A Comparative Study of Graph Neural Network Models for Drug-Target Interaction Prediction

Jaqueline Bitencourt[1], **Anderson Tavares**[1]

[1]Institute of Informatics – Federal University of Rio Grande do Sul (UFRGS)
Post Office Box 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

{jbcorreia, artavares}@inf.ufrgs.br

Abstract. *Accurately predicting drug-target interactions (DTI) is crucial for computational drug discovery, yet there's a research gap in evaluating existing graph neural network (GNN) models rather than developing novel architectures. This study provides a comparative analysis of three state-of-the-art GNN architectures – GraphSAGE, Graph Attention Network (GAT), and Graph Isomorphism Network (GIN) – for predicting interactions between chemical compounds and five protein targets. Using a dataset of 73,938 samples representing interactions between compounds and five protein targets derived from PubChem, we implement a robust evaluation framework with hyperparameter optimization and cross-validation. Our results show GraphSAGE achieves the highest accuracy (93%) and precision (79%), while GIN exhibits superior recall (72%). This work contributes to the field by: (1) providing a comprehensive evaluation framework for GNN models in DTI prediction; (2) offering empirical evidence of architecture-specific advantages for different application contexts; and (3) introducing a new benchmark dataset that facilitates reproducibility and further research in computational drug discovery.*

1. Introduction

Identifying a biological target is the cornerstone of modern drug discovery, as it enables the development of therapeutics that modulate disease mechanisms at the molecular level, either by inhibiting pathological activity or enhancing beneficial effects. A biological target is a protein or nucleic acid that interacts with a drug to induce a therapeutic effect [Wan et al. 2024]. Proteins are the most common biological targets due to their functional roles in disease progression. As large biomolecules composed of amino acid chains, proteins catalyze biochemical reactions, transmit cellular signals, and maintain structural integrity [Sachdev and Gupta 2019].

Computational drug discovery leverages machine learning to identify patterns in complex biological data by integrating diverse data types, including molecular properties (e.g., mass, logP, polarity), chemical structures (e.g., atomic composition, bonding patterns), and biological data (e.g., protein sequences, toxicity, metabolism). However, developing effective predictive models remains challenging due to data heterogeneity, high dimensionality, and inconsistencies in molecular representation. Notably, representation issues in notations like SMILES can introduce ambiguities and invalid molecules, affecting model reliability [Krenn et al. 2020].

To address these challenges, Graph Neural Networks (GNNs) have gained prominence due to their ability to effectively model the relational structure of molecular and

biological data [Reiser et al. 2022]. GNNs operate on graph-structured data, where nodes represent atoms and edges define molecular bonds or biological interactions. This structure naturally captures the topology and complex dependencies in biochemical interactions, making GNNs particularly well-suited for DTI prediction tasks.

Although several studies have proposed novel GNN architectures for DTI prediction, there remains a significant gap in the literature regarding the systematic and comparative evaluation of existing models. This research aims to address the following question: "Which GNN architectures are most suitable for different drug-target interaction prediction scenarios, and how do their performance characteristics vary in terms of precision, recall, and practical applicability?" This question is fundamental to guiding researchers and practitioners in selecting appropriate models based on specific application requirements, such as initial drug screening (where high precision is critical) or drug repurposing (where the balance between precision and recall is essential).

This study models DTI prediction as a multilabel graph classification problem, comparing three Graph Neural Network (GNN) models – GraphSAGE, Graph Attention Network (GAT), and Graph Isomorphism Network (GIN). These models use different approaches, such as neighborhood aggregation and attention mechanisms, to process molecular graph data. The models were trained on a custom dataset of 73,938 samples, including 66,631 DTIs from PubChem, representing chemical compounds (via SMILES) and their interactions with five target proteins. The dataset categorizes compounds as active or inactive for each protein target, including diverse molecular structures.

The study makes several key contributions to computational drug discovery. It provides a systematic comparison of three state-of-the-art GNNs, filling a gap in the literature where research focuses on novel architectures rather than rigorous evaluation. It introduces a high-quality dataset of chemical-protein interactions. The dataset and code are available at `https://github.com/JaquelineBitencourt/ProjetoDL.git`. The evaluation framework is methodologically robust, employing hyperparameter optimization, 5-fold stratified cross-validation, and three independent runs for reliability. The analysis reveals that GraphSAGE achieves the highest precision (79%), while GIN has better recall (72%), offering practical insights for researchers selecting models based on specific needs.

2. Theoretical Foundation

A graph is a mathematical structure composed of a set of vertices (or nodes) and edges (or links) that define connections between nodes, formally represented as G = (V, E). Atoms in molecular chemistry form the nodes, and chemical bonds define the edges. GNNs have been developed to process and extract knowledge from graph-structured data [Zhou et al. 2020].

A classic architecture for GNNs is the Message Passing Neural Network (MPNN) [Scarselli et al. 2008], in which the graph entities (nodes, edges, and the graph itself) have "hidden state" vectors, also called embeddings. These vectors encode a useful representation of the model. GNN layers process information through the operations below:

- **Message:** Each entity produces a message to its neighbors. A *message* typically contains the embedding of the entity. On the first layer, the embedding is the fea-

ture vector of the entity. Then, each entity receives the messages from its neighbors, including itself. It is common to use embeddings and exchange messages for not all graph entities; for example, if a dataset presents only node features, embeddings for edges and the entire graph are not defined or used;

- **Aggregation:** After receiving all the messages from its neighbors, each node produces a resulting vector with a function that aggregates the received messages. Typically, the aggregation function is an average or sum of the received messages (average- or sum-pooling).
- **Update:** This produces the next-layer embedding of the entity from the aggregation of the received messages. This is usually the trainable part of the GNN since the update function is typically a multi-layer perceptron (MLP).

After a defined number of layers performing these three operations, a readout operation extracts the useful information for the desired graph entity (node, edge, or entire graph) and outputs the value corresponding to the task at hand (e.g. a probability for classification). The readout operation is also commonly implemented with a trainable MLP. In a supervised learning setting, the trainable parameters of the GNN are adjusted via gradient descent on a loss function over the differences between GNN predictions and target values from the dataset.

Various GNN models have been proposed to improve upon the original approach. GraphSAGE [Hamilton et al. 2017] enables efficient learning on large-scale graphs and generalization to unseen nodes. Unlike traditional GNNs, which require full graph availability during training, GraphSAGE generates node embeddings for previously unseen nodes by leveraging a neighborhood aggregation strategy. Specifically, the model learns to aggregate feature information from a node's local neighborhood using one of three approaches: mean aggregation, LSTM-based aggregation, or max pooling. This inductive learning paradigm allows GraphSAGE to scale effectively to dynamic and evolving graphs.

Graph Attention Networks (GAT) [Veličković et al. 2017], integrates an attention mechanism into the message-passing process. Unlike previous models that uniformly aggregate neighbor information, GAT assigns learnable attention weights to each neighbor, allowing the network to focus on the most relevant connections. Additionally, multi-head attention stabilizes learning by computing multiple attention scores in parallel and obtaining the final embedding by concatenating the linear transformations of the embeddings obtained in each attention head. This mechanism has demonstrated superior performance in tasks involving heterogeneous graphs, where nodes and edges exhibit varying levels of importance.

Graph Isomorphism Network (GIN) [Xu et al. 2018] achieves maximal discriminative power by employing a learnable sum aggregator function with an MLP and a learnable parameter that controls the importance of the node's own embedding during its update. This ensures that structurally distinct graphs generate unique embeddings, addressing a fundamental limitation of previous approaches: GIN is as powerful as the Weisfeiler-Lehman graph isomorphism test [Weisfeiler and Leman 1968]. Consequently, GIN is particularly well-suited for graph classification tasks, where distinguishing between different graph topologies is crucial.

3. Related Work

Table 1 provides a comparative overview of deep learning approaches for DTI prediction. Most prior studies have focused on proposing new models or hybrid architectures—typically combining graph neural networks (GNNs) for drug representation with convolutional or recurrent networks for protein sequences. While these works contributed important architectural innovations, few have systematically evaluated existing GNN architectures under a consistent experimental setup.

Table 1. Comparative Summary of Deep Learning Models for DTI Prediction

Study	Comparison Type	Neural Network Architectures	Dataset Source & Size
[Öztürk et al. 2018]	Proposes a new model	CNN (without GNN component)	Davis, KIBA
[Tsubaki et al. 2019]	Proposes a new model	GNN for compounds, CNN for proteins	Human, C.elegans, DrugBank and Matador
[Elbasani et al. 2021]	Proposes a new model	GCN with Bi-LSTM/Bi-GRU for proteins, GNN for compounds	Annotated DTI
[Hasebe 2021]	Proposes a new model	Knowledge-embedded MPNN	ESOL, FreeSolv, Lipophilicity, Tg
[Nguyen et al. 2021]	Proposes a new model	GCN, GAT, GIN, GAT-GCN for drugs and CNN for proteins	Davis, KIBA
[Wang et al. 2021]	Proposes a new model	GAT for drug and protein representation	Human, C.elegans, DUD-E, PDBbind
[Cheng et al. 2021]	Proposes a new model	Multi-head self-attention for proteins, GAT for drugs	Human, C.elegans, DUD-E, DrugBank
[Tran et al. 2022]	Proposes a new framework	GENConv, GCNConv, and HypergraphConv for drugs; 1D CNN for proteins	Davis, KIBA, Allergy
[Liu et al. 2022]	Proposes a new model	Three-way transformer for drugs, proteins and complex	DUD-E, LIT-PCBA, MUV
[Li et al. 2022]	Proposes a new model	Multi-channel graph neural network (GCN + GAT)	Human, C.elegans
[Tang et al. 2024]	Proposes a new model	BiGRU with soft-attention for proteins, GraphSAGE for drugs, fusion by attention neural network	Davis, KIBA
[Dandibhotla et al. 2025]	Proposes a new model	Hybrid GNN with Random Forest and XGBoost	PDBbind v.2020, DUDE-Z v.2023
Our work	**Comparative GNN evaluation for DTI prediction**	**GraphSAGE, GAT, GIN**	**PubChem (73.938 samples)**

Early efforts like DeepDTA [Öztürk et al. 2018] used CNNs to process SMILES and protein sequences. [Tsubaki et al. 2019] extended this by integrating GNNs for compounds and CNNs for proteins. Subsequent work introduced architectural variations such as knowledge-enhanced MPNNs [Hasebe 2021], multi-channel GNNs [Li et al. 2022], attention mechanisms [Cheng et al. 2021, Tang et al. 2024], and even transformer-based frameworks [Liu et al. 2022, Su et al. 2024], aiming to improve generalization and interpretability in DTI and drug-drug prediction tasks.

Despite this progress, most studies emphasize novel model proposals rather than comparative benchmarking as shown in Table 1. For instance, GraphDTA [Nguyen et al. 2021] explored multiple GNN variants (GCN, GAT, GIN) but still within a model-centric development framework. In contrast, our work fills a gap in the literature by conducting a focused comparative evaluation of three widely used GNN architectures—GraphSAGE, GAT, and GIN—using a large, well-defined PubChem-derived dataset.

More specifically, our PubChem-derived dataset offers several distinct advantages: (1) it includes five well-defined protein targets (P01584, P00352, O60674, O43613,

and O15151) representing different functional classes;(2) it contains 66,361 unique compounds resulting in 73,938 drug-target interaction pairs; and (3) it provides binary interaction classifications instead of continuous binding affinity values, which simplifies the prediction task while maintaining biological relevance. This approach allows a focused evaluation of the performance of different GNN architectures in predicting drug-target interactions for this specific set of targets.

4. Materials and Methods

This section outlines the methodology adopted to use GNNs for biological target prediction, formulated as a multilabel graph classification task, where each graph, which represents a drug (chemical compound), can be associated with multiple targets. The following subsections detail the data and methodology of this work.

4.1. Data Collection and Preprocessing

A total of 1,841,145 records corresponding to five distinct classes of target proteins were retrieved from PubChem[1]. The proteins included in this study and their key biological functions are summarized in Table 2.

Table 2. Summary of Selected Target Proteins

UniProtID	Protein Name	Description
P01584	Interleukin-1 Beta (IL-1β)	A pro-inflammatory cytokine that regulates immune responses and promotes T-cell activation, cytokine production, and angiogenesis.
P00352	Aldehyde Dehydrogenase 1A1 (ALDH1A1)	A cytosolic enzyme that catalyzes the oxidation of aldehydes into carboxylic acids, including the conversion of retinaldehyde into retinoic acid for vitamin A metabolism.
O60674	Tyrosine-Protein Kinase JAK2	A non-receptor tyrosine kinase involved in cytokine and hormone signaling. It phosphorylates receptors and activates STAT proteins, regulating gene expression.
O43613	Orexin/Hypocretin Receptor Type 1 (HCRTR1)	A G-protein-coupled receptor (GPCR) that preferentially binds orexin-A, triggering intracellular calcium signaling.
O15151	p53-Binding Protein MDM4 (MDMX)	A negative regulator of p53 and p73 that suppresses their transcriptional activation and apoptotic functions.

4.2. Dataset Preparation

The dataset initially contained 1,841,145 records, each representing an interaction between a compound and a target protein. For the purpose of this study, only active records – defined as those indicating measurable biological activity against the target proteins – were considered. This refinement was necessary to ensure that the dataset exclusively contained interactions relevant to the classification task.

Each record in the dataset has several attributes. However, only the following features were retained for analysis:

- **Compound Identifier (CID):** A unique identifier assigned to each chemical compound.

[1] `https://pubchem.ncbi.nlm.nih.gov/#query=P01584\%20OR\%20P00352\%20OR\%20O60674\%20OR\%20O43613\%20OR\%20O15151\&tab\=protein`

- **SMILES Notation:** A textual representation of molecular structure converted to graph-based representations via RDKit, extracting atom-level features (atomic number, hybridization, aromaticity, charge, hydrogen count, degree, ring membership) and capturing molecular connectivity. Additionally, global properties (molecular weight, logP, rotatable bonds, hydrogen bond acceptors/donors) are incorporated.
- **Active Label:** A binary indicator that denotes whether the compound exhibits biological activity against the target protein.
- **Protein Identifier (UniProt ID):** The unique identifier for the target protein.

To remove redundancy, duplicate records were removed. The dataset details before and after preprocessing are presented in Table 3. Note that our dataset contains 66,361 unique compounds (drugs) and 5 distinct protein targets, resulting in 73,938 drug-target interaction pairs. Many drugs in the dataset interact with multiple targets simultaneously, as the number of interaction pairs exceeds the number of unique drugs. The dataset exhibits class imbalance: O15151 contains much fewer interactions than the majority class P01584 (10,413 active interactions versus 17,515).

To evaluate the impact of hyperparameter optimization while preventing data leakage, the dataset was split into 80% for training and 20% for testing. The 80% training portion was further subdivided into training and validation sets to evaluate different hyperparameter configurations. This data split ensures that no information from the test set is inadvertently used during hyperparameter tuning.

Table 3. Summary of Dataset Before and After Preprocessing

UniProt Class	Total Records (Before Deduplication)	Unique Records	Active Records
P01584	363,155	362,668	17,515
P00352	221,566	217,601	16,529
O60674	245,743	235,478	16,145
O43613	667,845	333,948	13,336
O15151	343,536	332,464	10,413
Total	**1,841,845**	**1,482,159**	**73,938**

4.3. Model Configuration and Training

To optimize the performance of the GNN models (GraphSAGE, GAT, and GIN – see Section 2), we employed Optuna [Akiba et al. 2019], a widely used hyperparameter optimization framework. The optimization process was performed over 50 trials, and each model was trained for up to 200 epochs, with early stopping applied using Optuna's built-in MedianPrune strategy to halt underperforming trials efficiently. Optuna employs Bayesian optimization, efficiently balancing the exploration and exploitation in the hyperparameter space. Table 4 presents the optimal configuration for the optimized hyperparameters for each architecture.

Batch normalization is applied after each graph convolutional layer, helping to mitigate internal covariate shifts. Additionally, dropout is incorporated to prevent overfitting. After extracting graph-level embeddings through global mean pooling for GraphSAGE and GAT and global sum pooling for GIN, a Multi-Layer Perceptron (MLP) is

used for classification, integrating both graph-derived features and molecular-level global descriptors. The MLP consists of multiple fully connected layers with LeakyReLU activations, progressively refining learned representations before the final output layer.

To ensure reliable performance estimation, we employ 5-fold stratified cross-validation, splitting the dataset into five subsets where each fold alternates between training and validation. During each fold, the model is trained using binary cross-entropy loss (BCEWithLogitsLoss) and optimized via AdamW, leveraging weight decay for better generalization. The training process runs for up to 500 epochs, with an early stopping mechanism (patience = 200 epochs) to prevent overfitting. To further enhance the robustness of our evaluation, each experiment is repeated three times, and we report average performance metrics. The metrics obtained across different runs showed minimal variation, indicating that the models produced consistent results across multiple executions.

Table 4. Optimized Hyperparameters for each GNN model

Hyperparameter	GraphSAGE	GAT	GIN
Hidden Dimension	200	142	223
Number of Layers	4	4	4
Dropout Rate	0.2	0.1	0.1
MLP Hidden Dimension	247	187	121
MLP Layers	2	2	1
Learning Rate	0.001233	0.000199	0.000703
Weight Decay	0.00019	3.9777	4.4564
Heads	Not applicable	4	Not applicable

5. Results

This section presents the results of the comparative evaluation of the GraphSAGE, GIN and GAT architectures for predicting drug-target interactions (DTI), focusing on how they vary in precision, recall and practical applicability. The analysis aims to guide the choice of model according to the use scenario, such as drug screening or repositioning. The results, presented through confusion matrices Figure 1 and classification metrics Table 5, provide useful empirical evidence for applications in drug discovery.

GIN demonstrated the most balanced performance profile, achieving the highest average F1-score and recall, while maintaining strong precision and accuracy. GIN's high expressiveness for molecular structures appears to translate into practical advantages for multilabel classification tasks. Its aggregation may provide an effective balance between precision and recall that is valuable for drug discovery applications requiring balanced performance metrics.

GAT, with a 4-head attention mechanism, delivered solid performance across all metrics (accuracy, precision, recall, F1-score). While slightly lower than the other models, GAT's performance remains competitive, suggesting that its attention mechanism successfully identifies relevant molecular features for interaction prediction.

GraphSAGE exhibited exceptional performance for target O60674, with only 3.1% false negatives compared to GIN's 6.8% and GAT's 7.4%. This suggests that GraphSAGE's neighborhood aggregation mechanism may be particularly effective at capturing the interaction patterns specific to O60674. Conversely, for target P01584, all models

showed relatively lower performances, indicating that this protein's interactions may be more challenging to predict regardless of the GNN architecture employed.

Analysis of the confusion matrices across the five protein targets revealed target-specific performance variations. This suggests that certain DTI patterns may be more amenable to capture by specific model architectures. The consistent performance across models for certain targets (particularly evident in the confusion matrices) indicates that some interaction patterns are robustly captured by GNNs generally, while others present challenges regardless of the specific architecture. For instance, all models achieved their highest true positive rates for target O60674, suggesting that drugs interacting with this protein may have more distinctive structural features that facilitate accurate prediction. Moreover, the lower performance for target P01584 suggests that traditional GNN approaches may need additional information (e.g., protein dynamics or sequence) to improve performance for challenging targets.

Our findings highlight the importance of selecting appropriate GNN architectures based on specific application requirements and target characteristics in drug discovery pipelines. The minimal performance differences between models (all achieving F1-scores within 1% of each other) also suggest that ensemble approaches combining multiple GNN architectures might offer further performance improvements by leveraging their complementary strengths.

Table 5. Overall Performance of Models in Classifying DTI. Best values for each metric are in bold.

Model	Accuracy	Precision	Recall	F1-Score
GraphSAGE	**93**%	**79**%	68%	**73**%
GIN	92%	76%	**72**%	**73**%
GAT	92%	74%	71%	72%

6. Discussion

DTI prediction is a challenging task due to the complex and often non-linear relationships between molecular structures (drugs) and protein targets. Our comparative analysis of GraphSAGE, GIN, and GAT reveals that all three architectures achieve promising performance metrics, with accuracy rates exceeding 92% and F1-scores between 72-73%. These GNNs demonstrate robust capabilities in capturing the complex molecular relationships underlying DTI. Our models performed well despite the class imbalance observed in our dataset (see Section 4.2). This suggests that the GNNs effectively capture the structural patterns characteristic of active compounds, allowing the models to generalize well even with fewer examples for certain classes. Next, we discuss each model.

GraphSAGE: Strong Generalization, Sampling Limitations. GraphSAGE's high accuracy and precision make it particularly useful in applications where false positives are costly, such as in early stages of the drug discovery pipelines, prioritizing highly confident interactions before experimental validation. However, in multilabel classification, where interactions with multiple targets must be identified simultaneously, GraphSAGE's reliance on neighborhood sampling may limit its ability to capture the full complexity of molecular interactions. This occurs because the sampling strategy might miss

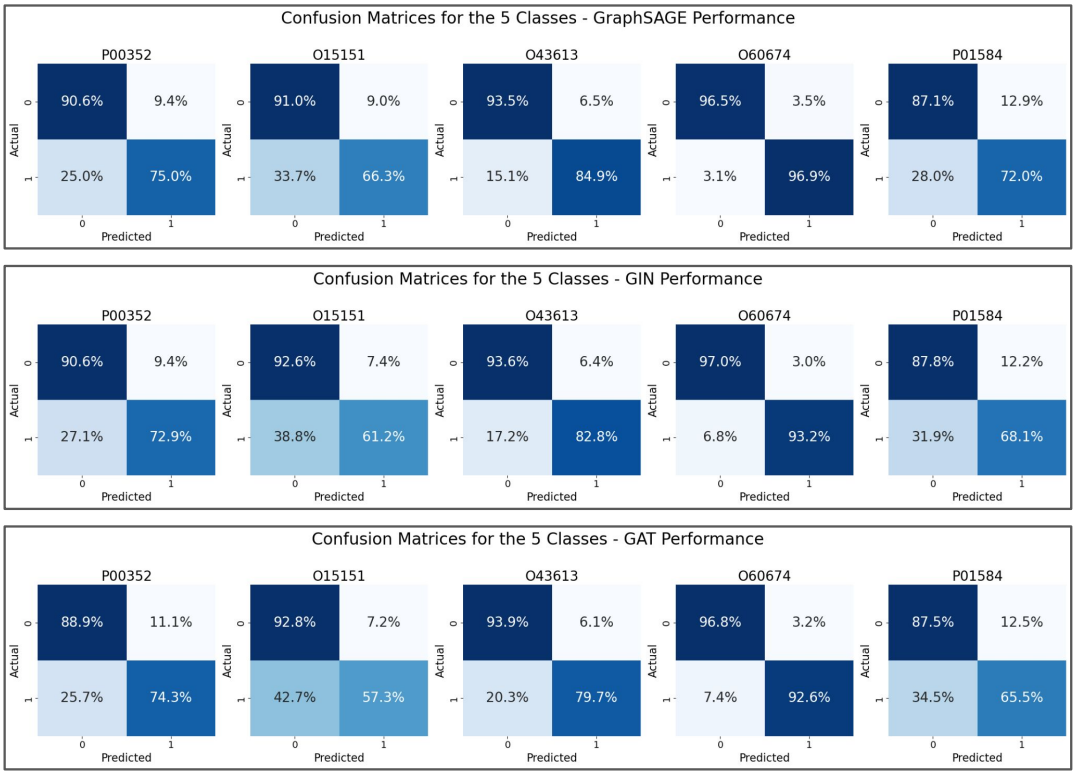

Figure 1. Confusion Matrices of GraphSAGE, GIN and GAT model. Note that, because of the multilabel nature of the problem, we use per-class confusion matrices (one-vs-others), and the values sum to 100% row-wise.

important structural features that contribute to interactions with certain targets. Increasing the number of sampled neighbors or incorporating additional molecular descriptors could help mitigate these limitations and improve recall.

GIN: High Expressiveness for Molecular Structures. The superior recall suggests that GIN effectively captures graph topological information, making it suitable for applications requiring high sensitivity, such as drug repurposing. In multilabel classification, where capturing shared and unique interactions across multiple targets is critical, GIN's aggregation allows it to differentiate structurally similar compounds interacting with different proteins. However, the aggregation method comes with a higher computational cost. GIN demonstrates a strong balance between recall and precision, making it a versatile choice for scenarios where discovering new interactions is a priority. This combination of high recall and precision plus its high accuracy is especially beneficial in comprehensive screening approaches, such as drug repurposing studies, where identifying a broad range of potential interactions is crucial.

GAT: Attention Mechanism Trade-offs. While the attention mechanism allows the GAT to focus on the most relevant molecular features, it may not be as effective in learning the full range of interactions due to its dependency on local attention weights. In multilabel classification, where understanding the interplay between different target interactions is crucial, GAT's focus on immediate neighborhood relationships may limit its capacity to model long-range dependencies. Integrating hierarchical attention networks [Wang et al. 2024] could improve its ability to capture broader interaction patterns

and better distinguish complex drug-target relationships. In this study, we used 4 attention heads, which provided richer feature extraction from neighboring nodes. However, while multiple heads improve expressiveness, they also increase computational complexity. On the other hand the attention mechanism enhances interpretability by highlighting molecular features that most influence predicted interactions – an advantage that can aid medicinal chemistry optimization.

In DTI prediction, achieving a balance between precision and recall is crucial. High precision ensures that selected interactions are reliable, minimizing wasted resources in experimental validation. Conversely, high recall ensures that potential DTIs are not overlooked, a critical factor in drug repurposing and polypharmacology studies.

An interesting aspect of our findings is how the models maintained consistent performance across targets despite the class imbalance in our dataset. The underrepresented target O15151, did not suffer from significantly poorer prediction metrics. This suggests that the GNNs effectively capture the essential structural determinants of DTIs regardless of class frequencies, a promising finding for applications to rare targets or orphan diseases where interaction data may be limited.

While the results are promising, limitations must be considered. Our study focused on only five protein targets, which may not fully represent the diversity of protein families and binding mechanisms relevant to drug discovery. The binary active/inactive classification scheme, while practical, simplifies the continuous nature of binding affinities and potentially overlooks important nuances in partial agonism or allosteric modulation. Additionally, our models rely on SMILES-derived graph representations that capture 2D molecular topology but may miss critical 3D conformational features that influence binding interactions.

From a methodological perspective, our validation approach, while rigorous with cross-validation, would benefit from external validation on independent datasets to better assess real-world generalizability. The models also focus primarily on compound representations without explicitly incorporating protein structural information, limiting their ability to model specific binding site interactions. This representation gap extends to temporal aspects of DTIs, such as binding kinetics and residence time, which are increasingly recognized as important determinants of drug efficacy.

Computational and interpretability challenges also persist. While we discuss relative performance differences between models, a more comprehensive analysis of computational efficiency, including training time and inference speed across different dataset scales, would provide valuable implementation insights. Furthermore, despite GAT offering some interpretability through attention weights, explaining predictions in terms meaningful to medicinal chemists remains challenging. Addressing these limitations in future work could further enhance the practical utility of GNN approaches in drug discovery applications.

7. Conclusion

Our evaluation suggests that GNN-based approaches achieve high accuracy in DTI prediction, with GraphSAGE achieving the highest accuracy (93%) and precision (79%), while GIN demonstrated superior recall (72%). The choice between these models depends on specific application requirements: GraphSAGE excels when minimizing false

positives, which is critical in early drug discovery phases, whereas GIN offers a balanced performance ideal for drug repurposing where discovering new interactions is prioritized. Despite the class imbalance in our dataset, all models maintained robust performance, confirming the effectiveness of graph-based representations for capturing molecular interaction patterns.

Future work could focus on ensemble approaches that combine the strengths of different GNN architectures, e.g. integrating GraphSAGE's precision with GIN's recall capabilities. Incorporating protein structural information alongside molecular graphs and extending beyond binary classification to predict binding affinities is also promising. Additionally, using interpretability techniques would increase the practical utility of these models in medicinal chemistry workflows. These advancements would further establish GNNs approaches as valuable tools in computational drug discovery.

8. Acknowledgments:

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. All experiments in this work used the PCAD infrastructure, `http://gppd-hpc.inf.ufrgs.br`, at INF/UFRGS.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Cheng, Z., Yan, C., Wu, F.-X., and Wang, J. (2021). Drug-target interaction prediction using multi-head self-attention and graph attention network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(4):2208–2218.

Dandibhotla, S., Samudrala, M., Kaneriya, A., and Dakshanamurthy, S. (2025). Gnnseq: A sequence-based graph neural network for predicting protein–ligand binding affinity. *Pharmaceuticals*, 18(3):329.

Elbasani, E., Njimbouom, S. N., Oh, T.-J., Kim, E.-H., Lee, H., and Kim, J.-D. (2021). Gcrnn: graph convolutional recurrent neural network for compound–protein interaction prediction. *BMC bioinformatics*, 22(Suppl 5):616.

Hamilton, W., Ying, Z., and Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

Hasebe, T. (2021). Knowledge-embedded message-passing neural networks: improving molecular property prediction with human knowledge. *ACS omega*, 6(42):27955–27967.

Krenn, M., Häse, F., Nigam, A., Friederich, P., and Aspuru-Guzik, A. (2020). Self-referencing embedded strings (selfies): A 100% robust molecular string representation. *Machine Learning: Science and Technology*, 1(4):045024.

Li, Y., Qiao, G., Wang, K., and Wang, G. (2022). Drug–target interaction predication via multi-channel graph neural networks. *Briefings in Bioinformatics*, 23(1):bbab346.

Liu, S., Wang, Y., Deng, Y., He, L., Shao, B., Yin, J., Zheng, N., Liu, T.-Y., and Wang, T. (2022). Improved drug–target interaction prediction with intermolecular graph transformer. *Briefings in Bioinformatics*, 23(5):bbac162.

Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., and Venkatesh, S. (2021). Graphdta: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8):1140–1147.

Öztürk, H., Özgür, A., and Ozkirimli, E. (2018). Deepdta: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17):i821–i829.

Reiser, P., Neubert, M., Eberhard, A., Torresi, L., Zhou, C., Shao, C., Metni, H., van Hoesel, C., Schopmans, H., Sommer, T., et al. (2022). Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1):93.

Sachdev, K. and Gupta, M. K. (2019). A comprehensive review of feature based methods for drug target interaction prediction. *Journal of Biomedical Informatics*, 93:103159.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80.

Su, X., Hu, P., You, Z.-H., Philip, S. Y., and Hu, L. (2024). Dual-channel learning framework for drug-drug interaction prediction via relation-aware heterogeneous graph transformer. In *Proceedings of AAAI*, volume 38, pages 249–256.

Tang, X., Lei, X., and Zhang, Y. (2024). Prediction of drug-target affinity using attention neural network. *International Journal of Molecular Sciences*, 25(10):5126.

Tran, H. N. T., Thomas, J. J., and Malim, N. H. A. H. (2022). Deepnc: a framework for drug-target interaction prediction with graph neural networks. *PeerJ*, 10:e13163.

Tsubaki, M., Tomii, K., and Sese, J. (2019). Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, 35(2):309–318.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wan, W., Silva, R., Odenweller, D. J., and Leeuwon, S. (2024). 6.03 - proprietary strategies in precision medicine. In Ramos, K. S., editor, *Comprehensive Precision Medicine (First Edition)*, pages 197–220. Elsevier, Oxford, first edition edition.

Wang, H., Zhou, G., Liu, S., Jiang, J.-Y., and Wang, W. (2021). Drug-target interaction prediction with graph attention networks. *arXiv preprint arXiv:2107.06099*.

Wang, X., Wen, Y., Zhang, Y., Dai, C., Yang, Y., Bo, X., He, S., and Peng, S. (2024). A hierarchical attention network integrating multi-scale relationship for drug response prediction. *Information Fusion*, 110:102485.

Weisfeiler, B. and Leman, A. (1968). A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Technicheskaya Informatsiya*, 2(9):12–16.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. (2018). How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., and Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81.