GraMDTA: Multimodal Graph Neural Networks for Predicting Drug-Target Associations

Jaswanth K. Yella
University of Cincinnati
Department of Electrical Engineering and Computer Science
OH, USA, *Division of Biomedical Informatics*
OH, USA
yellajk@mail.uc.edu

Sudhir K. Ghandikota
University of Cincinnati
Department of Electrical Engineering and Computer Science
OH, USA, *Division of Biomedical Informatics*
OH, USA
ghandisv@mail.uc.edu

Anil G. Jegga
Cincinnati Children's Hospital Medical Center
Division of Biomedical Informatics
OH, USA, University of Cincinnati
Department of Pediatrics,
OH, USA
anil.jegga@cchmc.org

Abstract— **Finding novel drug-target associations is vital for drug discovery. However, screening millions of small molecules for a select target protein is challenging. Several computational approaches have been developed in the past using Machine learning methods for computational drug-target association (DTA) prediction predominantly use structural data of drugs and proteins. Some of these approaches use knowledge graph networks and link prediction. To the best of our knowledge there have been no approaches that use both structural learning that offers molecular-based representations and knowledge graph-based learning which offers interaction-based representations for DTA discovery. Based on the premise that multimodal sources of information acting complimentarily could improve the robustness of DTA predictions, we developed GraMDTA, a multimodal graph neural network that learns both structural and knowledge graph representations utilizing multi-head attention to fuse the multimodal representations. We compare GraMDTA with other computational approaches for DTA prediction to demonstrate the power of multimodal fusion for discovery of DTA.**

Keywords—Graph Neural Networks, Multimodal, Drug Discovery, Attention, Virtual Screening

INTRODUCTION

The drugs approved for various indications often have unintended activities or drug-induced adverse events (AEs). This is often attributed to a drug's off-target effect, binding other known or unknown targets. While such off-target binding is one of the underlying causes for several drug-induced AEs, they can also be beneficial and may lead to drug repositioning or discovery of new indications for an approved drug. One of the important goals in drug discovery is to identify a drug lead that has minimal off-target effect with high binding affinity to selected target. Virtual screening is used for faster and economical discovery of potential drug-target associations (DTA), where computational screening of chemical libraries is performed using a target of interest.

In recent years, machine learning-based approaches are increasingly used for predicting DTA [1]. These methods typically compute drug and target similarities to predict the association or binding affinity scores. Recent advances in deep learning have enabled extraction of latent features directly from raw data. Previously to train deep learning models for predicting DTAs, the drug and target inputs, which are represented as SMILES[1] and FASTA[2] sequence respectively, are one-hot encoded. For instance, DeepDTA [2] and

DeepConv-DTI [3] applied 1D convolution to the 1-hot encoded sequences and aggregate potential patterns from raw sequence information to predict drug-target affinity [2].

Text-based representations are fixed and are inflexible for additional atom or bond level features. Whereas graphs are flexible to incorporate such auxiliary information. Recently, graph neural networks have demonstrated their effectiveness for computational drug discovery [4]–[6]. Hence as an alternative, representing drugs as molecular graphs have been proposed for feature rich molecule representations. GraphDTA [7] and Graph-CPI [8], for example, have leveraged graph neural networks to learn the graph representation of molecules along with their additional atom and bond features [9], [10] for DTA predictions. Alternatively, DTA problem can also be considered as a link prediction task utilizing the bi-partite drug-target association network and their heterogeneous annotations. Through integration of various heterogeneous data, several methods have been proposed for drug-target relationship discovery [11]–[14]. NeoDTI, for instance, integrated heterogeneous network data from diverse data sources and utilized GraphSAGE [15] for predicting DTAs.

In majority of these approaches, drugs and target proteins are represented either as text or graph modality to learn their respective representations. In other words, the drug embeddings are learnt through text, molecular graphs or through its heterogeneous neighbors using graph neural networks. Similarly, protein embeddings are learnt through its FASTA-text sequence representation using convolutional nets or through its heterogeneous neighbors using graph neural networks. Most of these approaches however rely on single modal representation of drugs and proteins. Multi-modal representation learning on the other hand offers robustness when any of the modality is incomplete or corrupted and further the modalities act complementary to each other [16]. Hence, we report **Gra**ph Neural Networks for **M**ultimodal biomedical datasets to predict **D**rug-**T**arget **A**ssociation (GraMDTA).

We collect structural and knowledge graph associations of drugs and proteins from various sources. We harmonize the collected modalities of the data. Our approach involves a pretraining phase where we train graph neural network on knowledge graph. Next, we incorporate the knowledge graph-based embeddings in GraMDTA. Further, for drugs

---

[1] Drugs are represented as SMILES (Simplified Molecular Input Line Entry System) which translates a 3-dimensional structure into a string of symbols.

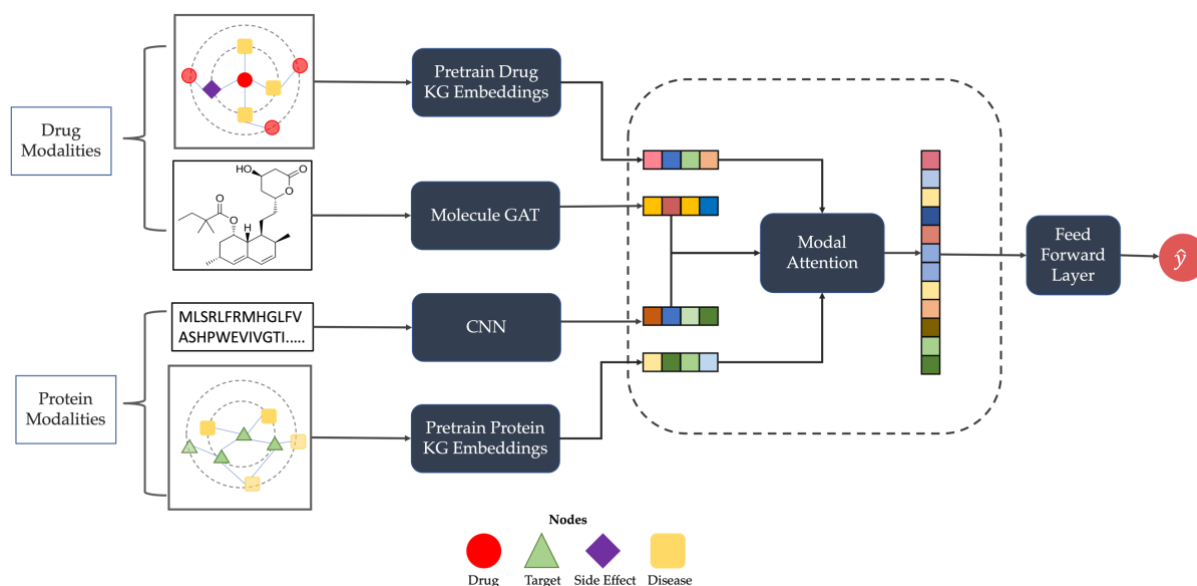[2] Proteins are represented as FASTA sequences (amino acids) in a text-based format.

Figure 1: Overview architecture of GraMDTA for predicting association between drug and target protein using multiple modalities of the entities.

represented as molecular graphs, we use graph attention network. For proteins represented as FASTA sequence, we use convolutional neural network for learning representations. Therefore, unlike previous works [2], [3], [7], [13], [17], [18], GraMDTA learns representations for drugs and proteins from structures and their corresponding knowledge graphs. We aggregate the multiple modalities of drugs and proteins using multi-head attention weighting mechanism to learn relevant information while eliminating the noisy information cascades. To assess the effectiveness of our model, we use it for drug target association prediction and drug target affinity score prediction problems by postulating the model as classification and regression methods. We perform systematic analysis to test the robustness of our framework on benchmark datasets through ablation studies. Our performance comparison with baseline models and ablation studies suggests that, multimodal learning has significant advantage over single modal learning. Our results from case study on select druggable target proteins from the Illuminating the Druggable Genome (IDG) [19], [20] database highlights the translational utility of GraMDTA.

MATERIALS AND METHODS

In this section, we describe our collected datasets i.e., drug-target association dataset, chemical structure dataset, protein sequence dataset, and relevant heterogeneous networks. Then we explain the components involved in pretraining and GraMDTA.

Datasets

Pretraining Datasets: Our pretraining datasets consists of knowledge graphs of drugs and proteins. We collected and normalized the heterogeneous associations of drugs and targets from publicly available datasets as described previously [11]. More specifically, we collected associations for all the heterogeneous entities from datasets such as DrugBank [21], RepoDB [22], DisGeNET (curated) [23], DrugCentral [24], STRING [25] and Pharos[19]. For entity normalization, we used DrugBank identifiers for drugs, UMLS concepts for diseases, UniProt identifiers for targets, and MESH identifiers for drug's categories (Table 1).

Table 1: Knowledge graph information for drugs and target knowledge graph datasets.

Dataset	Adjacency	Size	Edges
Drug KG	Drug-Diseases	6647 x 2958	8,957
	Drug-Categories	6647 x 2124	35,277
Target KG	Disease-Targets	5442 x 4294	32,379
	Target-Target (PPI)	4294x4294	13,889

Benchmark Datasets: The benchmark datasets consist of known DTAs from 3 resources, namely, Drugbank [21], KIBA [26], and DAVIS [27]. The Drugbank dataset consists of drug-target binary associations. Out of 35,022 interactions in Drugbank, 17,511 interactions are positive and remaining are random negative interactions. KIBA and DAVIS datasets consists of affinity scores between drug-targets (Table 2). As part of pre-processing, we use RDKit software [28] and convert the SMILES representation of drugs to molecular graph representation where nodes represent atoms and edges represent bonds. The features which characterize the atoms and bonds are encoded in one-hot fashion similar to [29] and are used along with the molecular graph. For protein sequence, we one-hot encode the FASTA sequence text similar to previous works [2], [3], [7].

Table 2: Benchmark datasets and their statistics

Dataset	#Drugs	#Targets	#Interactions
Drugbank	6655	4294	35,022
KIBA	2068	225	116,350
DAVIS	68	379	25,772

Pretraining

The pretraining stage consists of graph encoding and edge decoding phases. In the graph encoding phase, the feature information is aggregated from node neighborhood using graph neural network. In edge decoding phase, we perform

link prediction based on the learnt node representations. In Figure 2, we show the encoding and decoding phases for drug and protein KG as generic representation.

Graph Encoding: In the encoding phase of pretraining, we use GraphSAGE for learning node representation. GraphSAGE aggregates the features of target node neighborhoods and concatenates the node's current representation along with the aggregated neighborhood vector. This way the algorithm iteratively updates the node representation as:

$$h'_u = W_1 h_u + W_2 . Agg_{v \in N(u)}(h_v)$$

where $N(u)$ is the neighborhood of node of u, W_1 and W_2 are the weights associated to learn representation, and Agg(.) is the aggregate function such as mean or sum. We use mean aggregator in our implementation.

Similarly, we use CNN to encode the drug or protein sequence i.e., molecule or FASTA sequence using 1D convolutions. We use adjacently with knowledge graph encoding to alleviate the representation issue for isolated nodes in knowledge graph. For drugs or proteins that lack interactions, structural information is used and representations are generated. Our convolutional encoding mechanism is similar to protein sequence encoding mechanism discussed in the GraMDTA section.
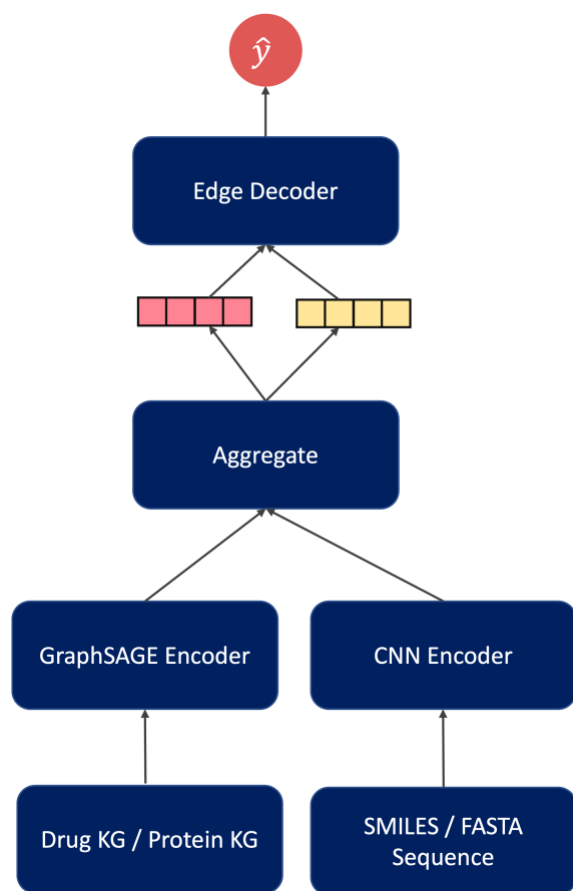

Figure 2: Overview architecture of pretraining model for training the knowledge graph

Edge Decoding: Here we define the edge decoding where we concatenate the representations of nodes and use feed forward layer to predict the link between them.

$$\hat{y} = f(h_x \oplus h_y) \qquad (1)$$

where $\oplus$ indicates concatenation of node embeddings h_x and h_y, and $f(.)$ is feed-forward neural network. We apply sigmoid activation to the output i.e., $\sigma(\hat{y})$, to predict the probability of link being associated between two nodes. We train the network using binary cross entropy loss function for both drug and protein knowledge graph datasets. The embeddings of drug and protein are extracted to further train with GraMDTA.

GraMDTA

GraMDTA has three types of encoding networks with each encoder corresponding to a modality. The molecular graphs are represented as homogeneous networks; protein FASTA sequence is represented as text; and drug and target-related annotations as knowledge graph embeddings of heterogeneous networks (Figure 1). We define graph neural networks and convolutional neural net encoders according to each modality and then aggregate the encoded representations through a multi-headed attention mechanism. Further, we perform classification on the aggregated embeddings predicting DTA as probabilities or affinity score based on benchmark dataset. In the following sections, we define the encoders and corresponding loss function.

Encoding Molecular Graphs

Molecular graphs are homogeneous networks where each node belongs to same node type (atoms) and edge type (bonds). To learn homogeneous graphs using neural networks, we define the GNN notations. A graph $G = (V, E)$, represented as an adjacency matrix A, consists of $n \in V$ nodes and $m \in E$ edges. The sparse adjacency A can be preprocessed by converting to normalized Laplacian $\hat{A}$ which is defined as a symmetric positive semidefinite matrix as $\hat{A} = I - D^{-1/2} A D^{-1/2}$. We use graph attention network (GAT) [30] to learn molecular representation which applies linear transformation for every node by a weight matrix W. Then for each node, attention coefficients are computed using their first-order neighbor nodes as follows

$$e_{ij} = a(W mol_i, W mol_j) \qquad (2)$$

Where i is the target node index, mol_i is the target node (atom) embedding of graph mol, mol_j is the neighbor node (atom), a is a function to compute attention coefficients, and e_{ij} is node j's influence over node i which quantifies the importance of the relationship. The scores are normalized using a SoftMax function as

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})} \qquad (3)$$

Using the normalized attention scores, node embeddings are computed using a non-linearity function σ as follows:

$$\widehat{mol_\iota} = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W mol_j\right) \qquad (4)$$

In this work, we use two layers of $GAT(.)$ to generate graph embedding $\widehat{mol}$. We further use a rectified linear unit (ReLU) as an activation function over the learnt embedding. The overall molecular graph embedding is simplified in the following notation

$$\widehat{mol} = ReLU(GAT(GAT(h))) \qquad (5)$$

Encoding FASTA Sequences

For 1D protein sequence, we one-hot encode and served as input for 1D-convolutions. The 1D convolutions are efficient in learning important local patterns within the sequence [2], [7]. The filters of convolutions slide through protein sequences effectively capturing relationships of an amino acids group. We denote the learnt representations of protein sequence (p) using convolution as

$$\hat{p} = ReLU\left(Linear\big(conv(p)\big)\right) \qquad (1)$$

Where $conv(.)$ is the convolutional operator, $Linear(.)$ is the fully connected linear layer, and $ReLU(.)$ is the activation function.

Encoding Knowledge Graphs

The pretrained embeddings of knowledge graph are fed to feed forward neural network as:

$$\widehat{k_d} = ReLU\big(f(k_d)\big) \qquad (6)$$

$$\widehat{k_p} = ReLU\big(f(k_p)\big) \qquad (7)$$

where k_d and k_p are knowledge graph-based representations of drugs and proteins, respectively.

Modal Attention – Fusing Multiple Modalities

Here we discuss the modal attention to fuse multiple modalities. Before we dive into the details of fusing modalities, we first summarize the notations of each modality. For a drug, there are two modalities i.e., molecular graph and heterogeneous network associations. The molecular graph representation of the drug is denoted $\widehat{mol}$ and the knowledge graph representation of the drug is represented as $\widehat{k_d}$. Similarly, the protein sequence has two modalities i.e., protein FASTA-text sequence and heterogeneous network associations. The protein sequence representation is denoted as $\hat{p}$ and the knowledge graph representation of protein is denoted $\widehat{k_p}$.

We first concatenate the multiple modal representations of drugs and proteins. Then we use multi-head attention akin to attention in sentence transformers [31] where each modality acts as a word and the concatenation of the modalities is a sentence. With this analogy, we represent the modalities to multi-head attention as

$$z = MHA(\widehat{mol} \oplus \widehat{k_d} \oplus \hat{p} \oplus \widehat{k_p})$$

where $\oplus$ indicates concatenation and MHA(.) is multi-head attention which learns key features within and across modalities from different representation subspaces. Given each input as query Q, key K, and value V, the single head attention is defined as

$$attention(Q, K, V) = softmax\left(\frac{QK^t}{\sqrt{d}}\right)V$$

Since multi-head attention computes k attention heads, the final multi-head attention would be concatenation of all the attention heads $(\alpha_1 \dots \alpha_k)$ with trainable parameter (W) i.e., $W.(\alpha_1 \oplus \alpha_2 \oplus \dots \oplus \alpha_k)$. We feed the final MHA(.) output i.e., $\hat{z}$ to 3-layer feed forward neural net along with dropout and ReLU activation function to predict the output.

Loss Functions

Based on the task at hand, we define loss function for the benchmark datasets. For classification task using Drugbank dataset, we use cross entropy loss which is given as

$$L(\hat{y}, y) = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

For regression task using KIBA and DAVIS datasets, we use mean squared error loss i.e.,

$$MSE(\hat{y}, y) = \frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2$$

EXPERIMENTAL SETUP AND RESULTS

In this section, we perform experiments on our collected benchmark datasets and compare with baseline models.

Hyperparameter setting: Our GraMDTA model is implemented in PyTorch [32]. We use scikit to split the datasets into training, test, and validation sets. We consistently used the same sets to train baselines and our model for fair comparison. Additionally, we used Adam optimizer [33] where the learning rate is varied using cosine annealing with learning rate scheduler. We set the initial learning rate (η) set to 0.01. We trained the model for 1000 epochs and used early stopping when F1 score did not improve for 10 consecutive epochs. We trained our baselines similarly during comparison.

Up-sampling Drugbank Dataset: While the associations between the drug and targets constitute the positive set, there is no negative set. To address the negative set issue, we pair random drug-target associations as negative set, similar to previous works [3], [7], [17]. This leads to a high-imbalance dataset depending on number of random parings being considered for each drug or target. Due to the high-imbalance of the associations, we limit to construct three datasets based on the up-sample ratios i.e., 1:1, 1:5, and 1:10. The 1:1 ratio has equal number of positives and negative associations where positive means there exists a drug-target association and negative association means no association between them. Then, we construct 1:5 where for every one positive association of a drug or target, we perform negative sampling of 5 associations with random pairing of target or drug respectively. We make sure that the randomly sampled drug-target association does not have any association across the training, validation, and test sets. Similarly, we construct a dataset having a ratio of 1:10 negative sampling where for one positive association we sample 10 negative samples.

Table 3: Classification performance comparison on Drugbank dataset

Model	1:1			1:5			1:10		
	AUPR	AUROC	F1	AUPR	AUROC	F1	AUPR	AUROC	F1
DeepDTA	0.8439	0.8462	0.7815	0.6826	0.8759	0.641	0.5621	0.8666	0.5479
GraphDTA - GCN	0.773	0.8131	0.776	0.6497	0.8577	0.6172	0.5626	0.8687	0.5623
GraphDTA - GAT	0.8114	0.8312	0.7752	0.6773	0.8735	0.647	0.591	0.8837	0.62
GraphDTA – GIN	0.783	0.8101	0.7666	0.5757	0.8313	0.6566	0.5232	0.843	0.5428
HyperAttention	0.8766	**0.8802**	0.8099	0.7398	**0.9079**	0.6946	0.6706	0.9054	0.6386
GraMDTA *(Ours)*	**0.8807**	0.8767	**0.803**	**0.7595**	0.9041	**0.7099**	**0.7245**	**0.9163**	**0.6861**

Baselines

We compare our results with existing works with all the aforementioned experimental settings. Following are our baseline models:

DeepDTA [2]: A convolutional neural net model for learning SMILES and FASTA sequences which are one-hot encoded. In their methodology, they concatenate the representations of drug and target, then use fully connected layers to learn to predict the association.

HyperAttentionDTI [17]: This work learns representations using convolutions and applies attentions within and across drug and target sequences.

GraphDTA [7]: This work utilizes graph neural networks for molecules by representing SMILES as molecular graph and apply convolutions on FASTA sequence. Variants of GraphDTA are proposed using graph convolutions [34], graph attention [30], and graph isomorphism network [9] to aggregate the representations of molecular graph. In our experiments, we use both the variants to compare as baselines. Fully connected layers are further used to concatenate molecule and protein sequence embeddings to predict association.

Results

We evaluate our results on benchmark datasets i.e., Drugbank, KIBA and DAVIS. We segregate the results in to two task types i.e., classification and regression tasks.

To quantitatively evaluate the models, we adopt the metrics based on the task. For classification task, we use the standard metrics such as precision-recall (AUPR), area under receiver operating characteristic curve (AUROC) and F1 scores. To compute F1-score, we identify optimal threshold through based on elbow method of precision-recall curve. For regression task, we compute mean squared error (MSE), and concordance index (CI).

Task 1 – Classification on Drugbank dataset: We conduct experiments with our model using three up-sampling ratios as three experiment settings and compared results with the baseline models. The results in Table 3 show that, with increasing negative sampling size, the performance of the models declines. This suggests that performance of all models is affected with data imbalance. GraphDTA performed consistently better than baseline models with increase in negative sample ratio. Our model achieved superior AUPR and F1 scores across all the sampling ratios. Although, GraMDTA is marginally inferior with respect to

AUROC in 1:1 and 1:5 sampling ratios, our model showed improvement over HyperAttentionDTI at 1:10 sampling. This demonstrates the model robustness with increase in negative samples.

Task 2 – Regression on KIBA and DAVIS datasets: Similar to Task 1, we conduct experiments evaluating the KIBA and DAVIS datasets as regression task. In Table 4, we show the performance results on both the datasets. With respect to KIBA dataset, GraMDTA achieved superior performance compared to all the baseline models. Our model achieved over 33% reduction in MSE and 6% improvement in CI when compared to the second-best reported work i.e., GraphDTA – GAT model. However, with respect to DAVIS dataset, HyperAttentionDTI had marginally improved performance with respect to MSE and CI. This suggests that there is room for improvement with respect to pretraining graph neural networks and GraMDTA encoding choices.

Ablation study: We conduct ablation studies on GraMDTA to understand the effectiveness and contributing components of the model by removing or substituting parts of the model. For the following ablation studies, we use 1:10 negative sampling from Drugbank dataset. The first row in Table 5, is the baseline score achieved from training Drugbank data with 1:10 negative sampling.

Table 4: Regression performance comparison on KIBA and DAVIS datasets (MSE: mean squared error; CI: concordance index)

Model	KIBA		DAVIS	
	MSE	CI	MSE	CI
DeepDTA	0.149	0.852	0.151	0.919
GraphDTA - GCN	0.19	0.823	0.192	0.901
GraphDTA - GAT	0.103	0.88	0.166	0.915
GraphDTA - GIN	0.214	0.811	0.166	0.923
HyperAttention	0.231	0.81	**0.108**	**0.947**
GraMDTA	**0.069**	**0.933**	0.113	0.942

First, we test our model by removing knowledge graph modalities of drug and target. We train the model with molecular graphs and FASTA sequence. This is equivalent to GraphDTA – GAT model along with modal attention component. The results in Table 5 suggest that removing knowledge graph embeddings (without KG) from the model

led to 3% decrease in performance. Despite the removal of pretrained KG embeddings, the performance of GraMDTA without KG is still 3% better than GraphDTA – GAT, and 2% better than HyperAttentionDTI. This suggests that multi-head based modal attention have promising impact in learning representations.

Second, we remove the non-KG components i.e., molecular graph representation and sequence-based representation. While preserving modal attention component, we train the model using only pretrained KG embeddings. As shown in Table 5, without non-KG components, the performance of the model goes down by about 20%. This suggests that molecular graphs and protein sequence information are extremely vital for our GraMDTA model performance.

Next, we delete the modal attention component from GraMDTA and train the network. Although there is a marginal decline, our model performed better than the rest of the baseline models even after removing modal attention. Thus, demonstrating that attention improves the GraMDTA performance.

Finally, we change the pretraining encoder to Graph Attention Network (GAT) and pretrain the model for drug and protein knowledge graphs. We extract the embeddings and train GraMDTA with new embeddings. With GAT pretraining, there is 2% decrease in performance. Thereby, our results from ablation studies suggest that multimodal learning

Table 5: Ablation results on GraMDTA

Model	F1 score
GraMDTA	0.686
without pretrained-KG weights	0.653
with pretrained-KG weights only	0.480
Without Modal Attention	0.66
Pretraining GAT + GraMDTA	0.665

complements with structure-based learning i.e., SMILES and FASTA sequences, and further improves the performance of the model.

Interpreting Modal Attention: We extract modal attention for naltrexone (FDA-approved drug used to prevent relapses into alcohol or drug abuse) and its predictions. Our choice of the drug is arbitrary, and the goal of this analysis is to understand the attention contribution across multiple modalities. In Figure 3, 'Molecule' represents the molecular graph representation, 'Drug KG' represents drug KG representation, 'FASTA seq' represents sequence representation of protein, and 'Protein KG' represents protein KG representation. Each row in the subplot represents the attention scores computed through SoftMax normalization. The attention score suggests the importance of knowledge graph especially protein knowledge graph being consistently important for naltrexone and its protein associations. Similarly, molecular self-attention followed by FASTA sequence attention has shown to be consistently important. As seen in the figure, the attention scores for targets OPRD1, OPRM1, and OPRK1 have similar heatmaps suggesting identical binding associations. To validate this, we ran
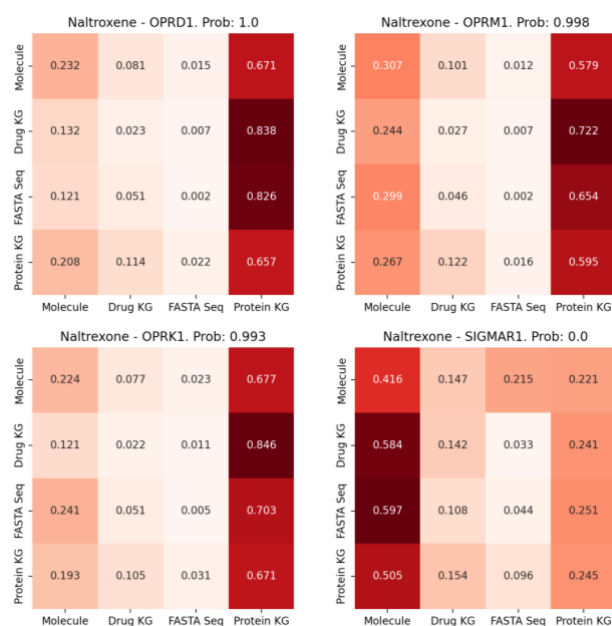

Figure 3: Attention scores for Naltrexone drug and its corresponding proteins ORPD1, OPRM1, OPRK1, and SIGMAR1. 'Prob.' in the subplot title indicates probability of drug and protein association.

enrichment analysis on these genes using ToppGene Suite [35]. We found that the selected genes are enriched (p-value<0.01; FDR Benjamini and Hochberg) for gene ontology functions targeting opioid receptor activity (G-protein-coupled opioid receptor activity G-protein-coupled opioid receptor signaling pathway (GO:0038003) suggesting that GraMDTA can identify binding capacity for homologous structures and predict efficiently. This could be in part due to the high protein KG attention with each modality for the targets. On the other hand, although naltrexone and target SIGMAR1 have high molecule attention, the predictive probability is zero. This suggests that although attention signifies the importance of the input modalities, it alone may not be enough for predicting DTAs or potential mechanism of action [36], [37].

CASE STUDY

To demonstrate the translational utility of GraMDTA, we selected minoxidil drug and its target predictions as our case study. Minoxidil was initially developed to treat high blood pressure or hypertension [38]. Later, the drug was repositioned to treat hair loss or alopecia [39]. In Figure 4, we show the network visualization of minoxidil predictions consisting of targets which are both known and predictions. We performed enrichment analysis on targets predicted with high confidence (probability score >= 90%) using ToppGene suite [35]. Our results from ToppGene enrichment show that GraMDTA-predicted targets not only recover known indications but also potential candidates for repositioning. For example, enrichment analysis of GraMDTA-predicted minoxidil targets showed hypertension and alopecia, known and repositioned indications of minoxidil, respectively. Additionally, among the target functional enrichment are other diseases or biological processes such as arterial stiffness, breast cancer, and cognition suggesting the repositioning potential of minoxidil for other diseases. Indeed, literature review showed published studies in support of these. For instance, a recent study using breast cancer cells, reported that minoxidil, a known potassium channel opener, could inhibit cellular invasiveness in breast cancer. This study also reported

that in combination with ranolazine, another ion channel blocker, minoxidil was synergistically effective as an anti-metastatic agent [40]. Another study on minoxidil revealed that the drug helps in reducing vascular or arterial stiffness

Figure 4: ToppGene enrichment analysis of targets predicted for minoxidil by GraMDTA. The rectangle nodes indicate indications (or diseases). The oval nodes indicate genes which are predicted with high probability (≥90% probability score). Pink-colored ovals are known targets of minoxidil. Approved known and repositioned indications (Hypertensive disease and alpoceia) are highlighted in pink and orange colors respectively. The other purple-colored rectangles represent potential novel indications (breast cancer, arterial stiffness, and cognitive disorders).

[41], known to be associated with early-onset cognitive impairment and dementia [42], [43]. Likewise, modulation of potassium channels is known to play an important role in the regulation of memory processes [44]. Although, this example demonstrates the utility of GraMDTA for discovery of potential drug repositioning opportunities, it should be noted that mere enrichment of disease or phenotype does not always suggest a novel indication and may also suggest potential drug-induced adverse events. For instance, while multiple evidence suggest benefits of minoxidil in cognitive disorders, there are also conflicting reports in the literature reporting that minoxidil may cause cognitive impairment issues [45]. Nevertheless, the discovery of novel drug-phenotype associations through prediction of DTAs using GraMDTA and similar approaches are powerful in formulating translation hypotheses for drug repositioning or characterizing drug-induced adverse events.

CONCLUSION

In this paper, we report GraMDTA, a novel multimodal graph neural network based screening for discovery of drug-target associations. When compared using benchmark datasets, multi-modal aggregation outperformed other existing approaches. Our ablation studies further suggest that GraMDTA is robust in performance and demonstrate that supplementing pretrained knowledge graph representations to structural representations can improve the performance of the model. Our case study using minoxidil demonstrates that GraMDTA predictions could be useful for hypothesizing drug-phenotype associations (drug repositioning or drug-induced adverse events). Although pretraining knowledge

graph using GraphSAGE supplemented GraMDTA, our pretraining could be further improved. An important future direction we plan to pursue is utilizing graph neural networks which are heterogeneous content aware [46]. More specifically utilizing context-specific gene expression (from human patients or animal models) information, and aggregating feature vectors accordingly. Secondly, we plan to expand our work on geometrical structures of protein as another modality in lieu with geometrical deep learning literature [47]–[49]. Finally, to encourage reproducibility of the work, we provide the source code and benchmark datasets at https://github.com/yellajaswanth/GraMDTA.

REFERENCES

[1] T. M. Nguyen, T. Nguyen, and T. Tran, "Learning to Discover Medicines." arXiv, Feb. 14, 2022. Accessed: May 31, 2022. [Online]. Available: http://arxiv.org/abs/2202.07096

[2] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: Deep drug-target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, Sep. 2018, doi: 10.1093/bioinformatics/bty593.

[3] I. Lee, J. Keum, and H. Nam, "DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences," *PLOS Computational Biology*, vol. 15, no. 6, p. e1007129, Jun. 2019, doi: 10.1371/journal.pcbi.1007129.

[4] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, and M. M. Hoffman, "Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities," *Information Fusion*, vol. 50, pp. 71–91, Oct. 2019, doi: 10.1016/j.inffus.2018.09.012.

[5] K. Huang, T. Fu, L. M. Glass, M. Zitnik, C. Xiao, and J. Sun, "DeepPurpose: a deep learning library for drug–target interaction prediction," *Bioinformatics*, vol. 36, no. 22–23, pp. 5545–5547, Dec. 2020, doi: 10.1093/bioinformatics/btaa1005.

[6] J. K. Yella and A. G. Jegga, "Magic bullets: Drug repositioning and drug combinations," in *Reference Module in Biomedical Sciences*, Elsevier, 2021. doi: https://doi.org/10.1016/B978-0-12-820472-6.00116-X.

[7] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh, "GraphDTA: predicting drug-target binding affinity with graph neural networks," *Bioinformatics*, vol. 37, no. 8, pp. 1140–1147, May 2021, doi: 10.1093/bioinformatics/btaa921.

[8] M. Tsubaki, K. Tomii, and J. Sese, "Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences," *Bioinformatics*, vol. 35, no. 2, pp. 309–318, Jan. 2019, doi: 10.1093/bioinformatics/bty535.

[9] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How Powerful are Graph Neural Networks?," *7th International Conference on Learning Representations, ICLR 2019*, Oct. 2018.

[10] D. K. Duvenaud *et al.*, "Convolutional Networks on Graphs for Learning Molecular Fingerprints," in *Advances in Neural Information Processing Systems*, 2015, vol. 28. Accessed: Mar. 01, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2015/hash/f9be311e65d81a9ad8150a60844bb94c-Abstract.html

[11] J. K. Yella and A. Jegga, "MGATRx: Discovering Drug Repositioning Candidates Using Multi-view Graph Attention," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1–1, 2021, doi: 10.1109/TCBB.2021.3082466.

[12] Y. Wang, G. Deng, N. Zeng, X. Song, and Y. Zhuang, "Drug-Disease Association Prediction Based on Neighborhood Information Aggregation in Neural Networks," *IEEE Access*, vol. 7, pp. 50581–50587, 2019.

[13] F. Wan, L. Hong, A. Xiao, T. Jiang, and J. Zeng, "NeoDTI: Neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions," *Bioinformatics*, vol. 35, no. 1, pp. 104–111, 2019, doi: 10.1093/bioinformatics/bty543.

[14] T. Ma, C. Xiao, J. Zhou, and F. Wang, "Drug similarity integration through attentive multi-view graph auto-encoders," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2018-July, pp. 3477–3483, 2018.

[15] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive Representation Learning on Large Graphs," in *Advances in Neural Information Processing Systems*, 2017, vol. 30. Accessed: Mar. 02, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7ebea9-Abstract.html

[16] D. Ramachandram and G. W. Taylor, "Deep Multimodal Learning: A Survey on Recent Advances and Trends," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 96–108, Nov. 2017, doi: 10.1109/MSP.2017.2738401.

[17] Q. Zhao, H. Zhao, K. Zheng, and J. Wang, "HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism," *Bioinformatics*, vol. 38, no. 3, pp. 655–662, Feb. 2022, doi: 10.1093/bioinformatics/btab715.

[18] H. Öztürk, E. Ozkirimli, and A. Özgür, "WideDTA: prediction of drug-target binding affinity." arXiv, Feb. 04, 2019. doi: 10.48550/arXiv.1902.04166.

[19] T. K. Sheils *et al.*, "TCRD and Pharos 2021: mining the human proteome for disease biology," *Nucleic Acids Research*, vol. 49, no. D1, pp. D1334–D1346, Jan. 2021, doi: 10.1093/nar/gkaa993.

[20] T. Sheils *et al.*, "How to Illuminate the Druggable Genome Using Pharos," *Current Protocols in Bioinformatics*, vol. 69, no. 1, p. e92, 2020, doi: 10.1002/cpbi.92.

[21] D. S. Wishart *et al.*, "DrugBank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic acids research*, vol. 34, no. suppl1, pp. D668–D672, 2006.

[22] A. S. Brown and C. J. Patel, "A standard database for drug repositioning," *Scientific Data*, 2017, doi: 10.1038/sdata.2017.29.

[23] J. Piñero *et al.*, "The DisGeNET knowledge platform for disease genomics: 2019 update," *Nucleic Acids Research*, 2019, doi: 10.1093/nar/gkz1021.

[24] O. Ursu *et al.*, "DrugCentral: online drug compendium.," *Nucleic acids research*, vol. 45, no. D1, pp. D932–D939, Jan. 2017, doi: 10.1093/nar/gkw993.

[25] D. Szklarczyk *et al.*, "The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets," *Nucleic Acids Res*, vol. 49, no. D1, pp. D605–D612, Jan. 2021, doi: 10.1093/nar/gkaa1074.

[26] J. Tang *et al.*, "Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis," *J Chem Inf Model*, vol. 54, no. 3, pp. 735–743, Mar. 2014, doi: 10.1021/ci400709d.

[27] M. I. Davis *et al.*, "Comprehensive analysis of kinase inhibitor selectivity," *Nat Biotechnol*, vol. 29, no. 11, Art. no. 11, Nov. 2011, doi: 10.1038/nbt.1990.

[28] "RDKit: Open-Source Cheminformatics Software." https://www.rdkit.org/ (accessed Aug. 21, 2022).

[29] Z. Xiong *et al.*, "Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism," *J. Med. Chem.*, vol. 63, no. 16, pp. 8749–8760, Aug. 2020, doi: 10.1021/acs.jmedchem.9b00959.

[30] P. Veličković, A. Casanova, P. Liò, G. Cucurull, A. Romero, and Y. Bengio, "Graph attention networks," *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018.

[31] A. Vaswani *et al.*, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[32] "[1912.01703] PyTorch: An Imperative Style, High-Performance Deep Learning Library." https://arxiv.org/abs/1912.01703 (accessed Aug. 21, 2022).

[33] "[1412.6980] Adam: A Method for Stochastic Optimization." https://arxiv.org/abs/1412.6980 (accessed Aug. 21, 2022).

[34] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, pp. 1–14, Sep. 2017.

[35] J. Chen, E. E. Bardes, B. J. Aronow, and A. G. Jegga, "ToppGene Suite for gene list enrichment analysis and candidate gene prioritization," *Nucleic Acids Research*, vol. 37, no. SUPPL. 2, 2009, doi: 10.1093/nar/gkp427.

[36] D. Pruthi, M. Gupta, B. Dhingra, G. Neubig, and Z. C. Lipton, "Learning to Deceive with Attention-Based Explanations." arXiv, Apr. 06, 2020. Accessed: Aug. 21, 2022. [Online]. Available: http://arxiv.org/abs/1909.07913

[37] S. Serrano and N. A. Smith, "Is Attention Interpretable?" arXiv, Jun. 09, 2019. Accessed: Aug. 21, 2022. [Online]. Available: http://arxiv.org/abs/1906.03731

[38] "Drugs@FDA: FDA-Approved Drugs." https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=018154 (accessed Aug. 21, 2022).

[39] "Drugs@FDA: FDA-Approved Drugs." https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=071344 (accessed Aug. 21, 2022).

[40] S. Qiu, S. P. Fraser, W. Pires, and M. B. A. Djamgoz, "Anti-invasive effects of minoxidil on human breast cancer cells: combination with ranolazine," *Clin Exp Metastasis*, vol. 39, no. 4, pp. 679–689, Aug. 2022, doi: 10.1007/s10585-022-10166-7.

[41] R. H. Knutsen *et al.*, "Minoxidil improves vascular compliance, restores cerebral blood flow, and alters extracellular matrix gene expression in a model of chronic vascular stiffness," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 315, no. 1, pp. H18–H32, Jul. 2018, doi: 10.1152/ajpheart.00683.2017.

[42] P. B. Gorelick *et al.*, "Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the american heart association/american stroke association," *Stroke*, vol. 42, no. 9, pp. 2672–2713, Sep. 2011, doi: 10.1161/STR.0b013e3182299496.

[43] A. Kearney-Schwartz *et al.*, "Vascular structure and function is correlated to cognitive performance and white matter hyperintensities in older hypertensive patients with subjective memory complaints," *Stroke*, vol. 40, no. 4, pp. 1229–1236, Apr. 2009, doi: 10.1161/STROKEAHA.108.532853.

[44] C. Ghelardini, N. Galeotti, and A. Bartolini, "Influence of potassium channel modulators on cognitive processes in mice," *Br J Pharmacol*, vol. 123, no. 6, pp. 1079–1084, Mar. 1998, doi: 10.1038/sj.bjp.0701709.

[45] A. Gupta, A. Sharma, A. Kumar, and R. Goyal, "Alteration in memory cognition due to activation of caveolin-1 and oxidative damage in a model of dementia of Alzheimer's type," *Indian J Pharmacol*, vol. 51, no. 3, pp. 173–180, 2019, doi: 10.4103/ijp.IJP_81_17.

[46] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, "Heterogeneous Graph Neural Network," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage AK USA, Jul. 2019, pp. 793–803. doi: 10.1145/3292500.3330961.

[47] B. Jing, S. Eismann, P. Suriana, R. J. L. Townshend, and R. Dror, "Learning from Protein Structure with Geometric Vector Perceptrons." arXiv, May 15, 2021. doi: 10.48550/arXiv.2009.01411.

[48] B. Dai and C. Bailey-Kellogg, "Protein interaction interface region prediction by geometric deep learning," *Bioinformatics*, vol. 37, no. 17, pp. 2580–2588, Sep. 2021, doi: 10.1093/bioinformatics/btab154.

[49] P. Gainza *et al.*, "Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning," *Nat Methods*, vol. 17, no. 2, Art. no. 2, Feb. 2020, doi: 10.1038/s41592-019-0666-6.