Survey Paper on Various Techniques of Drug Discovery by Graph Neural Networks

Rounak Saha
Computer Science & Engineering
Institute of Engineering and Management (IEM), University of Engineering and Management Kolkata
Kolkata, India
rounak.saha@iem.edu.in

Trishita Maity
Computer Science & Engineering in Artificial Intelligence
Institute of Engineering and Management (IEM), University of Engineering and Management Kolkata
Kolkata, India
trishitamaity2003@gmail.com

Akanksha Singh
Computer Science & Engineering in Artificial Intelligence
Institute of Engineering and Management (IEM), University of Engineering and Management Kolkata
Kolkata, India
akankshasingh030703@gmail.com

Gracy Kumari
Computer Science & Engineering
Institute of Engineering and Management (IEM), University of Engineering and Management Kolkata
Kolkata, India
gracykumari14@gmail.com

Arnab Chakraborty
Computer Science & Engineering in Internet of Things
Institute of Engineering and Management (IEM), University of Engineering and Management Kolkata
Kolkata, India
arnabchakraborty.sam@gmail.com

2024 4th International Conference on Computer, Communication, Control & Information Technology (C3IT) | 979-8-3503-8912-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/C3IT60531.2024.10829467

***Abstract* – This survey report identifies and evaluates various approaches that could contribute to success in drug development and tracking activities using Graph Neural Networks (GNN). This paper aims to determine the best drug discovery approach among several related papers to increase the efficiency and the quality of the product. GNNs were incorporated into all levels of the drug discovery process, from target molecule discovery to de novo drug design. GNNs have significantly advanced efficiency and reliability in these processes, though challenges like biased datasets persist, creating room for improvement of the existing GNN structures. It discusses the curation and categorization of existing literature that has helped us grasp the concept of GNNs in drug discovery processes. This paper presents a detailed analysis and results of various applications of GNNs in reducing health risks, refining drug formulations, and drug discovery. A collective comparison of theories from distinguished research papers is proposed, involving discussions and literature surveys on the implementation of the GNN approach to date.**

Keywords– Drug discovery, Graph Neural Networks, De Novo Drug Design, Biased datasets, Health risk reduction

I. INTRODUCTION

Since the 1870s, drug discovery has evolved significantly, with AI playing a crucial role in recent years. Drug-drug interactions are critical to investigate, as it's important to determine if the drugs being used are similar and if their interaction could cause future problems in the human body. Various methods have been developed to address this, including graph construction, matrix factorization, and sometimes combining or integrating different approaches. Interest in drug interactions has surpassed that of gene interactions. Over the years, various methods have been employed, with virtual screening becoming particularly popular. This method relies on the Drug-Target Affinity (DTA) principle, which assesses a drug's affinity for proteins. It's favored for its speed and ability to handle large datasets. Additionally, numerous deep-learning approaches have been proposed to predict and provide more accurate assessments of drug-drug interactions. To date, various experiments and methods have been applied to datasets of chemical compounds.

The cure of difficult diseases, such as cancer, has been facilitated by the use of polypharmacy, which in turn also increases unplanned drug-drug interactions (DDIs) which could be potential medical risks. Prediction of such DDI-related risks has been carried out using the Graph Neural Networks (GNN) approach which uses network embedding techniques to form drug-drug matrices and analyze the type of relationship. The GNN approach is beneficial in reducing medical risks, refining drug formulation, and discovering new applications for existing drugs.

The field of drug discovery has undergone a dramatic transformation with the rise of artificial intelligence, especially with the use of graph neural networks (GNNs), these advanced models are adept at molecular structures that it is difficult to analyze and process, and has led to advances in various fields of drug discovery. Another model, MolKGNN, addresses specific limitations of GNNs by incorporating chirality awareness and SE(3)-framework invariance. This allows for comprehensive and high-performance molecular positioning on datasets of target compounds. Another model, MG^2N^2, uses a modular architecture to sequence molecular graphs, achieves higher accuracy and outperforms benchmarks such as QM9 and ZINC GNN-SNGP. On the other hand, driving the confidence of cardiotoxicity prediction models improve by reducing incorrect predictions with overconfidence, thereby improving drug safety. A valuable analytical tool GNNs played a key role in the drug recycling efforts during the COVID-19 pandemic in 2010, where topology-optimized messaging neural networks showed promising results on COVID-19 datasets. The model is particularly effective in drug and target interaction tasks, using graph structures for feature extraction and proving its value in rapid drug discovery situations. Techniques for de novo automated drug synthesis have also been greatly improved with GNNs. Key improvements include:

1. Design of new molecules with desired properties and activities using both non-autoinhibitory and autoregressive mechanisms.

2. Analysis of chemical morphology and chemical activity by molecular scoring, which is crucial for successful drug design.
3. Optimizing generated molecules to enhance their properties, meeting specific criteria for biological activity and pharmacokinetics.
4. Integrating synthesis planning into the workflow to ensure feasibility in the laboratory.

Despite these advancements, challenges remain, such as biases in benchmark datasets and synthesis difficulties of generated molecules. Issues like 'over smoothing' in GNNs prompt further exploration into deeper GNN architectures. The review highlights the transformative role of GNNs in drug discovery, advocating for broader empirical studies and developing quantitative measures to assess distance preservation within learned representations.

Overall, the integration of GNNs in drug discovery has resulted in more reliable and efficient methodologies, significantly advancing the field. These models continue to evolve, promising further improvements and broader applications in the future.

II. LITERATURE SURVEY

[1] The MoLGNN model enhances drug discovery by pretraining with unlabeled chemical data, reducing latent space sparsity. It uses a GINVAE to integrate node features and motifs. To demonstrate the utility of MoLGNN a screening drug test occurs which mainly targets the Janus kinase. Validation includes screening drugs targeting JAK family members (JAK1, JAK2, JAK3) for COVID-19 treatment and effectively capturing substructure information for efficient drug discovery.

[2] In graph neural networks, drugs are depicted as nodes and their interactions as links, positioning the prediction of drug-drug interactions as a link prediction problem. Conditional random fields (CRF) are effective in GNN technology for structured prediction but have limited representation ability due to their linear model nature. Structured Prediction Energy Networks (SPENs) were introduced to expand the capabilities of structure prediction models. They employ neural networks to define an energy function, allowing optimization over a continuous relaxation of labels to capture high-arity interactions among output labels in graph-based data.

Figure 1: Warfarin increases bleeding risk when taken with antibiotics like Cefixime or Fluconazole, and NSAIDs such as Celecoxib or Naproxen

[3] In a comparative study of graph neural network architectures for drug repurposing, GDRnet approaches the task by framing it as a link prediction problem, focusing on representing drugs as nodes and drug-drug interactions as links within a graph neural network framework. It constructs a four-layered graph integrating diseases, genes, drugs, and anatomical structures, enabling analysis of both inter-layered and intra-layered connections. In this framework, connections between drugs and diseases signify potential treatment avenues. GDRnet comprises an encoder generating node embeddings and a decoder using a score function to predict drug-disease interactions. Evaluation involves ranking actual treatment drugs in prediction lists for diseases from testing sets based on score function outputs.

[4] Bi-level graph Neural Networks have been employed to predict biological connections, including drug-drug interactions. In this approach, the data is structured as a bi-level graph where the upper level represents the interaction graph and the lower level corresponds to the representation graph. The lower-level graph neural network expands the intrinsic representation of each biological entity in the dataset. Simultaneously, the upper-level graph neural network leverages the detailed representations from the lower level to examine interactions between these entities. This approach results in a fully integrated network that generates the final scores.

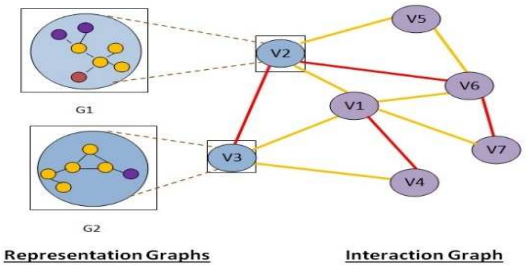

Figure 2: A bi-level graph illustrating drug-drug interactions, with representation graphs showing molecular element types and an interaction graph depicting drug interaction types

[5] MolKGNN represents a novel molecular representation model designed specifically for drug discovery, overcoming GNN limitations with chirality awareness and SE(3)- conformation invariance. It employs molecular graph convolution with learnable kernels to capture meaningful subgraph patterns and message aggregation for similarity propagation. MolKGNN excels on diverse drug target datasets, using interpretable kernels aligned with domain knowledge. Switching between convolution and message-passing layers, MolKGNN captures chemical patterns across multiple hops and excels in identifying active molecules, extending its applicability beyond drug discovery.

[6] MG^2N^2, a novel sequential molecular graph generator using Graph Neural Networks (GNNs), incrementally constructs molecular graphs node by node and their neighbors. It features a modular architecture optimizing information capacity and flexibility, employing BFS ordering and parallel edge generation for efficiency. MG^2N^2 surpasses benchmarks in unconditional molecule generation, accurately replicating original dataset characteristics. The GNN iteratively updates vertex states through message passing and aggregation, with task segmentation into three classification subtasks guided by an MLP. Future research should focus on broader generalization, conditional generation techniques, and enhancing performance on larger molecule datasets like Zinc.

[7] Rapid COVID-19 vaccine development underscores AI's role in drug discovery, especially in graph neural networks (GNNs). Using topology adaptive GNNs and message-passing neural networks, this method screens and repurposes drugs for COVID-19, leveraging graph structures for feature extraction. The GNN architecture combines TAGCN and MPNN, adapting its structure and message passing. This framework reviews drug-target interactions and introduces a novel method merging TAGCN and MPNN, showcasing GNNs' potential in COVID-19 drug discovery.

[8] The review explores automated drug design, emphasizing Graph Neural Networks' pivotal role in revolutionizing drug discovery. It explores GNNs' versatility and effectiveness in

molecular scoring, property prediction, and identifying drug-likeness and pharmacological activity. The text discusses their impact on antibiotic discovery, ADMET prediction, and recent advancements in computer-aided synthesis planning. Challenges such as biased datasets, synthesis difficulties, and issues like 'over-smoothing' in GNNs are addressed, underscoring the need for improved architectures and deeper exploration. Overall, the review provides comprehensive insights into automated drug design, emphasizing GNNs' transformative potential in this field.

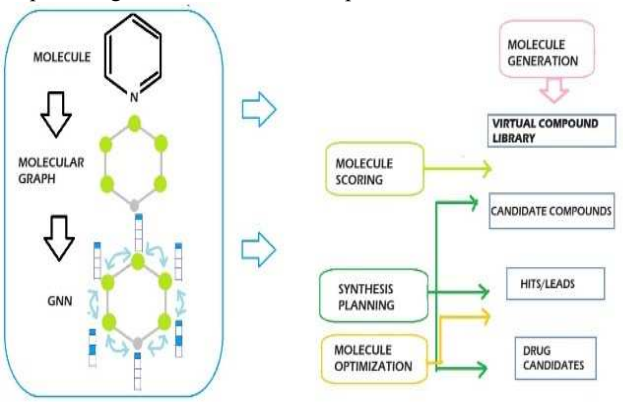

Figure 3: The applications of graph neural networks (GNNs) in all stages of automated de novo drug design

[9] Cardiotox provides a standard for drug cardiotoxicity studies, which is important for drug safety assessment. The introduced GNN-SNGP addresses overconfidence predictions and increases the reliability of prediction of cardiotoxicity. GNN-GP and GNN-SNGP combine distance-awareness techniques such as Gaussian process layers and Spectral Normalization to improve feature extraction and prevent drops. Detailed analysis of benchmarks shows improved accuracy, robustness and calibration, Drug-Research focusing on the practical value of the discovery pipeline highlights the effectiveness of GNN-GP and GNN-SNGP to refine predictions while maintaining accuracy and robustness, advocating for further research with observations and quantitative measures to improve drug discovery.

[10] The COVID-19 outbreak in 2019 spread to 213+ countries, prompting AI-based drug design using Graph Neural Networks (GNN). GNN models focused on covalent bonds and atomic interactions, utilizing element-specific graph representations for topology analysis. Multiphysical Molecular Graph Networks (MP-GNN) split data into head and tail sections: heads generated feature vectors, while tails assessed binding affinity. MP-GNN aided SARS-CoV-2 drug development by evaluating complexes and identifying effective binding particles like carbon and nitrogen, balancing cost and accuracy with techniques like ensemble learning, stacking, and boosting.

[11] Cancer remains incurable, causing significant mortality annually. Recent advancements use datasets, matrix factorization, and machine learning to identify cancer cell components. Multi-omic data processing aids cancer profiling, while graph-based methods analyze drug-cancer interactions. Genomic transcription helps develop new compounds. Graph CDR, utilizing Graph Neural Networks (GNN), predicts cancer lineage by forming nodes from specific data types. Representation learning integrates datasets, with cross-validation and independent testing enhancing accuracy. Graph CDR acts as a predictive model and regularizer in multitasking frameworks.

[12] High-performance computers greatly enhance drug design, especially through virtual screening. Drug Target Affinity (DTA) assessments expedite matching drugs with targets. Graph DTA, developed with PyTorch, is a key advancement. It evaluates drug affinity toward targets, often proteins, using flexible graphs with residues as nodes. Deep DTA manages large datasets of proteins and constants, describing drugs with SMILES notation and self-loops. Encoding proteins into vectors and selecting node features streamline representation. Separate models and graphs simplify implementation. GNNs combine training data from both domains for DTA prediction, with two layers enhancing predictability during testing.

[13] Precision medicine, tailored to individual lifestyles and genetic profiles, faces challenges in diseases like cancer due to the complex genotype-phenotype relationship. Past methodologies, such as matrix factorization, aimed to identify cells and drugs by reducing matrix levels. Accurate prediction and understanding of drug composition necessitate methodologies such as Message Passing Neural Networks and the integration of Graph Neural Networks into Target-Guided Structure-Activity models. TGSA integrates gene, drug, and cell line representations, enabling fine-tuning for better predictions. Challenges include limited resources and structural disparities between datasets, addressed by pre-training GNNs on large-scale molecule datasets and semantic matching strategies. Despite challenges, TGSA finds numerous clinical applications, enhancing drug precision in treatment.

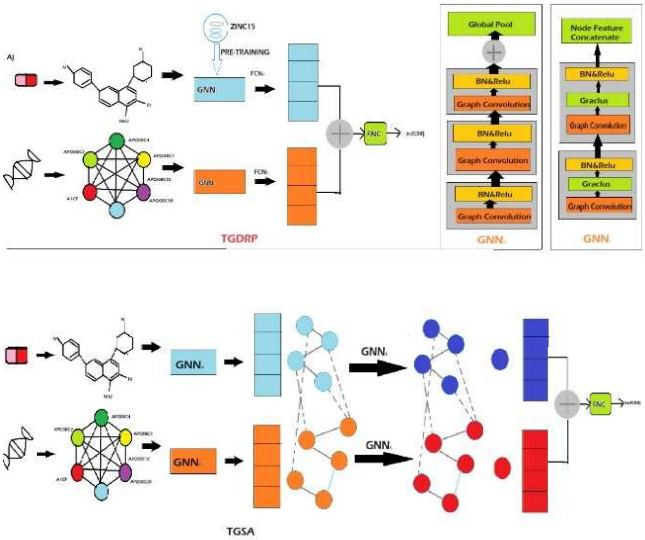

Figure 4: An overview of the TGSA framework

[14] Cancer and polypharmacy are common today, especially in the elderly, requiring careful treatment to avoid fatal drug interactions. Detecting drug-drug interactions is crucial but can be costly and tedious. There are four main approaches: similarity-based, matrix factorization, network analysis-based, and deep learning. The proposed method utilizes graph neural networks for making predictions. It involves data collection, forming a diverse network, and employing the deep learning model to predict interactions. This is achieved by reducing the dimensionality of the matrix and integrating embedded vectors. This method outperforms others, addressing dataset imbalances and improving DDI prediction accuracy.

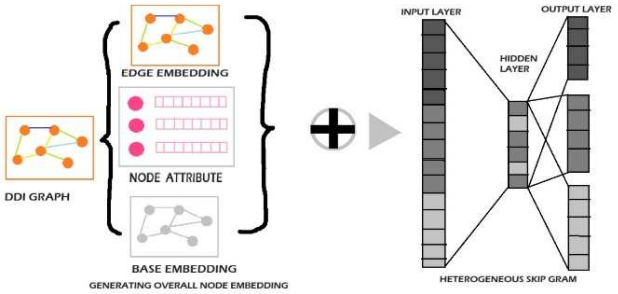

Figure 5: An overview of the embedding model

[15] The prediction of drug-drug interactions involves two tasks: identifying the interactions and determining their type. The k-fold cross-validation method partitions the dataset into k equally sized subsets, using each subset once as a test set while the other k−1 subsets are used for training in each iteration. Evaluation metrics like the Area Under the Precision-Recall Curve and the Area Under the ROC Curve are frequently used to evaluate the performance of machine learning models. This proposed method has been tested with various input data sources to verify the accuracy of the evaluation method.

[16] The DeepDDS model utilizes deep learning techniques to predict drug combination synergies. It analyzes drug interactions, uncovering potential novel synergies for pharmaceutical R&D. Initially, it converts drug chemical structures into molecular graphs via RDKit, using SMILES notation. These graphs depict atoms as vertices and bonds as edges. DeepDDS computes drug embeddings with a graph convolutional network, capturing essential drug characteristics. These embeddings encode structural information for deep learning analysis. By integrating genomic and pharmaceutical features, DeepDDS identifies synergistic drug combinations based on chemical structures and gene expression patterns. Overall, DeepDDS employs advanced graph-based deep learning to predict drug synergies accurately.

TABLE 1. COMPARISON TABLE

Author	Year	Approach	Description
Tengfei Ma et al.	2019	Structure prediction approach in Graph Neural Network	Introduces a Conditional Random Field approach for structure prediction in GNNs but enhances it with SPENs for better representation.
Mingjian Jiang et al.	2020	Predicting drug-target affinity using contact maps	Advocates for the use of computer applications in drug design, emphasizing drug affinity to accelerate development and promote advancements in deep learning and predictive modeling.
Mark Cheung et al.	2020	Graph Neural Networks for Discovering Drugs for COVID-19	Presents a GNN predicting SARS-CoV-2 inhibiting antibiotics, combining adaptive topology, message-passing, and graph convolution.
Yunsheng Bai et al.	2020	Hierarchical Graph Neural Networks	Proposed Bi-Level GNN which had an interaction graph and a representation graph to study and predict the drug-drug interactions.
Yiheng Zhu et al.	2021	Twin Graph Neural Network	Proposes TGDRP using GNN, demonstrating TGSA's effectiveness in precision medicine for cell lines.
Jiacheng Xiong et al.	2021	Automated de novo drug design	Explores GNNs in automated drug design, using GraphVAE, GraphNVP, and receptor-based scoring methods.
Pietro Bongini et al.	2021	Graph Neural Network for Molecular Generation	Introduces MG²N², a GNN-based molecular graph generator, emphasizing its efficiency, interpretability, and cost-saving potential.
Kehang Han et al.	2021	Robust Graph Neural Network	Introduces GNN-SNGP, a distance-aware GNN for drug discovery, addressing overconfident predictions and cardiotoxicity.
Xiaoke Shen et al.	2022	Self-Supervised Motif Learning	MoLGNN pretrains on unlabeled chemical data,

			reducing latent space sparsity. It integrates node features via GINVAE, excelling in drug property prediction and validated by JAK family COVID-19 drug tests.
Xiao-Shuang Li et al.	2022	Graph Neural Network for Multiphysical Applications	Using virtual screening enhances drug design and finds solutions for potential drugs and better treatments.
Jinxian Wang et al.	2022	Deep Graph Neural Network incorporating an attention mechanism	Employs gnn to transform drug structures into embeddings, allowing it to predict synergistic effects through the integration of genomic data.
Xuan Liu et al.	2022	Contrastive Learning.	Introduces graph CDR, predicting cancer cell responses to specific drugs using diverse cancer cell lines and drug datasets.
Mohammad Hussain Al Rabeah et al.	2022	Feature Attraction to predict drug-drug interaction	Proposed two main methods for predicting drug-drug interactions: gathering resources and combining vectors. This approach proved more efficient than others.
Mohammad Hussain Al_Rabeah	2022	method of K-fold cross-validation	Proposed using cross-validation with k folds to predict drug-drug interactions and evaluating models with AUPR, tested for accuracy with various data inputs.
Siddhant Doshi et al.	2022	GDRnet for drug repurposing.	Employed a four-layered graph to identify biological entity interactions, utilizing GDRnet with encoder and decoder components for prediction and analysis.
Yunchao (Lance) Liu et al.	2023	Graph Neural Network with Interpretable Chirality Awareness	Proposed MolKGNN, a SE(3)-equivariant Graph Neural Network with chirality awareness for drug discovery, emphasizing molecular kernel convolution and message-passing mechanisms.

BI-GNN [4] proposes a bi-level architecture containing two graphical representations for the presentation of 93% accuracy. Another approach, using chirality for 3D accounting and focusing on cardiotoxicity, reported accuracy of 91% [5]. The GNN-based protocol for drug design [8] involved GNNs in candidate molecule generation and scoring and reported an accuracy of 90.5%. The GNN-SNGP model [9] increases the awareness of distance to lessen overconfident mispredictions, providing accuracy of 90% and yielding the AUROC value between 0.894 to 0.935. For drug-drug interaction prediction, a GNN-based method [14] reaches the accuracy of 82.5%. Finally, GNN applied to COVID-19 drug discovery [10] reaches an ROC-AUC of 0.72 with 78% accuracy, thus showing prospects in urgent healthcare situations.

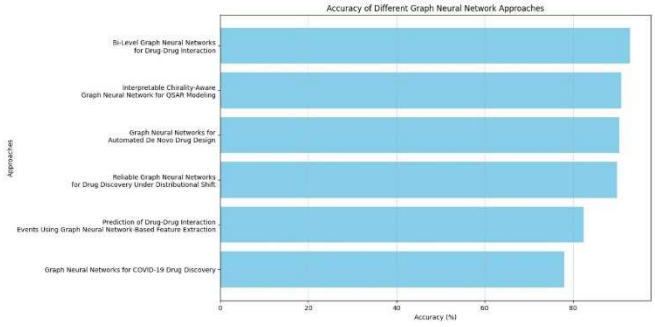

Figure 6: This graph is a useful tool for comparing the effectiveness of various GNN approaches in terms of their accuracy in drug discovery and prediction tasks.

III. NOVELTY

The precise study and understanding of the graph neural network approach provide accurate and varied information about drug-drug interactions and their effects. This paper presents a detailed analysis and results of various applications of GNNs in reducing health risks, refining drug formulations, and drug discovery. A collective comparison of theories from distinguished research papers is proposed, involving discussions and literature surveys on the implementation of the GNN approach to date. This paper offers readers a comprehensive overview of drug-drug interactions and a

concise understanding of how these interactions are utilized in advancing the discovery of new drug combinations. All procedures followed in utilizing the GNN approach have been extracted from numerous research papers and are mentioned in this paper.

IV. LIMITATION

This paper is an in-depth review based on eight of the 22 key papers available that have made significant contributions to the field of Graph Neural Networks (GNNs). These selected papers provide a solid research foundation, covering important work and it has an impact on the project. However, it is not possible to fully examine the broader field of GNN research by focusing only on these eight papers. Although the paper is comprehensive and integrates key concepts and successes, there is still untapped knowledge in other papers. These documents can provide new perspectives or address emerging trends that could significantly improve the industry. Although the study brings together important data, further research will be conducted in the near future. It is important to keep up to date with developments in order to stay at the forefront of GNN development.

V. CONCLUSION

This paper explores various drug discovery methods, highlighting Graph Neural Networks for forecasting drug-target interactions and automating drug design. GNNs have significantly advanced efficiency and reliability in these processes, though challenges like biased datasets persist. Future research should focus on refining GNN architectures to address these limitations and enhance their efficacy in drug discovery.

REFERENCES

[1] Y. L. Y. W. L. X. Xiaoke Shen, "MoLGNN: Self-Supervised Motif Learning Graph Neural Network for Drug Discovery," in Machine Learning for Molecules Workshop at NeurIPS, 2020.

[2] J. S. C. X. J. S. Tengfei Ma, "GENN: Predicting Correlated Drug-drug Interactions with Graph Energy Neural Networks," arXiv, pp. 1-2, A1, ii-xi, 2019.

[3] S. P. C. Siddhant Doshi, "A computational approach to drug repurposing using graph neural networks," Computers in Biology and Medicine, vol. 150, p. 105992, 2022.

[4] K. G. Y. S. W. W. Y Bai, "Bi-Level Graph Neural Networks for Drug-Drug Interaction Prediction," arXiv preprint arXiv:2006.14002, 2020.

[5] Y. W. O. V. R. M. B. B. J. M. T. D. Yunchao (Lance) Liu, "Interpretable Chirality-Aware Graph Neural Network for Quantitative Structure-Activity Relationship Modeling in Drug Discovery," in Proceedings of the AAAI Conference on Artificial Intelligence, 2023.

[6] M. B. F. S. Pietro Bongini, "Molecular generative Graph Neural Networks for Drug Discovery," Neurocomputing, vol. 450, pp. 242-252, 2021.

[7] J. M. F. M. Mark Cheung, "Graph Neural Networks for COVID-19 Drug Discovery," in 2020 IEEE International Conference on Big Data (Big Data), 2020.

[8] Z. X. K. C. H. J. M. Z. Jiacheng Xiong, "Graph neural networks for automated de novo drug design," Drug Discovery Today, vol. 26, pp. 1382-1393, 2021.

[9] B. L. J. L. Kehang Han, "Reliable Graph Neural Networks for Drug Discovery Under Distributional Shift," arXiv, 25 November 2021.

[10] X. L. L. L. X.-S. H. Y. C. K. X. Xiao-Shuang Li, "Multiphysical graph neural network (MP-GNN) for COVID-19 drug design," Briefings in Bioinformatics, vol. 23, no. 4, pp. 1-10, 2022.

[11] C. S. F. H. H. F. W. X. W. Z. Xuan Liu, "GraphCDR: a graph neural network method with contrastive learning for cancer drug response prediction," Briefings in Bioinformatics, vol. 23, no. 1, p. 1–9, 2022.

[12] Z. L. S. Z. S. W. X. W. Q. Y. Z. W. Mingjian Jiang, "Drug–target affinity prediction using graph neural network and contact maps," RSC Advances, vol. 10, no. 35, pp. 20701-20712, 2020.

[13] Z. O. W. C. R. F. D. Z. C. J. C. J. W. Yiheng Zhu, "TGSA: protein-protein association-based twin graph neural networks for drug response prediction with similarity augmentation," Bioinformatics, vol. 38, no. 2, pp. 461-468, 2022.

[14] M. H. &. A. Lakizadeh, "Prediction of drug-drug interaction events using graph neural networks based feature extraction," Scientific Reports, vol. 12, no. 1, p. 15590, 2022.

[15] M. H. A. A. Lakizadeh, "GNN-DDI: a new data integration framework for predicting drug-drug interaction events based on graph neural networks," 2022.

[16] X. L. S. S. L. D. a. H. L. Jinxian Wang, "DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations," Briefings in Bioinformatics, vol. 23, no. 1, 2022.