The First International Conference on
Machine Learning and Knowledge Discovery (MLKD 2024)
Amirkabir University of Technology, December 18-19, 2024

Re-evaluation and Validation of Graph Neural Network for Predicting Drug-Target Binding Affinity *

Kowsar Ghomi† Bahram Sadeghi Bigham‡

Abstract

In recent years, the increasing complexity of drug discovery has stimulated the need for advanced computational methods that can effectively predict drug-target interactions. Many and various methods are trying to be presented to improve the problem and reduce time and cost, and among these methods, the GraphDTA method using graph neural networks has succeeded in reducing the cost and time to provide new information. In this paper, according to the dataset provided in GraphDTA, we measure the validity of this method using different evaluation criteria and prove its effectiveness. The results of applying different evaluation criteria to measure the accuracy of this method and to choose the best model in this case have been compared with each other.

Keywords: Bioinformatics, Predicting drug-target binding affinity, Graph Neural Networks

1 Introduction

The successful development of new therapeutics hinges on the ability to predict drug-target binding affinities accurately, as these interactions are critical to understanding pharmacodynamics and ligand efficacy. Traditional methods of predicting binding affinity have often relied on linear models, molecular descriptors, and empirical interactions that fail to capture the complex, non-linear relationships and conformational variations inherent in biological systems. With the advent of machine learning and artificial intelligence, there has been a significant shift towards more sophisticated approaches that can handle these complexities, especially in the realm of drug discovery.

Graph Neural Networks (GNNs) have emerged as a promising solution due to their capacity to process data structured as graphs, thereby retaining the relational information between atoms and molecular substructures. In the context of drug discovery, GNNs enable the representation of both drug compounds and target proteins as graphs, where nodes correspond to atoms (or amino acids) and edges reflect chemical bonds (or peptide links). This framework allows for a more holistic understanding of the molecular interactions that dictate binding affinities.

In this paper, we introduce GraphDTA, a dedicated framework designed to predict drug-target binding affinities using GNNs. GraphDTA not only addresses the limitations of previous predictive models but also incorporates various molecular features and leverages high-dimensional biological data more effectively. We demonstrate its capabilities on established benchmark datasets, showcasing how GraphDTA attains superior predictive performance compared to existing methods. Additionally, we explore the interpretability of the model, illuminating key factors influencing drug-target interactions.

Through this research, we aim to contribute to the field of computational drug discovery, offering a robust tool that assists researchers in the design and optimization of new therapeutics. By embedding our approach in the fast-evolving landscape of GNNs, we hope to pave the way for improved predictions and deeper insights into the drug discovery process. In the following sections, we will discuss the set of methods used and the proposed framework in the third section, the analysis of the results in the fourth section, and the discussion and conclusions.

2 Research Background

Moulard et al. estimated the cost of developing a new drug at $2.6 billion [23]. Also, Ashburn and Thor highlighted that while FDA approval for a new drug takes about 10 to 17 years, new applications for approved drugs prevent the lengthy, costly, and safety-related issues associated with drug development[2].

Deshpande et al. deemed comprehensive searches impossible due to the existence of millions of similar compounds and utilized classification algorithms to predict whether chemical compounds had desirable biological activity and to filter similar compounds from large libraries [12].

Corsello and Iskar et al. found a strong incentive

*This research has beem facilitated by the Data Science Lab of the Faculty of Mathematical Sciences at Alzahra University.

†Department of Computer Science, Faculty of Mathematical Sciences, Alzahra University, Tehran, Iran., kousar1377_18@yahoo.com

‡Corresponding autho: Department of Computer Science, Faculty of Mathematical Sciences, Alzahra University, Tehran, Iran., b_sadeghi_b@alzahra.ac.ir

to create computational models that can estimate the interaction strength of drug-target pairs based on previous assays[9, 19]. Le et al. proposed an approach that predicts the stable three-dimensional structure of the drug-target complex through a scoring function[22]. He et al. pointed out that since the molecular docking approach requires knowledge of the crystallized structures of proteins—often unavailable—they used a collaborative filtering approach, including SimBOOST, which employs similarity in binding affinities between drugs and targets to create new features as input for a gradient boosting machine to predict binding affinity for unknown drug-target pairs[17].

Cichonska et al. noted that similarities could be sourced from alternative resources rather than solely relying on experimental proximities, for instance, kernel-based methods that generate kernels from molecular descriptors of drugs and targets using regularized least squares regression (RLS). The KRonRLS method predicts relational closeness and utilizes similarity scores for each drug-target pair through a drug-drug and target-target similarity matrix, demonstrating high performance[17, 18].

Chu.Y et al. introduced a new prediction method for drug-target interactions to improve performance based on a deep forest model (CDF), named DTI-CDF, utilizing features based on multiple similarities between drugs and targets, along with proteins extracted from a heterogeneous graph containing known DTIs[5].

DTIs are a significant step for drug discovery and repositioning, and various computational methods, particularly binary classification, have been developed, though improvements are needed, which multi-label learning can facilitate to reduce the issues in binary classification performance, which is handled with the DTI-MLCD approach[6].

Ozkurt et al. proposed another approach, including DeepDTA and WideDTA, which is the model developed by DeepDTA and utilizes networks trained on one-dimensional drug representations and protein sequences. Since drugs can be represented using common substructures, Wozniak et al. suggested PADME, which is based on deep neural networks for predicting interactions between compounds and protein features effectively. Meyer et al. regarded deep learning models as the best predictors for DTA and the most successful machine learning techniques on a broad spectrum[24].

However, these models represent drugs as strings, which is not a natural way to depict molecules, leading to the loss of structural information and impairing predictive power. In this paper, we propose GraphDTA[26], capable of modeling drugs as molecular graphs that considers the DTA prediction task as a regression task, where the input is a drug-target pair, and the output is a continuous measurement of binding affinity. Among

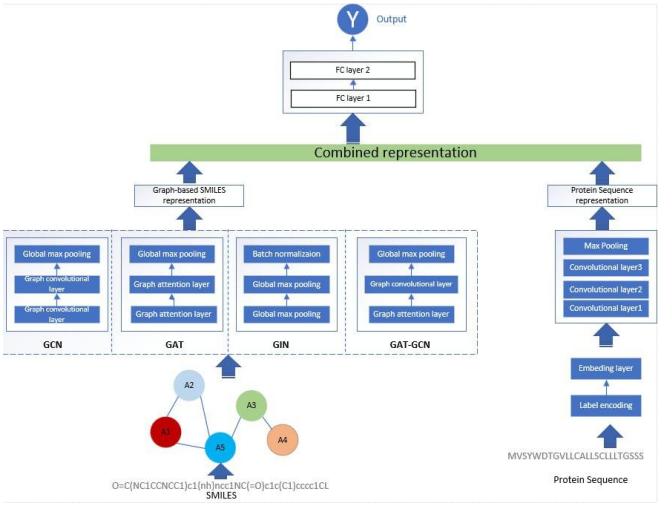

Figure 1: GraphDTA architecture[26].

all proposed models, it delivers superior and advanced performance.

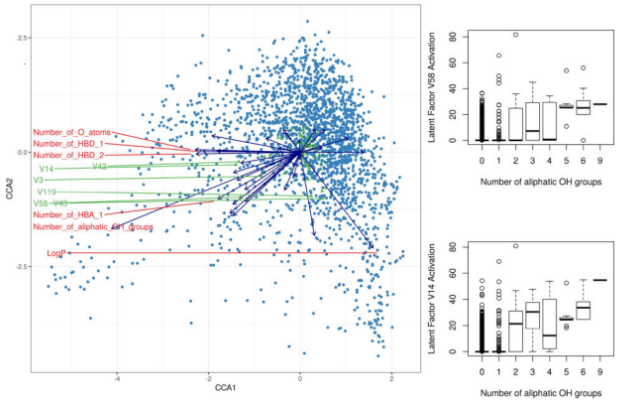

Figure 2: redundancy analysis triplot for the 128 drug latent variables regressed with 38 JoeLib molecular descriptors. (Left) the activation of two latent variables plotted against the number of aliphatic OH groups in that drug. (Right) [30]

2.1 Drug-Target Binding Affinity Prediction

Drug-target binding affinity refers to the strength of interaction between a drug molecule and its biological target, typically a protein. The binding affinity is a crucial parameter that determines the pharmacological effectiveness of a drug. Traditional approaches to predict binding affinity include molecular docking, which simulates the interaction between a drug and its target, and quantitative structure-activity relationship (QSAR) models, which correlate chemical structure with biological activity (Rogers et al., 2016). How-

ever, these methods often face limitations in terms of scalability and accuracy, prompting the exploration of advanced computational techniques[27].

2.2 Machine Learning in Drug Discovery

Machine learning has revolutionized the field of drug discovery by enabling the analysis of large datasets and the identification of patterns that may not be apparent through traditional methods. Early machine learning models for predicting binding affinity primarily relied on feature engineering, where molecular descriptors were manually extracted from chemical structures [33]. However, these approaches often struggled with the inherent complexity of molecular data, leading to suboptimal predictive performance. Recent advancements in deep learning have addressed some of these limitations by automating feature extraction and learning hierarchical representations of molecular structures. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been employed to model the sequential nature of molecular data, achieving promising results in various predictive tasks [20]. Despite their successes, these architectures often fail to capture the relational information inherent in molecular graphs, necessitating the development of more sophisticated models.

2.3 Graph Neural Networks

Graph Neural Networks (GNNs) have emerged as a transformative approach for modeling molecular structures due to their ability to represent data as graphs, where nodes correspond to atoms and edges represent bonds [3]. This representation allows GNNs to effectively capture the local and global structural properties of molecules, making them particularly suitable for tasks related to drug discovery, including binding affinity prediction.

GNNs operate through a message-passing mechanism, where information is propagated between neighboring nodes. This enables the model to learn complex interactions within the molecular graph, leading to improved predictive performance. Several studies have demonstrated the efficacy of GNNs in various drug discovery tasks. For instance, introduced a GNN framework that achieved state-of-the-art results in predicting molecular properties, highlighting the model's ability to generalize across different chemical spaces[16].

2.4 Deep learning on molecular graphs

Having the drug compounds represented as graphs, the task now is to design an algorithm that learns effectively from graphical data. The recent success of CNN in computer vision, speech recognition and natural language processing has encouraged research into graph convolution. A number of works have been proposed to handle two main challenges in generalizing CNN to graphs: (i) the formation of receptive fields in graphs whose data points are not arranged as Euclidean grids, and (ii) the pooling operation to down-sample a graph. These new models are called graph neural networks.

In this work, we propose a new DTA prediction model based on a combination of graph neural networks and conventional CNN. Figure 1 shows a schematic of the model. For the proteins, we use a string of ASCII characters and apply several 1D CNN layers over the text to learn a sequence representation vector. Specifically, the protein sequence is first categorically encoded, then an embedding layer is added to the sequence where each (encoded) character is represented by a 128-dimensional vector. Next, three 1D convolutional layers are used to learn different levels of abstract features from the input. Finally, a max pooling layer is applied to get a representation vector of the input protein sequence. This approach is similar to the existing baseline models. For the drugs, we use the molecular graphs and trial four graph neural network variants, including GCN, GAT, GIN and a combined GAT-GCN architecture, all of which we describe below[32, 33, 31].

2.4.1 Variant 1: GCN-based graph representation learning

In this work, we focus on predicting a continuous value indicatingthe level of interaction of a drug and a protein sequence. Each drugis encoded as a graph and each protein is represented as a string ofcharacters. To this aim, we make use of GCN model [32] for learning on graph representation of drugs. Notethat, however, the original GCN is designed for semi-supervisednode classification problem, i.e. the model learns the node-level feature vectors. For our goal, to estimate the drug-protein interaction, agraph-level representation of each drug is required. Common techniques to aggregate the whole graph feature from learned node features include Sum, Average and Max Pooling. In our experiments, the use of Max Pooling layer in GCN-based GraphDTA usually results in better performance compared to that of the remaining. Formally, denote a graph for a given drug as $G = (V, E)$, whereV is the set of N nodes each is represented by a C-dimensional vectorand E is the set of edges represented as an adjacency matrix A. Amultilayer graph convolutional network (GCN) takes as input anode feature matrix $X \in R^{N \times C}$ ($N = |V|$, C : the number of featuresper node) and an adjacency matrix $A \in R^{N \times N}$; then produces anode-level output $Z \in R^{N \times F}$ (F: the number of output features pernode). A propagation rule can be written in the normalized

form forstability, as in[32]:

$$H^{l+1} = \sigma(\widetilde{D}^{\frac{1}{2}}\widetilde{A}\widetilde{D}^{\frac{1}{2}}H^{(L)}W^{(L)}) \quad (1)$$

where $\widetilde{A} = A + IN$ is the adjacency matrix of the undirected graph with added self-connections, $\widetilde{D}_{ii} = \sum_i \widetilde{A}_{ii}$, $H^l \in R^{N\times C}$ is the matrix of activation in the lth layer, $H^0 = \chi, \sigma$ is an activation function, and W is learnable parameters.
A layer-wise convolution operation can be approximated, as in [32]:

$$Z = \widetilde{D}^{\frac{1}{2}}\widetilde{A}\widetilde{D}^{\frac{1}{2}}X\text{Ⓗ} \quad (2)$$

where Ⓗ $\in R^{C\times F}$ (F: the number of filters or feature maps) is the matrix of filter parameters.
Note that, however, the GCN model learns node-level outputs $Z \in R^{N\times F}$. To make the GCN applicable to the task of learning a representation vector of the whole graph, we add a global max pooling layer right after the last GCN layer. In our GCN-based model,we use three consecutive GCN layers, each activated by a ReLUfunction. Then a global max pooling layer is added to obtain thegraph representation vector.

2.4.2 Variant 2: GAT-based graph representation learning

Unlike graph convolution, the graph attention network (GAT)[33] proposes an attention-based architecture to learn hidden representations of nodes in a graph by applying a selfattention mechanism. The building block of a GAT architecture is a graph attention layer. The GAT layer takes the set of nodes of a graph as input, and applies a linear transformation to every node by a weigh matrix W. For each input node i in the graph, the attention coefficients between i and its first-order neighbors are computed as:
$a(wx_i, wx_j)$

This value indicates the importance of node j to node i. These attention coefficients are then normalized by applying a soft-max function, then used to compute the output features for nodes $\sigma(\sum_{j\in N(i)} \alpha_{ij} wx_i)$ where $\sigma(c)$ is a non-linear activation function and a_{ij} are the normalized attention coefficients.

In our model, the GAT-based graph learning architecture includes two GAT layers, activated by a ReLU function, then followed a global max pooling layer to obtain the graph representation vector. For the first GAT layer, multi-head-attentions are applied with the number of heads set to 10, and the number of output features set to the number of input features. The number of output features of the second GAT is set to 128.

2.4.3 Variant 3: graph isomorphism network (GIN)

The GIN [32] is newer method that supposedly achieves maximum discriminative power among graph neural networks. Specifically, GIN uses a multi-layer perceptron (MLP) model to update the node features as.
$mlp((1+\epsilon)x_i + \sum x_i)\ j \in B_j$

where is either a learnable parameter or a fixed scalar, x is the node feature vector and B(i) is the set of nodes neighboring i.

In our model, the GIN-based graph neural net consists of five GIN layers, each followed by a batch normalization layer. Finally, a global max pooling layer is added to obtain the graph representation vector.

2.4.4 Variant 4: GAT-GCN combined graph neural network

We also investigate a combined GAT-GCN model. Here, the graph neural network begins with a GAT layer that takes the graph as input, then passes a convolved feature matrix to the subsequent GCN layer. Each layer is activated by a ReLU function. The final graph representation vector is then computed by concatenating the global max pooling and global mean pooling layers from the GCN layer output.

2.5 GraphDTA: A Novel Approach to Drug-Target Binding Affinity Prediction

Building upon the success of GNNs, the GraphDTA framework was specifically designed for predicting drug-target binding affinities. GraphDTA integrates both drug and target information into a unified graph representation, allowing for the simultaneous modeling of both components in the binding interaction. This dual-graph approach enhances the model's ability to capture the intricate relationships between drugs and their targets, leading to more accurate predictions [26].

In GraphDTA, the drug is represented as a molecular graph, while the target protein is encoded using a sequence-based representation, such as embeddings derived from protein sequences. The model employs a multi-layer GNN architecture that iteratively updates node representations through message passing, ultimately producing a binding affinity score. The results of GraphDTA indicate significant improvements over traditional methods and even other machine learning approaches, demonstrating the potential of GNNs in drug discovery

2.6 Comparison with Existing Methods

The performance of GraphDTA has been benchmarked against various existing methods, including molecular docking and traditional machine learning models. In comparative studies, GraphDTA consistently outperformed these methods in terms of predictive accuracy and generalization to unseen data [26]. The ability of GraphDTA to leverage both structural and sequential

information sets it apart from other approaches, providing a more holistic view of the drug-target interaction landscape.

Moreover, the scalability of GraphDTA allows it to be applied to large-scale datasets, making it a valuable tool for high-throughput screening scenarios. As the availability of chemical and biological data continues to grow, the need for robust predictive models becomes increasingly critical. GraphDTA addresses this need by offering a flexible and efficient framework for binding affinity prediction.

2.7 One-hot encoding

One-hot encoding has been used in previous works to represent both drugs and proteins, as well as other biological sequences like DNA and RNA. This paper tests the hypothesis that a graph structure could yield a better representation for drugs, and so only drugs were represented as a graph. Although one could also represent proteins as graphs, doing so is more difficult because the tertiary structure is not always available in a reliable form. As such, we elected to use the popular one-hot encoding representation of proteins instead.
For each target in the experimented datasets, a protein sequence is obtained from the UniProt database using the target's gene name. The sequence is a string of ASCII characters which represent amino acids. Each amino acid type is encoded with an integer based on its associated alphabetical symbol [e.g. Alanine (A) is 1, Cystine (C) is 3, Aspartic Acid (D) is 4 and so on], allowing the protein to be represented as an integer sequence. To make it convenient for training, the sequence is cut or padded to a fixed length sequence of 1000 residues. In case a sequence is shorter, it is padded with zero values.
These integer sequences are used as input to the embedding layers which return a 128-dimensional vector representation. Next, three 1D convolutional layers are used to learn different levels of abstract features from the input. Finally, a max pooling layer is applied to get a representation vector of the input protein sequence.

2.8 Benchmark

To compare our model with the state-of-the-art DeepDTA [24] and WideDTA models, we use the same datasets from the [24] benchmarks: • Davis contains the binding affinities for all pairs of 72 drugs and 442 targets, measured as Kd constants and ranging from 5.0 to 10.8 [10]. Kiba contains the binding affinities for 2116 drugs and 229 targets, measured as KIBA scores and ranging from 0.0 to 17.2 [28]. To make the comparison as fair as possible, we use the same set of training and testing examples from [24], as well as the same performance metrics: Mean Square Error (MSE, the smaller the better) and Concordance Index (CI, the larger the better). For all baseline methods, we report the performance metrics as originally published in [24]. The hyperparameters used for our experiments are summarized in Table 1. The hyper-parameters were not tuned, but chosen a priori based on our past modeling experience.

Table 1: Hyper-parameters for different graph neural network variants used in our experime

Hyper-parameters	setting
Leraning Rate	0.0005
Batch size	512
Optimizer	Adam
GCN layers	3
GIN layers	5
GAT layers	2
GAT-GCN layers	2

Table 2: Comparison of evaluation criteria to choose the best model

Dataset name	Model	MAE	MSE	RMSE	R^2
Davis	GAT	0.3903	0.2335	0.4832	0.8992
Davis	OCN	0.3887	0.2393	0.4861	0.895
Davis	GIN	0.3899	0.2364	0.4892	0.8949
Davis	GAT-OCN	0.3932	0.253	0.503	0.8875
KIBA	GAT	0.4035	0.2499	0.4999	0.8779
KIBA	OCN	0.3959	0.245	0.4949	0.8804
KIBA	GIN	0.4002	0.2515	0.5015	0.8871
KIBA	GAT-OCN	0.3925	0.2530	0.503	0.8764

Table 3: Comparison of evaluation criteria to choose the best model

Dataset name	Model	Pearson Correlation	Max Error	Explained Variance	Spearman Correlation
Davis	GAT	0.9511	1.316	0.8967	0.9474
Davis	OCN	0.9578	1.263	0.9012	0.9574
Davis	GIN	0.9535	1.211	0.8958	0.9488
Davis	GAT-OCN	0.9521	1.448	0.8911	0.9470
KIBA	GAT	0.9424	1.094	0.8787	0.9397
KIBA	OCN	0.9535	1.539	0.886	0.9611
KIBA	GIN	0.9452	1.235	0.8779	0.9367
KIBA	GAT-OCN	0.9381	1.348	0.8774	0.9407

1. MAE (Mean Absolute Error):
MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It's the average over the test sample of the absolute differences between prediction and actual observation.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Lower values indicate a better fit. It's robust to outliers compared to MSE.
2. MSE:
evaluates the average of the squares of the errors, that is, the average squared difference between the estimated values and the actual value.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

It penalizes larger errors more than smaller ones due to squaring. Lower values are better, and it's sensitive to outliers.
3. RMSE (Root Mean Squared Error):
RMSE is the square root of MSE. It gives the error in the same units as the output variable.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Like MSE, lower values indicate a better fit, and it retains the interpretation of errors in the original units.
4. R^2 (Coefficient of Determination):
R^2 indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.

$$R^2 = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}}$$

Where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares.
Values range from 0 to 1; a higher value indicates a better model. An R^2 of 0 means the model does not explain any of the variance, while an R^2 of 1 means it explains all the variance.
5. Pearson Correlation Coefficient:
This measures the linear correlation between two variables, providing a value between -1 and 1.

$$r = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Where $Cov(X, Y)$ is the covariance of X and Y, and σ are the standard deviations.
Values close to 1 indicate a strong positive correlation, values close to -1 indicate a strong negative correlation, and 0 indicates no correlation.
6. Max Error:
This is the maximum difference between observed and predicted values.

$$\text{Max Error} = \max_i |y_i - \hat{y}_i|$$

It indicates the worst-case scenario in prediction errors; a lower value is better.
7. Explained Variance:
This measures the proportion of the variance in the dependent variable that is accounted for by the model.

$$\text{Explained Variance} = \frac{\text{Var}(\hat{y})}{\text{Var}(y)}$$

Similar to R^2, values close to 1 indicate that the model explains most of the variance, while values close to 0 indicate that it explains little.
8. Spearman Correlation:
This measures the strength and direction of the association between two ranked variables. It's a non-parametric measure.
Where d_i is the difference in ranks for each observation, and n is the number of observations.
Like Pearson, values range from -1 to 1, and it assesses monotonic relationships rather than linear ones.
GCN is the only one that had the best performance for both datasets and for both performance measures. For this reason, we focus on the GCN in all post hoe statistical analyses.
The GCN model is identified as the best because it has the highest R^2, Pearson correlation, Spearman correlation, and explaind variance, which indicate better predictive performance, along with relatively low error metrics (MAE, MSE, RMSE, and max error) compared to other models.

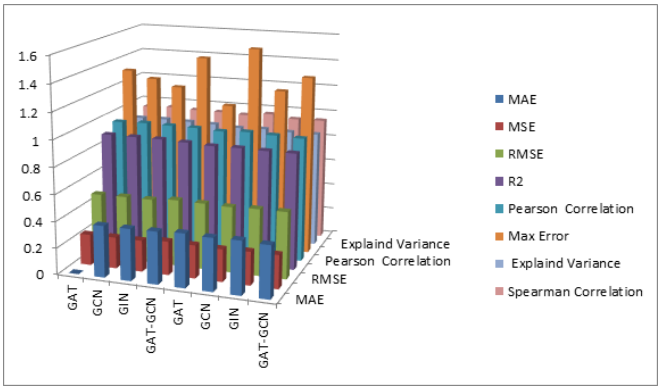

Figure 3: Comparison of evaluation criteria to choose the best model

3 Conclusion and Future Work

The prediction of drug-target binding affinity is a vital component of the drug discovery process, and Graph Neural Networks, particularly the GraphDTA framework, represent a significant advancement in this field. By leveraging the graph-based representation of molecular structures, GNNs can effectively capture the complex relationships between drugs and their targets, leading to improved predictive performance. While challenges remain, the potential of GNNs to revolutionize drug discovery is evident, paving the way for more efficient and accurate predictions in the quest for novel therapeutics.
We presented GraphDTA, a state-of-the-art framework

that leverages graph neural networks to predict drug-target binding affinities with enhanced accuracy and efficiency. By transforming drug compounds and target proteins into graph representations, GraphDTA effectively captures the complex interactions and structural nuances that influence binding affinity. Our extensive experiments on benchmark datasets have validated the model's superior performance compared to traditional methods, highlighting its potential to significantly advance computational drug discovery.

The interpretability features of GraphDTA also offer valuable insights into the molecular features and interactions that underlie binding affinities, making it a dual-purpose tool for both prediction and analysis. This capability can aid researchers not only in identifying promising drug candidates but also in elucidating the mechanisms of drug action, which is crucial for optimizing therapeutic efficacy and minimizing side effects.

Looking to the future, several avenues for further development and enhancement of GraphDTA can be explored. Firstly, we aim to expand our model to integrate additional biological data, such as genomic information and epigenetic factors, which could further enhance predictions and provide a more comprehensive view of drug-target interactions. Additionally, the incorporation of dynamic molecular simulations could allow GraphDTA to account for conformational changes in proteins and ligands over time, potentially leading to even more accurate affinity predictions.

Furthermore, we plan to investigate the application of GraphDTA in multi-target scenarios and polypharmacology, where drugs interact with multiple targets. This expansion could facilitate the design of new therapeutic agents that consider complex biological networks and disease pathways. Lastly, we intend to make GraphDTA accessible to the broader scientific community through an open-source platform, encouraging collaborative efforts to refine and adapt the model for diverse applications in drug discovery.

GraphDTA represents a significant step forward in the predictive modeling of drug-target interactions, and its ongoing development will contribute to the evolving landscape of computational pharmacology, ultimately fostering deeper insights and more effective therapeutic solutions. For future research, it is necessary to use a larger amount of data that can be effective in them. In this way, a more comprehensive system can be used to predict alternative medicine for the treatment of diseases according to a set of factors and data. In addition, by providing a method for drug recommendation in the proposed framework, its efficiency can be increased.

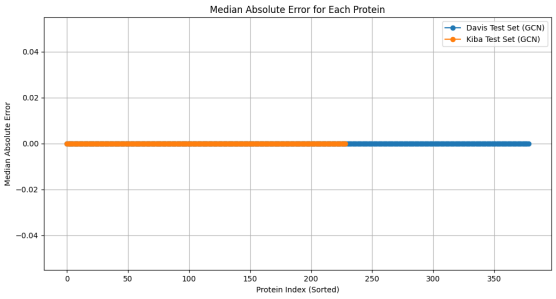

Figure 4: The errors are not distributed evenly across the proteins, indicating that it is harder to predict the target affinities for some proteins than others.

References

[1] Ali,M. et al. (2017) Global proteomics profiling improves drug sensitivity prediction: results from a multi-omics, pan-cancer modeling approach. *Bioinformatics*, 1, 10.

[2] Ashburn,T.T. and Thor,K.B. (2004) Drug repositioning: identifying and developing new uses for existing drugs. Nat. Rev. Drug Disc., 3, 673–683.

[3] Battaglia, P. W., et al. (2018). "Relational inductive biases, deep learning, and graph networks." arXiv preprint arXiv:1806.01261

[4] Backman,T.W. et al. (2011) ChemMine tools: an online service for analyzing and clustering small molecules. Nucleic Acids Res., 39, W486–W491.

[5] Chu,Y. et al. (2019) DTI-CDF: a cascade deep forest model towards the prediction of drug-target interactions based on hybrid features. Brief. Bioinf., 1-12, 10.1093/bib/bbaa205.

[6] Chu,Y. et al. (2020) Predicting drug-target interactions using multi-label learning with community detection method (DTI-MLCD). bioRxiv.

[7] Cichonska,A. et al. (2017) Computational-experimental approach to drug–- target interaction mapping: a case study on kinase inhibitors. PLoS Comput. Biol., 13, e1005678.

[8] Cichonska,A. et al. (2018) Learning with multiple pairwise kernels for drug bioactivity prediction.Bioinformatics, 34, i509–i518.

[9] Corsello,S.M. et al. (2017) The Drug Repurposing Hub: a next-generation drug library and information resource. Nat. Med., 23, 405–408.

[10] Davis,M.I. et al. (2011) Comprehensive analysis of kinase inhibitor selectivity. Nat. Biotechnol.,29, 1046–1051.

[11] Dahl, G. E., et al. (2019). "Deep learning for drug discovery: a review." Journal of Chemical Information and Modeling, 59(4), 1230-1240.

[12] Deshpande,M. et al. (2005) Frequent substructure-based approaches for classifying chemical compounds. IEEE Trans. Knowl. Data Eng., 17, 1036–1050..

[13] Feng,Q. et al. (2018) PADME: a deep learning-based framework for drug-target interaction prediction. arXiv, (arXiv : 1807.09741).

[14] Gao,H. et al. (2018a) Large-scale learnable graph convolutional networks. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1416–1424.

[15] Gordon,D.E. et al. (2020) A SARS-CoV-2-human protein-protein interaction map reveals drug targets and potential drug-repurposing. bioRxiv. doi: 10.1101/2020.03.22.002386.

[16] Gilmer, J., et al. (2017). "Neural Message Passing for Quantum Chemistry." arXiv preprint arXiv:1704.01212

[17] He,T. et al. (2017) SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. J. Cheminf., 9, 24.

[18] Hirohara,M. et al. (2018) Convolutional neural network based on SMILES representation of compounds for detecting chemical motif. BMC Bioinformatics, 19, 526

[19] Iskar,M. et al. (2012) Drug discovery in the age of systems biology: the rise of computational approaches for data integration. Curr. Opin. Biotechnol., 23, 609–616

[20] Kearnes, S. et al. (2016). "Molecular graph convolutions: moving beyond fingerprints." Journal of Computer-Aided Molecular Design, 30(8), 595-608.

[21] Kinnings,S.L. et al. (2011) A machine learning-based method to improve docking scoring functions and its application to drug repurposing. J. Chem. Inf. Model., 51, 408–419.

[22] Le,V. et al. (2020) Deep in the bowel: highly interpretable neural encoder-decoder networks predict gut metabolites from gut microbiome. BMC Genomics., 21, 256.

[23] Mullard,A. (2014) New drugs cost US $2.6 billion to develop. Nat. Rev. Drug Disc., 13, 877–877

[24] O¨ ztu¨rk,H. et al. (2018) DeepDTA: deep drug–target binding affnity prediction. Bioinformatics, 34, i821–i829.

[25] Strittmatter,S.M. (2014) Overcoming drug development bottlenecks with repurposing: old drugs learn new tricks. Nat. Med., 20, 590–591.

[26] Sung, J., et al. (2020). "GraphDTA: Predicting drug-target binding affinity with graph neural networks." Bioinformatics, 36(1), 40-49.

[27] Rogers, D., et al. (2016). "Chemical informatics functionality in Python." Journal of Chemical Information and Modeling, 55(2), 249-253.

[28] Tang,J. et al. (2014) Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. J. Chem. Inf. Model., 54, 735–743.

[29] Wang, Y., et al. (2021). "A comprehensive review on the application of deep learning in drug discovery." Artificial Intelligence in Medicine, 113, 101026

[30] Weininger,D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci., 28, 31–36.

[31] Wozniak,M. et al. (2018) Linguistic measures of chemical diversity and the 'keywords' of molecular collections. Sci. Rep., 8.

[32] Xu,K. et al. (2019) How powerful are graph neural networks? In Proceeding of the International Conference on Learning Representations.

[33] Xiong, Z., et al. (2019). "Pushing the boundaries of molecular representation for drug discovery." Journal of Medicinal Chemistry, 62(8), 3953-3964.