capability in modeling complex biomedical relationships and integrating heterogeneous data sources.

Despite promising results, challenges related to data quality,

scalability, interpretability, and clinical validation remain open. Addressing these issues will be critical for translating computational predictions into real-world therapeutic outcomes. Future research integrating multimodal data, explainable AI, and large-scale validation holds significant promise for accelerating drug discovery and improving global healthcare accessibility.

## 10. REFERENCES

- [1] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. Yu, "A Comprehensive Survey on Graph Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.
- [2] X. Pan, Y. Liu, J. Guan, and S. Zhou, "AI-DrugNet: A Graph Neural Network Framework for Drug–Disease Association Prediction," *IEEE Journal of Biomedical and Health Informatics*, 2023.
- [3] A. Singh, "Artificial Intelligence for Drug Repurposing Against Infectious Diseases," *Briefings in Bioinformatics*, 2024.
- [4] K. Prasad and R. Kumar, "AI-driven Drug Repurposing for SARS-CoV- 2 Using Deep Learning and Molecular Docking," *Journal of Biomedical Informatics*, vol. 115, 2021.
- [5] G. Jin and S. T. C. Wong, "Toward Better Drug Repositioning: Prioritizing and Integrating Existing Approaches," *Drug Discovery Today*, vol. 19, no. 5, pp. 637–644, 2014.
- [6] F. Iorio, R. Tagliaferri, and D. di Bernardo, "Identifying Network of Drug Mode of Action by Gene Expression Profiling," *Nature Methods*, vol. 6, pp. 761–767, 2009.
- [7] M. Amiri, H. R. Rabiee, and M. Jalili, "IDDI-DNN: A Deep Learning Framework for Drug–Disease Interaction Prediction," *IEEE Access*, vol. 11, pp. 24561–24572, 2023.
- [8] Q. Wen, Z. Wang, and Y. Li, "A Clinical Connectivity Map for Drug Re- purposing Using Electronic Health Records," *Nature Communications*, vol. 12, 2021.
- [9] V. Kulkarni, "A Review on Computational Drug Repurposing Approaches," *Computers in Biology and Medicine*, vol. 155, 2023.
- [10] H. Wang, "DrugRepo: A Scalable Scoring Framework for Drug Repurposing," *Bioinformatics*, vol. 38, no. 4, pp. 1031–1039, 2022.
- [11] M. Firoozbakht, S. Ahmed, and J. Loscalzo, "Network-based Drug Re- purposing for Breast Cancer Subtypes," *PLOS Computational Biology*, vol. 17, no. 6, 2021.
- [12] M. Zhang, Z. Cui, M. Neumann, and Y. Chen, "An End-to-End Deep Learning Architecture for Graph Classification," *AAAI Conference on Artificial Intelligence*, 2018.
- [13] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "GNNExplainer: Generating Explanations for Graph Neural Networks," *NeurIPS*, 2019.
- [14] Y. Lin et al., "Drug Repurposing for COVID-19 Using AI and Network- based Methods," *IEEE Access*, vol. 9, pp. 123047–123057, 2021.
- [15] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling Polypharmacy Side Effects with Graph Convolutional Networks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, 2018.