Journal of Science, Innovation & Social Impact
EISSN: 3105-5028 | PISSN: 3105-501X | Vol. 1, No. 2 (2025)

Review

Integrating Graph Neural Networks into Computational Drug Design for Pediatric Diseases

Lina Zhang [1,*] and Wei Zhou [1]

[1] School of Life Sciences, Central South University, Changsha, China
* Correspondence: Lina Zhang, School of Life Sciences, Central South University, Changsha, China

Abstract: Computational drug design (CDD) has emerged as a powerful tool for accelerating the discovery of novel therapeutics, particularly for diseases affecting pediatric populations. However, traditional CDD methods often struggle with the complexities of pediatric diseases, including unique biological targets and developmental considerations. Graph Neural Networks (GNNs), a class of deep learning models capable of processing graph-structured data, have shown remarkable promise in various CDD tasks, such as drug-target interaction prediction, molecular property prediction, and de novo drug design. This review provides a comprehensive overview of the integration of GNNs into CDD for pediatric diseases. We begin with a historical overview of CDD and its applications in pediatrics, highlighting the limitations of traditional approaches. We then delve into the core concepts of GNNs and their specific adaptations for CDD. We discuss the application of GNNs across diverse pediatric diseases, including cancers, genetic disorders, and infectious diseases, focusing on how GNNs can address specific challenges such as data scarcity and target heterogeneity. A comparative analysis of different GNN architectures and their performance in pediatric CDD is presented, along with a discussion of current challenges and limitations, such as the need for improved interpretability and validation. Finally, we explore future perspectives and opportunities for GNN-driven CDD in pediatrics, including the integration of multi-omics data, the development of personalized medicine approaches, and the application of explainable AI techniques. This review aims to provide a valuable resource for researchers and practitioners interested in leveraging GNNs to accelerate the development of safe and effective treatments for pediatric diseases.

Keywords: Graph Neural Networks (GNNs); computational drug design (CDD); pediatric diseases; drug discovery; Deep Learning; drug-target interaction; molecular property prediction

Received: 25 November 2025
Revised: 13 January 2026
Accepted: 25 January 2026
Published: 31 January 2026

1. Introduction

1.1. Background and Motivation

Computational drug design (CDD) has become an indispensable tool in modern drug discovery, accelerating the identification of promising drug candidates and reducing development costs. However, pediatric drug development faces unique hurdles, including limited patient populations, ethical considerations in clinical trials, and age-related differences in drug metabolism and efficacy. Graph Neural Networks (GNNs) offer a powerful approach to address these challenges by leveraging the structural information of molecules and biological networks to predict drug-target interactions and optimize drug properties specifically for pediatric populations, potentially accelerating the development of safe and effective treatments [1].

1.2. Scope and Objectives

This review focuses on the application of Graph Neural Networks (GNNs) within Computational Drug Design (CDD) specifically targeting pediatric diseases. We will cover GNN-based methods for target identification, drug-target interaction prediction, de novo molecule design, and drug repurposing. The objectives are to summarize current research employing GNNs in these areas, identify key challenges such as data scarcity and model validation, and outline promising future directions, including the development of more robust and interpretable GNN models tailored for pediatric drug discovery, considering factors like p_k and IC_{50} values.

2. Historical Overview of CDD in Pediatrics

2.1. Early Approaches to Pediatric Drug Design

Early computational drug design (CDD) efforts in pediatrics largely mirrored adult approaches, employing structure-based and ligand-based methods (Table 1). Structure-based design utilized protein structures, when available, to identify potential drug candidates that bind to target proteins implicated in pediatric diseases. This involved techniques like molecular docking and scoring to predict binding affinity. Ligand-based design, conversely, relied on known active molecules to build pharmacophore models or quantitative structure-activity relationship (QSAR) models, relating chemical structure to biological activity (IC_50, K_i, etc.). Successes included the identification of compounds with promising activity in vitro. However, these early methods often failed to adequately address the unique physiological characteristics of children, such as differences in drug metabolism, distribution, and target expression. This resulted in limited in vivo efficacy and highlighted the need for pediatric-specific CDD strategies.

Table 1. Comparison of Traditional CDD Methods in Pediatrics.

Feature	Structure-Based Design	Ligand-Based Design
Basis	Protein structure	Known active molecules
Techniques	Molecular docking, scoring	Pharmacophore models, QSAR models
Input Data	3D structure of target protein	Chemical structures and activity data (IC_{50}, K_i, etc.)
Output	Potential drug candidates that bind to target proteins	QSAR models relating structure to activity; identification of promising compounds.
Successes	Identification of compounds with promising *in vitro* activity	Identification of compounds with promising *in vitro* activity.
Limitations in Pediatrics	Fails to adequately address unique pediatric physiological characteristics, such as drug metabolism, distribution, and target expression. This leads to limited *in vivo* efficacy.	Fails to adequately address unique pediatric physiological characteristics, such as drug metabolism, distribution, and target expression. This leads to limited *in vivo* efficacy.

2.2. Challenges in Pediatric Drug Development

Pediatric drug development faces unique hurdles beyond those encountered in adult populations. Significant differences in physiology, including organ system maturation, body composition (V_{body}), and renal function (GFR), impact drug absorption, distribution, metabolism, and excretion (ADME). Metabolic pathways, particularly those involving cytochrome P450 enzymes ($CYP450$), exhibit age-dependent activity, leading to

unpredictable drug clearance and potential toxicity. Furthermore, disease presentation can vary significantly between children and adults, necessitating age-specific clinical trial designs and endpoints. These factors, compounded by ethical considerations in pediatric research and the relatively small market size for individual pediatric indications, underscore the urgent need for innovative approaches, such as computational drug design, to accelerate the development of safe and effective therapies for children [2].

3. GNNs for Drug-Target Interaction Prediction in Pediatric Diseases

3.1. GNN Architectures for DTI Prediction

GNNs have emerged as powerful tools for DTI prediction, offering various architectures suited for different aspects of drug and target representation. Graph Convolutional Networks (GCNs) utilize spectral convolutions to aggregate information from neighboring nodes, effectively capturing local structural patterns in both drug and protein graphs. The aggregation process can be represented as $h_i^{(l+1)} = \sigma(\sum_{j \in N(i)} \frac{1}{c_{ij}} W^{(l)} h_j^{(l)})$, where $h_i^{(l)}$ is the node feature at layer l, $N(i)$ is the neighborhood of node i, c_{ij} is a normalization constant, and $W^{(l)}$ is a trainable weight matrix. GCNs are computationally efficient but can be limited in capturing long-range dependencies.

Graph Attention Networks (GATs) address this limitation by introducing attention mechanisms, allowing nodes to selectively attend to their neighbors based on learned weights. This is particularly relevant for pediatric targets, where subtle structural variations can significantly impact drug binding. The attention coefficient e_{ij} is calculated as $a(Wh_i, Wh_j)$. While GATs offer improved expressiveness, they can be more computationally demanding than GCNs.

Message Passing Neural Networks (MPNNs) provide a general framework encompassing GCNs and GATs (Table 2). MPNNs consist of a message passing phase and a readout phase. The message passing phase iteratively updates node representations by aggregating information from their neighbors. The readout phase then uses these updated representations to predict DTI. The flexibility of MPNNs allows for customization to specific pediatric disease contexts, but requires careful design of the message and update functions.

Table 2.: Performance Comparison of GNN Architectures for DTI Prediction.

GNN Architecture	Advantages	Disadvantages	Relevant to Pediatric Targets?	Computational Cost
Graph Convolutional Networks (GCNs)	Computationally efficient, effectively captures local structural patterns.	Limited in capturing long-range dependencies.	Useful for capturing basic structural features.	Low
Graph Attention Networks (GATs)	Improved expressiveness, allows nodes to selectively attend to neighbors, suitable for capturing subtle structural variations.	More computationally demanding than GCNs.	Particularly relevant due to the need to capture subtle structural variations in pediatric targets.	High
Message Passing Neural Networks (MPNNs)	General framework encompassing GCNs and GATs, high flexibility and customization.	Requires careful design of message and update functions.	Highly relevant due to the ability to customize to specific pediatric disease contexts.	Medium to High (depending on implementation)

3.2. Application to Specific Pediatric Diseases

GNNs are increasingly being applied to predict drug-target interactions (DTIs) for pediatric diseases, offering a promising avenue for accelerating drug discovery [3]. In childhood cancers, for example, GNNs have been employed to identify potential drug candidates for leukemia. By representing both drugs and target proteins as nodes in a graph, and their interactions as edges, GNN models can learn complex relationships and predict the likelihood of a drug binding to a specific target implicated in leukemia development. Features such as gene expression data or protein sequence information can be incorporated as node attributes, enhancing the predictive power of the model. Similarly, in neuroblastoma research, GNNs can be used to prioritize compounds that target proteins involved in tumor growth and metastasis. The ability of GNNs to handle heterogeneous data, including genomic, proteomic, and chemical information, is particularly beneficial in understanding the complex interplay of factors contributing to pediatric cancers. Furthermore, GNNs can be adapted to predict DTIs for genetic disorders affecting children, where identifying drugs that modulate the activity of specific disease-causing proteins is crucial. The benefits of using GNNs in these applications include improved prediction accuracy compared to traditional methods, the ability to identify novel drug targets, and the potential to personalize treatment strategies based on individual patient profiles.

4. GNNs for Molecular Property Prediction and De Novo Design in Pediatrics

4.1. Predicting ADMET Properties with GNNs

Graph Neural Networks (GNNs) offer a powerful approach to predict ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties of drug candidates, crucial for pediatric drug development. Accurately predicting these properties can significantly reduce the risk of adverse drug reactions in children, a population particularly vulnerable due to their developing physiology. GNNs can learn complex relationships between molecular structure and ADMET outcomes by representing molecules as graphs, where nodes represent atoms and edges represent bonds. These models can then be trained on existing ADMET data to predict these properties for novel compounds [4].

Specifically, GNNs can predict parameters like oral bioavailability (F), volume of distribution (V_d), metabolic clearance (CL), and potential for drug-drug interactions mediated by cytochrome P450 enzymes (Table 3). Furthermore, GNNs can be used to assess potential toxicities, such as hepatotoxicity and cardiotoxicity. By integrating these predictions early in the drug discovery process, researchers can prioritize compounds with favorable ADMET profiles, leading to safer and more effective medications for pediatric patients and reducing the need for extensive in vivo testing.

Table 3. GNN-based ADMET prediction performance for Pediatric Drug Candidates.

ADMET Property	GNN Prediction	Significance for Pediatric Drug Development
Oral Bioavailability (F)	Predicts the fraction of drug absorbed into systemic circulation after oral administration.	Crucial for determining appropriate oral dosages, considering children's variable absorption rates.
Volume of Distribution (V_d)	Estimates the extent to which a drug distributes throughout the body.	Essential for calculating loading doses and understanding drug exposure in different tissues, especially in the context of age-related body composition changes.
Metabolic Clearance (CL)	Quantifies the rate at which a drug is removed from the body.	Critical for determining dosing intervals, considering that children's metabolic

		pathways may be immature or different from adults.
Cytochrome P450 (CYP) Inhibition	Predicts the potential for a drug to inhibit CYP enzymes, leading to drug-drug interactions.	Important for minimizing the risk of adverse drug reactions resulting from altered drug metabolism, as children may be more susceptible to these interactions.
Hepatotoxicity	Assesses the potential for a drug to cause liver damage.	Essential for identifying and avoiding drugs that may pose a risk to children's developing livers.
Cardiotoxicity	Estimates the risk of adverse effects on the heart.	Critical to prevent any cardiac complications that can severely affect children.

4.2. De Novo Drug Design Using GNNs

GNNs offer a powerful approach to de novo drug design, enabling the generation of novel molecules tailored for pediatric diseases. These models can be trained on datasets of existing drugs and bioactive compounds, learning the complex relationships between molecular structure and desired properties like target affinity and ADMET profiles. By sampling from the learned chemical space, GNNs can propose new molecular structures predicted to possess specific characteristics relevant to treating childhood illnesses.

However, applying de novo design to pediatric drug development presents unique challenges. Ensuring safety and efficacy in children requires careful consideration of age-related physiological differences, such as variations in drug metabolism and organ development. GNN-generated molecules must be rigorously evaluated for potential off-target effects and toxicity in pediatric-specific models. Furthermore, the limited availability of pediatric-specific data for training GNNs necessitates the development of transfer learning strategies and data augmentation techniques to improve model accuracy and reliability. The use of appropriate scoring functions that incorporate pediatric-specific parameters is also crucial for prioritizing promising drug candidates [5].

5. Comparison of GNNs with Other Methods and Current Challenges

5.1. Comparison with Traditional Machine Learning Methods

GNNs offer distinct advantages over traditional machine learning methods like support vector machines (SVMs) and random forests (RF) in computational drug design (CDD) for pediatric diseases (Table 4). Traditional methods often require manual feature engineering, a time-consuming and potentially biased process. GNNs, conversely, learn directly from the graph structure of molecules, capturing complex relationships between atoms and bonds without explicit feature definition. This is particularly beneficial when dealing with limited data, a common challenge in pediatric drug development. However, traditional methods can be more computationally efficient and require less training data than GNNs [6]. Furthermore, the interpretability of models like RF can be higher compared to the "black box" nature of some GNN architectures. The choice between GNNs and traditional methods depends on the specific application, data availability, and the need for interpretability versus predictive accuracy.

Table 4. Comparative Analysis of GNNs vs. Traditional ML in CDD for Pediatrics.

Feature	GNNs	Traditional ML (SVM, RF)
Feature Engineering	Automates feature extraction from graph structure.	Requires manual, often time-consuming, feature engineering.
Data Requirements	Can handle limited data but generally benefit from more data.	Can perform well with less training data.

Computational Efficiency	Can be computationally expensive, especially for large graphs.	Generally more computationally efficient.
Interpretability	Often considered "black box" models with lower interpretability, although progress is being made in explainable GNNs.	Can offer higher interpretability (e.g., RF feature importance).
Capability to capture complex relationships	Excellent at capturing complex relationships between atoms and bonds.	Limited ability to capture complex relationships without extensive feature engineering.
Suitability for Limited Data (Pediatrics)	Particularly beneficial when dealing with limited data, a common challenge in pediatric drug development.	Can be effective with limited data, but performance may be limited by the need for manual feature engineering.
Predictive Accuracy	Potentially higher predictive accuracy by learning directly from the graph structure.	Predictive accuracy depends heavily on the quality of hand-engineered features.

5.2. Challenges and Limitations of GNNs in Pediatric CDD

GNNs, while promising, face significant hurdles in pediatric computational drug design (CDD) (Table 5). Data scarcity is a primary concern, as pediatric-specific datasets for drug activity and safety are often limited compared to adult data [7]. This can lead to overfitting and poor generalization of GNN models. Furthermore, the "black box" nature of many GNN architectures hinders interpretability. Understanding why a GNN predicts a certain outcome is crucial for building trust and facilitating rational drug design. Improved methods for explaining GNN predictions, such as attention mechanisms or feature importance analysis, are needed. Finally, rigorous validation of GNN predictions is essential [8]. Due to the limited availability of pediatric clinical trial data, alternative validation strategies, such as physiologically-based pharmacokinetic (PBPK) modeling and in vitro studies, are necessary to ensure the safety and efficacy of predicted drug candidates. Addressing these challenges is critical for realizing the full potential of GNNs in pediatric CDD [9,10].

Table 5. Challenges in Applying GNNs to Pediatric Drug Discovery.

Challenge	Description
Data Scarcity	Limited pediatric-specific datasets for drug activity and safety compared to adult data. This can lead to overfitting and poor generalization of GNN models.
Lack of Interpretability	The "black box" nature of many GNN architectures hinders understanding *why* a GNN predicts a certain outcome, which is crucial for building trust and facilitating rational drug design. Improved methods for explaining GNN predictions, such as attention mechanisms or feature importance analysis, are needed.
Validation Difficulties	Limited availability of pediatric clinical trial data necessitates alternative validation strategies, such as physiologically-based pharmacokinetic (PBPK) modeling and *in vitro* studies, to ensure the safety and efficacy of predicted drug candidates.

6. Future Perspectives

6.1. Integrating Multi-Omics Data

Integrating multi-omics data holds immense promise for enhancing GNN-driven drug discovery for pediatric diseases [11]. By incorporating genomics, transcriptomics,

proteomics, and metabolomics data, GNNs can gain a more holistic understanding of disease mechanisms specific to children. For instance, integrating gene expression data (x_i) with protein-protein interaction networks can reveal key regulatory pathways dysregulated in pediatric cancers [12]. However, significant challenges exist. These include the inherent heterogeneity and high dimensionality of multi-omics data, requiring sophisticated feature engineering and data normalization techniques. Furthermore, effectively integrating data with varying scales and formats, while accounting for patient-specific variations, remains a crucial hurdle. Overcoming these challenges will be essential to fully realize the potential of multi-omics informed GNNs in pediatric drug development [13,14].

6.2. Personalized Medicine Approaches

GNNs hold immense promise for personalized medicine in pediatric diseases, enabling treatment strategies tailored to individual patient profiles [15]. By integrating multi-omic data, such as genomics, transcriptomics, and proteomics, into graph-based representations, GNNs can identify patient-specific disease mechanisms [16]. For instance, a patient's genetic mutations (g_i), combined with clinical data (c_i), can be encoded as node features within a patient-specific graph. GNNs can then learn complex relationships between these features and predict individual treatment responses [17]. This approach allows for the identification of optimal drug candidates and dosages for each patient, minimizing adverse effects and maximizing therapeutic efficacy. Furthermore, GNNs can facilitate the discovery of novel biomarkers for disease stratification and treatment monitoring, paving the way for more precise and effective interventions in pediatric populations [18].

7. Conclusion

7.1. Summary of Key Findings

This review highlights the promising potential of Graph Neural Networks (GNNs) in revolutionizing Computational Drug Design (CDD) for pediatric diseases. Our analysis reveals that GNNs excel at capturing complex relationships within biological data, leading to improved prediction accuracy in tasks such as drug-target interaction prediction, toxicity assessment, and de novo molecule generation. Specifically, GNNs demonstrate a superior ability to model intricate molecular structures and biological networks compared to traditional methods. However, the application of GNNs in pediatric CDD remains relatively nascent. Significant challenges persist, including the limited availability of pediatric-specific data and the need for GNN architectures tailored to the unique biological characteristics of children. Further research is crucial to address these limitations and unlock the full potential of GNNs in developing safer and more effective treatments for pediatric diseases. Future work should focus on developing novel GNN architectures and incorporating multi-omics data to improve prediction accuracy and robustness.

References

1. Y. S. Cai, "Organizational restructuring of fintech enterprises: A strategic study balancing compliance and innovation," Financial Economics Insights, vol. 3, no. 1, pp. 1–10, 2026.
2. C. Abate, S. Decherchi, and A. Cavalli, "Graph neural networks for conditional de novo drug design," Wiley Interdisciplinary Reviews: Computational Molecular Science, vol. 13, no. 4, e1651, 2023.
3. R. Satheeskumar, "Enhancing drug discovery with AI: Predictive modeling of pharmacokinetics using Graph Neural Networks and ensemble learning," Intelligent Pharmacy, vol. 3, no. 2, pp. 127-140, 2025.
4. J. Jin, T. Zhu, and C. Li, "Graph Neural Network-Based Prediction Framework for Protein-Ligand Binding Affinity: A Case Study on Pediatric Gastrointestinal Disease Targets," Journal of Medicine and Life Sciences, vol. 1, no. 3, pp. 136–142, 2025.
5. C. Wang, G. A. Kumar, and J. C. Rajapakse, "Drug discovery and mechanism prediction with explainable graph neural networks," Scientific Reports, vol. 15, no. 1, 179, 2025.
6. J. Xiong et al., "Graph neural networks for automated de novo drug design," Drug discovery today, vol. 26, no. 6, pp. 1382-1393, 2021.

7. R. Luo, X. Chen, and Z. Ding, "SeqUDA-Rec: Sequential user behavior enhanced recommendation via global unsupervised data augmentation for personalized content marketing," arXiv preprint arXiv:2509.17361, 2025.
8. Y. Ding, Y. Shen, K. Yi, and Y. G. Wang, "GNNs in drug design," in Deep Learning in Drug Design, Academic Press, 2026, pp. 31-50.
9. K. Wang et al., "Recent advances from computer-aided drug design to artificial intelligence drug design," RSC Medicinal Chemistry, vol. 15, no. 12, pp. 3978-4000, 2024.
10. P. Bongini, "Graph neural networks for drug discovery: An integrated decision support pipeline," in 2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE), 2023, pp. 218-223.
11. V. Ponzi and C. Napoli, "Graph Neural Networks: Architectures, Applications, and Future Directions," IEEE Access, 2025.
12. J. L. Zhao, "Graph-based deep dive on AI startup revenue composition and venture capital network effect," Economics and Management Innovation, vol. 3, no. 1, pp. 27–36, 2026.
13. G. Corso et al., "Graph neural networks," Nature Reviews Methods Primers, vol. 4, no. 1, 17, 2024.
14. Y. Wang, Z. Li, and A. Barati Farimani, "Graph neural networks for molecules," in Machine learning in molecular sciences, Cham: Springer International Publishing, 2023, pp. 21-66.
15. S. Li, K. Liu, and X. Chen, "A context-aware personalized recommendation framework integrating user clustering and BERT-based sentiment analysis," Journal of Computer, Signal, and System Research, vol. 2, no. 6, pp. 100–108, 2025.
16. R. R. Saxena and R. Saxena, "Applying graph neural networks in pharmacology," Authorea Preprints, 2024.
17. Y. Li, G. Zhang, P. Wang, Z. G. Yu, and G. Huang, "Graph neural networks in biomedical data: a review," Current Bioinformatics, vol. 17, no. 6, pp. 483-492, 2022.
18. C. L. Cheong, "Research on AI security strategies and practical approaches for risk management," Journal of Computer, Signal, and System Research, vol. 2, no. 7, pp. 98–115, 2025.