Enhancing Drug-Drug Interaction Predictions: Novel Dual-Contrasting Framework and Self-Attention in Message-Passing Neural Networks

Chenlin Chai

Abstract

Drug-drug interactions (DDIs) occur when multiple drugs react with each other when taken together. They can lead to unintended side effects that may be harmful to patients. Developing an efficient and accurate computational model for DDI predictions is highly important to assist healthcare professionals in making better prescription decisions.

The proposed GMPNN-CS++ model in this paper employs the Self-attention mechanism and a residual memory network after GMPNN-CS's message-passing module to enhance the extracted representation of cross-substructure pairs within two interacting drug molecules. Previous neural network models for DDI prediction mainly consider two types of samples: positive samples and negative non-interaction samples. I introduce a novel dual-contrasting sampling approach in GMPNN-CS++ to include a third type, negative mislabeled-interaction samples, representing an overlooked common negative scenario. A dual-contrasting loss function is designed to make the neural network distinguish between positive samples and both types of negative samples, thereby widening the model's applicability. Through dual-contrasting, the proposed method GMPNN-CS++ demonstrates an ability to capture additional features and improves performance in predicting both positive and negative DDI cases, achieving an overall accuracy of 97%, compared to the baseline GMPNN-CS.

Keywords: Drug-drug Interaction, Message-passing Neural Network, Dual-contrasting Sampling, Self-attention, Residual Memory Network

1 Introduction

This paper investigates the prediction of drug-drug interactions (DDIs) using message-passing neural networks. In the GMPNN-CS model, a gated message-passing neural network mechanism is proposed to learn substructures around every atom in drug molecules (Nyamabo et al., 2022). This paper utilizes this mechanism and extends it by incorporating the Self-attention mechanism (Vaswani et al., 2017) and a residual memory network (He et al., 2016) to enhance the extracted cross-substructure representations. A dual-contrasting framework is further proposed to enable the model to identify a new negative case, in which the given drug pair does interact, while the interaction type is mislabeled, thus widening its applicability.

1.1 Drug-drug Interactions

Drug-drug interactions are the reactions between two or more drugs when taken together by patients. They can lead to inhibition or induction of drug effects, which may be adverse side effects that can lead to unexpected morbidity; the inhibition of drug effects can also cause treatment inefficacy (Deng et al., 2021; Hunta et al., 2015). However, synergistic effects produced by interactions between multiple drugs, on the contrary, can be utilized in treatment to increase the efficiency of drugs consumed on purpose (Vijayan et al., 2022). These facts suggest the importance of studying drug-drug interactions to not only mitigate morbidity but also improve treatment efficacy. Because many patients have multiple illness conditions, polypharmacy has become more common in recent years and led to a higher incidence of DDI (Jiang et al., 2022). According to the Centers for Disease Control and Prevention (CDC), 21.3% of the U.S. population took three or more prescriptions in the past 30 days, while another study reveals a 175% DDI-prescription ratio with 4.33% of DDIs being serious and 66.12% significant (Ferooqui et al., 2018). Predicting drug-drug interactions and informing biomedical professionals of potential DDI effects before drugs are prescribed to patients by doctors is thus necessary.

1.2 Related Work

Due to the extensive number of existing drugs and possible drug combinations, computational methods have a huge advantage for the task of predicting drug-drug interactions because they can generate the results of prediction for a large number of samples quickly and at a low cost. Accordingly, many machine learning and deep learning methods have been proposed to predict these interactions and achieve success. Early machine learning methods are often based on the calculation of similarities between drugs to conduct DDI predictions (Qiu et al., 2022). For example, Ferdousi et al. (2017) constructed a computational model based on functional similarity of drugs. In 2018, Ryu et al. (2018) proposed DeepDDI, a new deep learning framework, which calculates the structural similarity profile of these drugs as an input for their feed-forward deep neural network.

In recent years, more models have been observed to use graph data to represent drug molecules in graph neural networks (GNN). A graph is composed of various nodes and edges, and it stores information (in feature vectors) for each node and edge, as well as the positional relationships, including which nodes or edges are adjacent to each other. Wang et al. (2020) developed the GoGNN model that treats all drugs as a single network by creating an interaction graph of these drugs. In this way, the drugs are the nodes of a graph, and the DDIs are the edges connecting drug pairs that would have interactions. Different

from GoGNN, Nyamabo et al.'s (2021) SSI-DDI model views drugs as independent entities instead of a single network. Drugs are converted to independent graphs, in which nodes are atoms and edges are the bonds of the molecules. A drug pair's two corresponding graphs will be passed through a graph attention network and a co-attention layer to compute the interaction score of this drug pair for prediction.

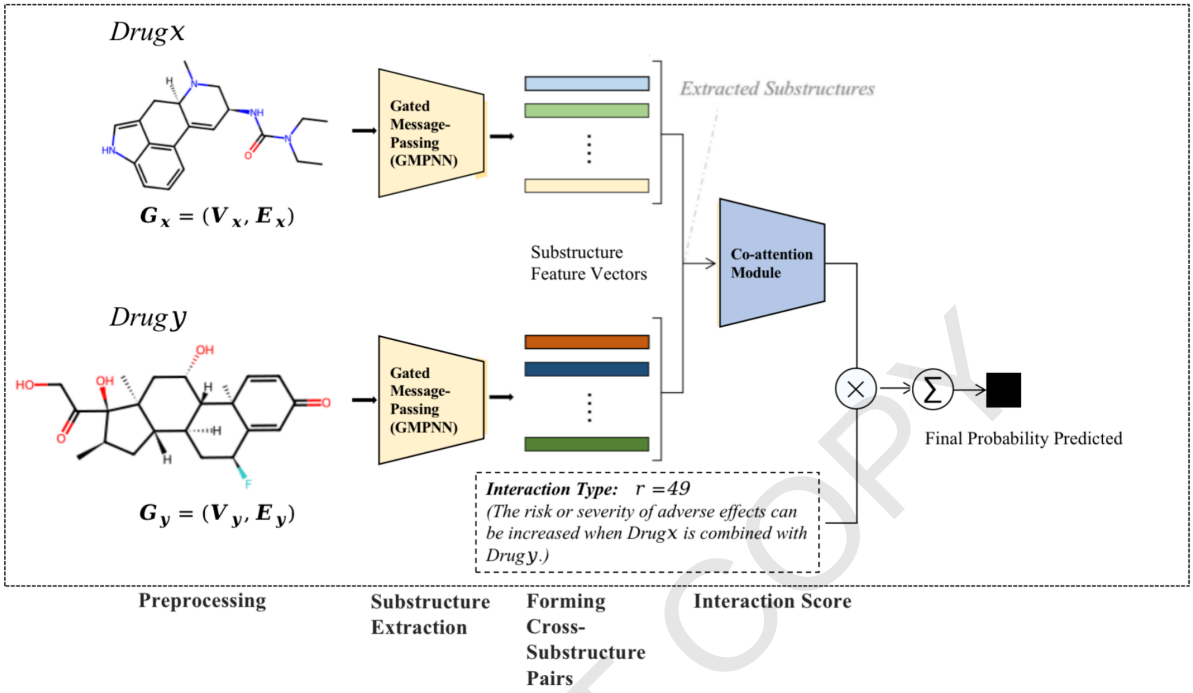

Figure 1: An Overview of GMPNN-CS (All molecular graphs are generated using RDKit)

In 2022, Nyamabo et al. further developed the GMPNN-CS model, which replaces the graph attention network in SSI-DDI with a novel gated message-passing neural network. This message-passing neural network they proposed provides new insight into directly learning the substructural information within drug molecules by summing up an atom's features with the features of all its adjacent atoms and bonds. In this way, each atom in a drug molecule is treated as the center of a substructure and contains the information of its surroundings. Message-passing neural networks enable substructure learning, which is crucial for predicting DDIs because molecules with similar substructures are more likely to have similar biological activities and physicochemical properties (Safizadeh et al., 2021).

1.3 Research Objectives

The GMPNN-CS model generates a substructure representation for every single atom in drug molecules. This, however, leads to a redundancy in calculations, as pointed out by themselves in the discussion section of their paper (Nyamabo et al., 2022). Often a drug molecule contains many similar substructures, and the extracted substructure feature vectors for all the atoms within these substructures can be nearly identical, leading to redundancies in calculations for a group of very similar vectors and affecting the performance of the model for predicting DDIs of those drugs that contain multiple similar substructures. Hence, the first objective of this research is to solve this redundancy issue, and the Self-attention mechanism and a residual memory network are specifically implemented after the co-attention layer to enhance the representations of cross-substructure pairs.

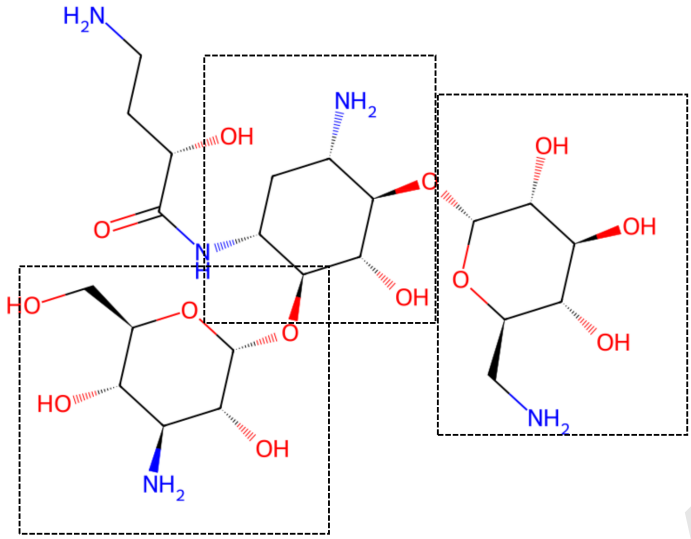

Figure 2: Amikacin (C22H43N5O13), an example of a DrugBank molecule that contains multiple atoms (boxed) surrounded by similar substructures. Figure generated with RDKit.

Another potential deficiency of the GMPNN-CS model is that it only contains two types of samples for training the model: positive samples and negative non-interaction samples. It ignores another common negative case, the interaction-mislabeled case. The missing of this negative sample limits the model's real-world applicability. A dual-contrasting framework is proposed in this paper, along with a dual-contrasting loss function for distinguishing the positive samples from both of the negative cases. A description of the original sampling approach and a detailed discussion of the proposed dual-contrasting approach is in Section 2.5.

2 Methods

2.1 Tasks

Given a drug pair (D_x, D_y) and a specific interaction type r, the GMPNN-CS++ model proposed in this paper aims to compute the probability that this certain type of interaction will occur between the two drugs: $f(D_x, D_y, r) \rightarrow [0, 1]$.

The task of our model is identical to the baseline model GMPNN-CS's (Nyamabo et al., 2022). It can be regarded as a binary classification task. When the predicted probability is larger than 0.5, the drug pair will be predicted to produce the labeled interaction; if it's lower than 0.5, the prediction will be that this labeled interaction type will not occur between the two drugs (Nyamabo et al., 2022).

2.2 Data Source

Data are obtained from the DrugBank dataset, which contains 191,808 DDI samples, 1704 drugs, and 86 interaction types. Table 1 provides three examples of DDI samples from the DrugBank dataset. There are 86 interaction types, with indexes from 1-86 to represent them. For instance, the Type 49 interaction, as shown in the table, means that "The risk or severity of adverse effects can be increased when #Drug1 is combined with #Drug2." The SMILE sequences, line notations to represent molecules' three-dimensional structures, are also obtained and will be used to get the features of drug molecules.

D_x ID	D_y ID	Interaction Type #1	Interaction Type Description	D_x SMILE sequence	D_y SMILE sequence
DB04571	DB00460	1	#Drug1 may increase the photosensitizing activities of #Drug2.	CC1=CC2=CC3=C(OC(=O)C=C3C)C(C)=C2O1	COC(=O)CCC1=C2NC(\C=C3/N=C(/C=C4\N\C(=C/C5=N/C(=C\2)/CCCC(O)=O)=C5C)C(C=C)=C4C)C2=CC=C([C@@H](C(=O)OC)[C@@]32C)C(=O)OC)=C1C
DB00193	DB13025	49	The risk or severity of adverse effects can be increased when #Drug1 is combined with #Drug2.	COC1=CC=CC(=C1)[C@@]1(O)CCCC[C@@H]1CN(C)C	CCN(CC)CCNC(=O)C1=CC(=CC=C1OC)S(C)(=O)=O
DB00437	DB00691	86	The risk of a hypersensitivity reaction to #Drug2 is increased when it is combined with #Drug1.	OC1=NC=NC2=C1C=NN2	CCOC(=O)[C@H](CCC1=CC=CC=C1)N[C@@H](C)C(=O)N1CC2=CC(OC)=C(OC)C=C2C[C@H]1C(O)=O

Figure 3: Examples of DrugBank DDI samples

2.3 Data Preprocessing

This research uses the same preprocessing strategy in GMPNN-CS (Nyamabo et al., 2022). The Python library RDKit (Landrum, 2023) enables obtaining the atom features and bond features of the DrugBank molecules represented in SMILES. Table 2 lists all the features used in this research.

Atom Features	Size	Bond Features	Size
Atom Symbol	45	Bond Type	4
Atom degree	10	Conjugated	1
Implicit valence	7	In Ring	1
Formal charge	1		
Number of radical electrons	1		
Hybridization	5		
Aromacity	1		

Figure 4: Extracted Molecule Features. Atom and bond features.

These features, in the form of numbers, are encoded using One-hot encoding to generate graph data with the set of node (atom) feature vectors V and the set of edge (bond) feature vectors E for every drug. Note that for features with a size of 1, this research simply uses 0 or 1 to represent the feature in vectors instead of using the One-hot method. Here is an example to illustrate the One-hot encoding method. From Table 2 we can see that Hybridization has a size of 5. This is because there are five types of hybridizations for an atom (sp, sp2, sp3, sp3d, or sp3d2). If an atom belongs to the sp2 hybridization, one-hot encoding assigns a value of 1 to the position corresponding to sp2 in the vector, and a value of 0 for the rest of the positions, and the number of positions equals the size of the feature. Therefore, in this case, an atom in sp2 hybridization's Hybridization feature will be represented as [0, 1, 0, 0, 0] in the atom feature vector, where the position with a value of 1 is the position corresponding to sp2.

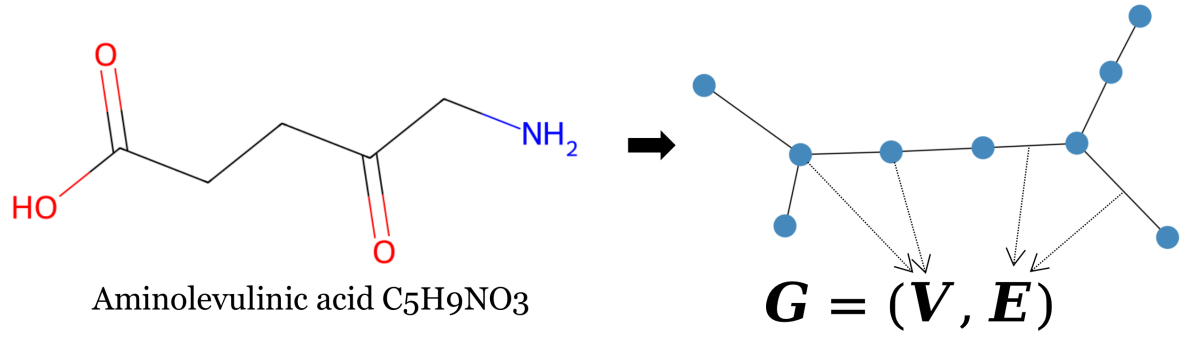

Figure 5: A drug molecule being preprocessed into a graph.

After obtaining the set of atom feature vectors **V** and the set of bond feature vectors **E** for a molecule, the molecule is converted to a graph. In a graph, nodes represent atoms, and edges represent bonds. Each node in the graph contains an atom feature vector of that corresponding atom, and each edge contains a bond feature vector of the corresponding bond. Figure 3 shows that graph data still preserves the positional relationships of the atoms and bonds in a molecule by storing the connectivity of atoms. This is vital for passing adjacent atoms' information to the center atom in order to extract its surrounding substructure.

2.4 GMPNN-CS++ Framework

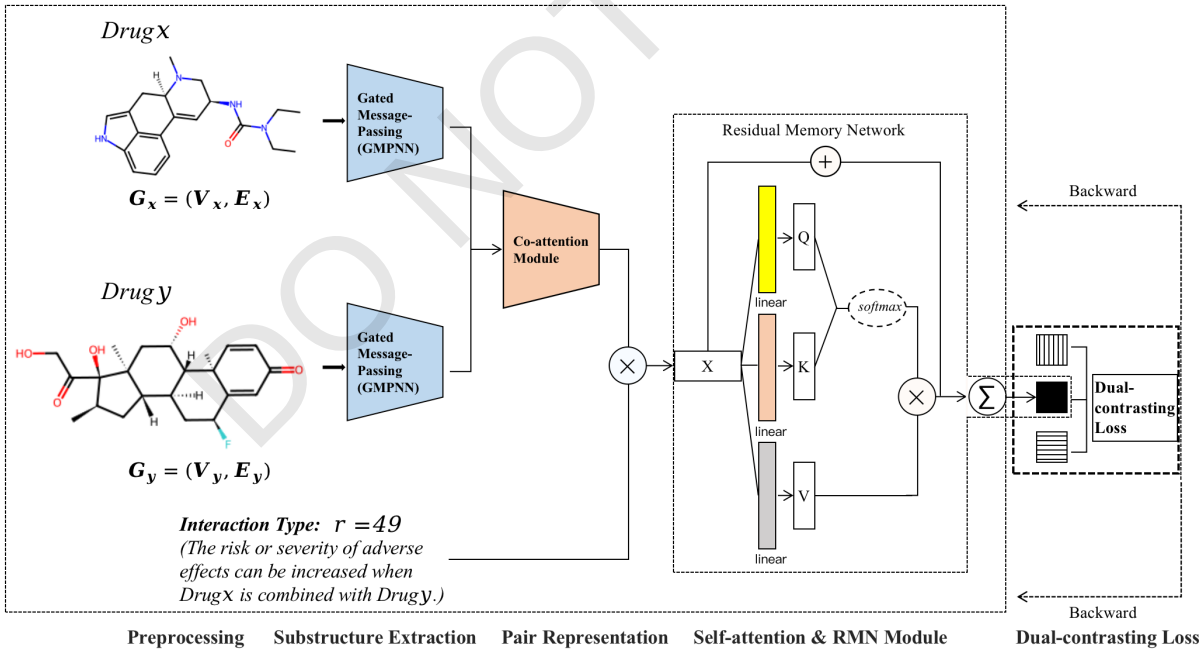

Figure 6: GMPNN-CS++ framework

As discussed in Section 2.1, the task of the proposed GMPNN-CS++ model is to predict the probability that a certain interaction would occur between a drug pair, so the input of the model has to be (G_x, G_y, r), a triplet that contains two drugs' graph data and a labeled interaction type. The GMPNN (gated message-passing neural network) and Co-attention modules proposed in GMPNN-CS are used to obtain cross-

substructure pair representations for the two drugs (Nyamabo et al., 2022). The Self-attention and residual memory network mechanism are implemented after the message-passing neural network and Co-attention module to enhance those pair representations. Most importantly, the GMPNN-CS++ model designs a dual-contrasting framework to make the model learn to distinguish a new type of negative case. This is achieved by the proposed sampling method and the dual-contrastive loss function.

2.5 GMPNN and Co-attention: Brief Explanation

This section provides a brief explanation of GMPNN-CS's gated message-passing neural network and Co-attention module (Nyamabo et al., 2022).

The gated message-passing neural network extracts the substructures around every atom in a molecule. Figure 5 illustrates how an atom's surrounding substructure is extracted using this approach. The information of atoms and bonds close to the central atom (painted black) will be multiplied by learnable weights and then added to the original atom feature vector of the central atom so that this substructure feature vector not only contains the features of the central atom but also contains information of nearby atoms and bonds. This extracts the substructure surrounding this atom, and it is done for every atom in a molecule.

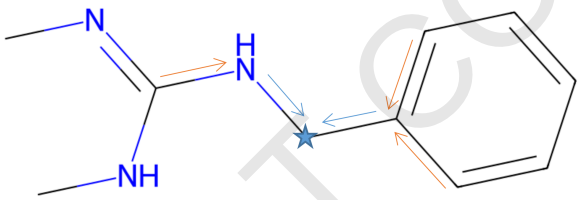

Figure 7: Gated Message-Passing Process.

Then, each substructure feature vector in one drug will form a cross-substructure pair with a substructure from the other drug in the drug pair. These cross-substructure pairs will be passed through the Co-attention module, which includes linear transformation layers and a softmax layer, to obtain the pair representations of them.

2.6 Self-attention Mechanism and Residual Memory Network

This paper proposes to use the Self-attention mechanism (Vaswani et al., 2017), along with a residual memory network structure (He et al., 2016), to fix the redundancy issue of substructure extraction in the message-passing neural network.

The Self-attention mechanism contains three linear transformation layers of X (Vaswani et al., 2017), which in our case are the pair representations of the cross-substructures. The linear transformation layer multiples weights to each element in the vectors, and this highlights the most important elements in those pair representations and diminishes the effects of redundancy.

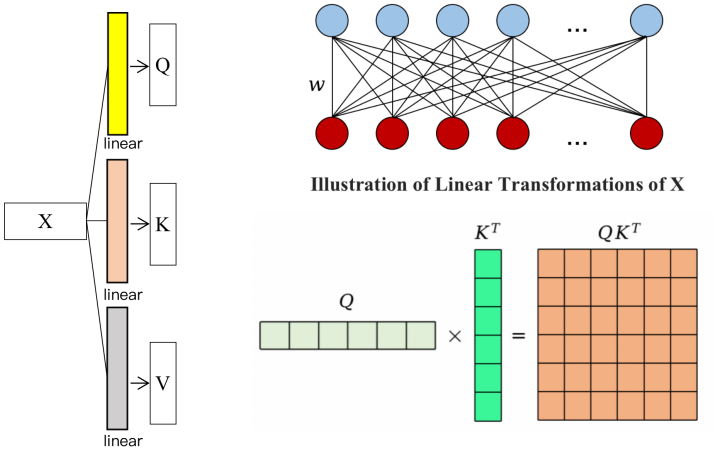

Figure 8: Self-attention Mechanism

The entire process of the Self-attention mechanism can be written as the equation below:

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_K}})V, \quad (1)$$

in which the softmax function transfers all the elements in the matrix QK^T into numbers within the range of [0,1]. Then, it is multiplied by V, the pair representation vector that received one linear transformation, for the ultimate transformation. $\sqrt{d_K}$ is the square root of K's dimension. QK^T is divided by $\sqrt{d_K}$ for normalization.

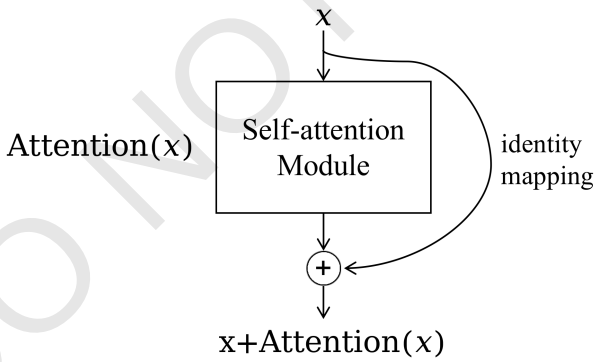

Figure 9: Residual Memory Network in Self-Attention

The residual memory network (RMN) structure (He et al., 2016) is further integrated with the Self-attention mechanism to achieve better pair representations. RMN reflects a technique of identical mapping by adding the input of a module in the neural network to the output of this module. It is proposed to deepen the structure of the neural network by adding two values obtained after they are passed through some number of layers together and has been demonstrated to lead to higher efficiency of neural networks while minimizing the percentage errors (Moniz et al., 2016). Drawing on this idea, this research finds that by adding the original pair representations to the transformed pair representations, which are the outputs of the Self-attention module, the pair representations would be further enhanced, and the results are shown in Section 3.

After these two modules, all pair representations for a drug pair will be added together and the sigmoid function will be used to convert it to a number between 0 and 1 for final probability prediction.

2.7 Dual-contrasting Framework

This paper has finished discussing every module inside the GMPNN-CS++ model except the loss function. Two things are left for discussion: What values should be fed into the loss function, and what should the loss function be?

After each epoch of training, the neural network model needs to generate a loss value, and minimizing this loss value is always the goal for training the neural networks. The neural networks will keep changing the values of the weights in every layer to achieve a lower loss value. Hence, a loss function determines the objectives of a neural network.

GMPNN-CS model's loss function takes in two values: the probability predicted for positive samples and the probability predicted for negative non-interaction samples, and its loss function serves to make the model learn the features and maximize the probability predicted for positive samples, while minimizing the probability predicted for negative non-interaction samples, thereby distinguishing the positive samples from non-interaction samples (Nyamabo et al., 2022).

This former sampling approach of GMPNN-CS uses positive samples and negative non-interaction samples in the model. Positive samples record true DDI incidents, while negative samples record incidents that would not occur in reality.

Assume $\mathcal{D}$ is the set of all drug pairs recorded in the Drugbank dataset that are known to have interaction, while $\mathcal{R}$ is the set of interaction types in the dataset. The positive samples are directly obtained from the DrugBank dataset. They are triplets that contain two drugs (D_x, D_y) and a correctly labeled interaction type r (86 interaction types total in the dataset) for this drug pair: (D_x, D_y, r), expressing a DDI incident. This means that a certain interaction type r does occur between (D_x, D_y), as recorded in the positive samples $\mathcal{P}$ (note here that in the DrugBank dataset, every drug pair only has one correct interaction type):

$$\forall (D_x, D_y) \in \mathcal{D}, \exists r \in \mathcal{R} \text{ such that } (D_x, D_y, r) \in \mathcal{P}.$$

The negative non-interaction samples are formed by replacing one drug (either D_x or D_y) with a third drug D_z randomly chosen from the dataset that does not belong to the drug pair. The DDI incident (D_x, D_z, r) recorded in the non-interaction case would not occur (which is called "negative") because (D_x, D_z) is not a drug pair from the dataset and thus will not interact, and thus will definitely not produce the type r interaction:

$$\forall (D_x, D_z) \notin \mathcal{D}, \text{ and } r \notin \mathcal{R}_{(D_x, D_z)}.$$

This sampling approach ignores another common negative case, the interaction-mislabeled case. The interaction-mislabeled case can be expressed as $(D_x, D_y, r*)$, and we can see that the drug pair is the same as that of the positive sample, meaning that the drug pair does have interaction in the real world. However, the interaction type is altered by replacing the correct interaction type with another interaction randomly chosen from the rest of the 85 interaction types in DrugBank. In other words, the interaction-mislabeled case means that a drug pair does interact, but the labeled interaction type in the sample is wrong:

$$\forall (D_x, D_y) \in \mathcal{D}, \text{ and } r^* \in \mathcal{R} \setminus \{r\}.$$

This case is very common in real-world settings because when a biomedical professional obtains a drug pair, he or she does not know the correct interaction type for this drug pair and often needs to try different types of interaction to see which one has the highest probability to occur between these two drugs. The missing of this negative case during learning makes the GMPNN-CS model less applicable to distinguish the positive samples from these interaction-mislabeled samples.

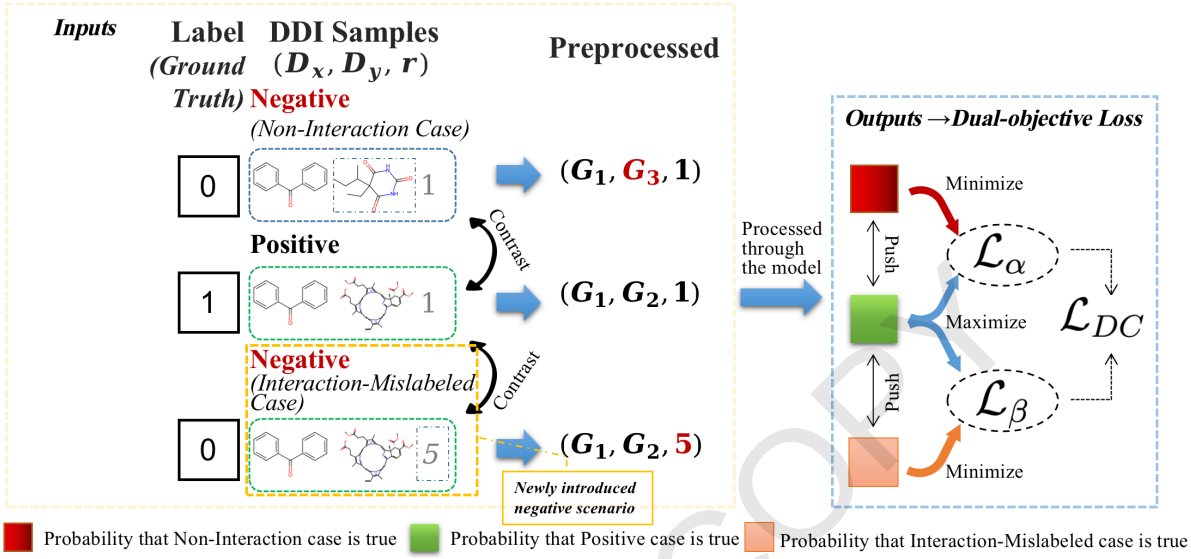

Figure 10: Ilustration of the Dual-contrasting framework. For each training, three samples consisting of positive, non-interaction, and interaction-mislabeled samples would be the inputs of the model separately. Three predicted probabilities, corresponding to these three samples, will thus be generated.

In contrast, the proposed dual-contrasting framework has two objectives: (1) to distinguish Positive samples from Non-interaction samples, and (2) to distinguish Positive samples from Interaction-mislabeled samples. Specifically, the dual-contrasting framework contains a newly added interaction-mislabeled negative case in the DDI samples and a dual-contrasting loss function that consists of two binary cross-entropy loss functions, designed for balanced recognition of both negative cases.

A dual-objective loss function will then be used to maximize the probability predicted for the positive sample and minimize the probabilities of the non-interaction and interaction-mislabeled negative samples by contrasting the two negative samples with the positive sample separately.

$$\mathcal{L}_\alpha = -\frac{1}{N}\sum_{i=1}^{N}\left(\log(p_i) + \log(1 - p_i')\right) \tag{2}$$

$$\mathcal{L}_\beta = -\frac{1}{N}\sum_{i=1}^{N}\left(\log(p_i) + \log(1 - p_i*)\right) \tag{3}$$

$$\mathcal{L}_{DC} = 0.5 \cdot \mathcal{L}_\alpha + 0.5 \cdot \mathcal{L}_\beta \tag{4}$$

Specifically, the loss function is formed by adding two equally weighted sub-loss functions. p_i is the probability predicted for the positive sample, p_i' is the probability predicted for the non-interaction sample, and p_i* is the probability predicted for the interaction-mislabeled sample. $\mathcal{L}_\alpha$ contrasts the non-interaction sample with the positive sample, which is the final loss function used in the baseline

model GMPNN-CS (Nyamabo et al., 2022). $\mathcal{L}_\beta$ contrasts the interaction-mislabeled sample with the positive sample. Adding them together with equal weights forms the final loss function $\mathcal{L}_{DC}$. The larger the probability predicted for the positive sample and the lower the probabilities predicted for the two negative samples, the lower the loss value. By using this loss function, every epoch the model will be trained to compute predictions closer to real-world results.

2.8 Setup

The number of hidden layer features in the neural network is set to 64. The learning rate is 1e-3, the weight-decay rate is 5e-4, the batch size is 512, the number of epochs is 100, and the number of iterations in the message-passing neural network is 10. The 191,808 DDI samples are divided into 6 folds. The training: validation: testing sets ratio is 6:2:2.

3 Results

The results of the methods proposed in this paper are evaluated and compared with the baseline GMPNN-CS model. As Table 1 on the next page shows, The dual-contrasting framework, the Self-attention mechanism, and the residual memory network (RMN) modules are evaluated separately and observed to have actual effects. Whenever a module is added, the model performs generally better than when it is not added. The final GMPNN-CS++ model uses all three modules and has the best overall result.

The proposed dual-objective loss function has been proven to be very effective in dual discrimination (Positive vs. Non-interaction & Positive vs. Interaction-mislabeled) tasks. The most important thing is that even if we only employ the dual-contrasting framework with the absence of the Self-attention and RMN modules, the recall score still increases by 0.67% compared with the baseline GMPNN-CS model. The recall score is calculated by dividing the number of positive samples classified correctly as positive by the total number of positive samples and is used to evaluate a model's performance in predicting positive samples. This improved recall score indicates that after introducing new negative samples and by contrasting the new negative case with the positive samples, the model has learned more information that contributes to higher accuracy in predicting positive samples.

Models	ACC	AUROC	F1	P	R	AP
GMPNN-CS (baseline)	95.09	98.86	90.82	85.28	97.14	96.40
(w/Dual-contrasting)	96.58	99.12	95.03	92.25	97.78	97.67
(w/Dual-contrasting + Self-attention)	96.86	99.16	95.40	93.12	97.80	**97.80**
(w/Dual-contrasting + Self-attention + RMN)	**96.94**	**99.18**	**95.52**	**93.33**	**97.81**	97.79

Table 1: Model evaluation on Accuracy, AUROC, F1, Precision, Recall and Average Precision scores. With the dual-contrasting framework, Self-attention, and RMN modules all implemented, the proposed GMPNN-CS++ obtains its best performance.

Figures 11-14 demonstrate the robust performance of the proposed GMPNN-CS++ model, which contains newly added Self-attention, RMN, and dual-contrasting methods.

The confusion matrix (Figure 16 on the next page) suggests a very balanced recognition of positive and negative samples.

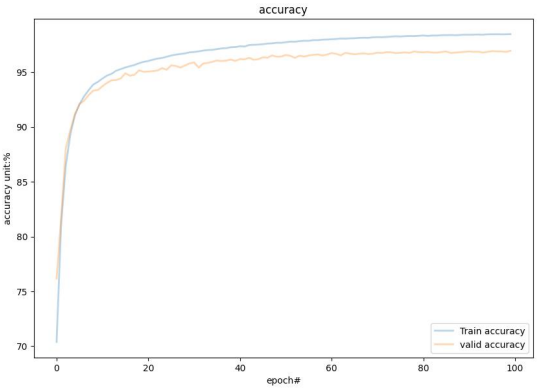

Figure 11: Accuracy

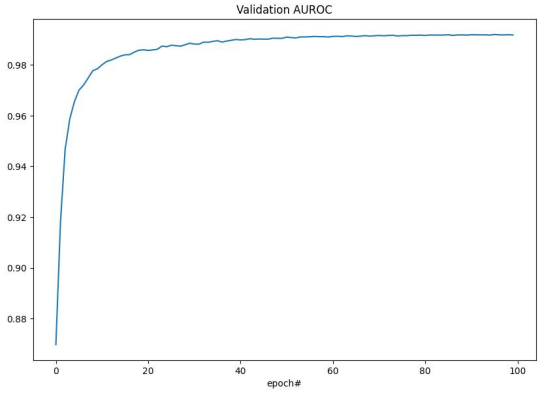

Figure 12: AUROC

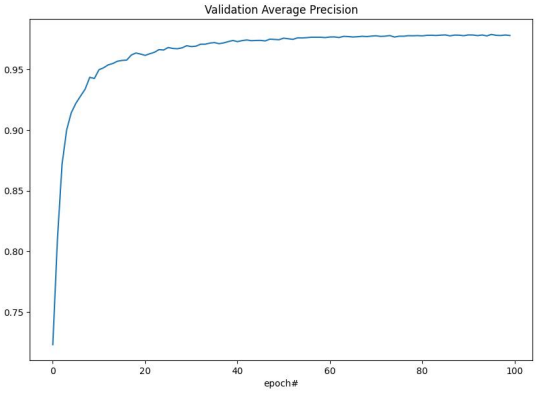

Figure 13: AP

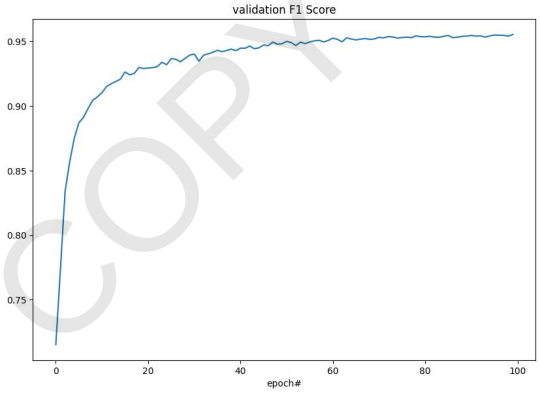

Figure 14: F1

Figure 15: Summary of Metrics

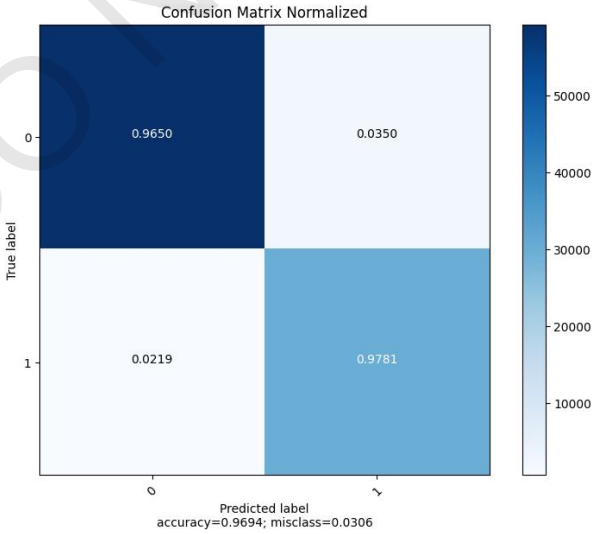

Figure 16: Confusion Matrix Normalized

4 Discussion

Based on the good performance of the dual-contrasting approach in drug-drug interaction prediction tasks, this dual-contrasting approach may also be useful in other networks and tasks to improve the

prediction accuracy of positive samples. In the future, we can try to employ this approach in other models and test its performance. This research also proves the effectiveness of Self-attention & RMN in enhancing the feature representations of drug molecule pairs in message-passing neural networks.

5 Conclusion

This research proposes a dual-contrasting framework based on message-passing neural networks for DDI predictions. A critical interaction-mislabeled case is added to the original positive and non-interaction cases and a dual-objective loss function is designed to achieve a balanced recognition of both negative cases. With further implementation of the Self-attention module and a residual memory network, this model GMPNN-CS++ is compared with the baseline GMPNN-CS model and demonstrates both better performance and wider application.

GMPNN-CS++ allows medical personnel to identify mislabeled interaction types of a drug pair and determine the correct interaction by changing different interaction types in the sample for testing. By bringing in more precise DDI prediction and a wider application, GMPNN-CS++ can inform health-care professionals of possible adverse DDIs before they occur, ensuring patient safety and improving personalized medicine.

6 Ackowledgement

I would like to thank Dr. Posey for teaching me some basic organic chemistry knowledge and for giving me valuable paper feedback. I also want to thank her for bringing together students who love science in our school's innovation lab as the head of our school's research program and for providing us with an excellent academic environment. I want to thank my parents and my sister for their encouragement throughout the project.

References

Centers for Disease Control and Prevention. (2023, September 19). FASTSTATS - therapeutic drug use. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/fastats/drug-use-therapeutic.htm

Deng, Y., Yu, S., Deng, L., Liu, H., Liu, X., & Luo, Y. (2021). A multi-task graph convolutional network modeling of drug-drug interactions and synergistic efficacy. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). https://doi.org/10.1109/bibm52615.2021.9669575

Ernstmeyer, K., & Christman, E. (Eds.). (1970, January 1). Chapter 1 pharmacokinetics & Pharmacodynamics. Open Resources for Nursing (Open RN). Nursing Pharmacology [Internet]. 2nd edition. https://www.ncbi.nlm.nih.gov/books/NBK595006/

Farooqui, R., Hoor, T., Karim, N., & Muneer, M. (2018). Potential drug-drug interactions among patient's prescriptions collected from medicine out-patient setting. Pakistan Journal of Medical Sciences, 34(1). https://doi.org/10.12669/pjms.341.13986

Ferdousi, R., Safdari, R., & Omidi, Y. (2017). Computational prediction of drug-drug interactions based on drugs functional similarities. Journal of Biomedical Informatics, 70, 54–64. https://doi.org/10.1016/j.jbi.2017.04.021

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2016.90

Hunta, S., Aunsri, N., & Yooyativong, T. (2015). Drug-drug interactions prediction from enzyme action crossing through machine learning approaches. 2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). https://doi.org/10.1109/ecticon.2015.7207126

Jiang, H., Lin, Y., Ren, W., Fang, Z., Liu, Y., Tan, X., Lv, X., & Zhang, N. (2022). Adverse drug reactions and correlations with drug–drug interactions: A retrospective study of reports from 2011 to 2020. Frontiers in Pharmacology, 13. https://doi.org/10.3389/fphar.2022.923939

Landrum, Greg. (2023). Open-source cheminformatics software. RDKit. https://www.rdkit.org/

Moniz, J., & Pal, C. (2016). Convolutional residual memory networks. arXiv preprint arXiv:1606.05262.

Nagata, K., Tsuji, T., Suetsugu, K., Muraoka, K., Watanabe, H., Kanaya, A., Egashira, N., & Ieiri, I. (2021). Detection of overdose and underdose prescriptions-An unsupervised machine learning approach. PloS one, 16(11), e0260315. https://doi.org/10.1371/journal.pone.0260315

Nyamabo, A. K., Yu, H., & Shi, J.-Y. (2021). SSI–DDI: Substructure–substructure interactions for drug–drug interaction prediction. Briefings in Bioinformatics, 22(6). https://doi.org/10.1093/bib/bbab133

Nyamabo, A. K., Yu, H., Liu, Z., & Shi, J.-Y. (2021). Drug–drug interaction prediction with learnable size-adaptive molecular substructures. Briefings in Bioinformatics, 23(1). https://doi.org/10.1093/bib/bbab441

Pichini, S., Di Trana, A., García-Algar, O., & Busardò, F. P. (2023). Editorial: Drug-drug interactions in pharmacology. Frontiers in Pharmacology, 14. https://doi.org/10.3389/fphar.2023.1155738

Qiu, Y., Zhang, Y., Deng, Y., Liu, S., & Zhang, W. (2022). A comprehensive review of computational methods for Drug-Drug Interaction Detection. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 19(4), 1968–1985. https://doi.org/10.1109/tcbb.2021.3081268

Ryu, J. Y., Kim, H. U., & Lee, S. Y. (2018). Deep learning improves prediction of drug–drug and drug–food interactions. Proceedings of the National Academy of Sciences, 115(18). https://doi.org/10.1073/pnas.1803294115

Safizadeh, H., Simpkins, S. W., Nelson, J., Li, S. C., Piotrowski, J. S., Yoshimura, M., Yashiroda, Y., Hirano, H., Osada, H., Yoshida, M., Boone, C., & Myers, C. L. (2021). Improving measures of chemical structural similarity using machine learning on chemical-genetic interactions. Journal of Chemical Information and Modeling, 61(9), 4156-4172. https://doi.org/10.1021/acs.jcim.0c00993

Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. Neural Information Processing Systems.

Vijayan, A., & B.S, C. (2022). Advance single stage convolutional neural network for Drug-Drug Interactions. 2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP). https://doi.org/10.1109/ccip57447.2022.10058636

Wang, H., Lian, D., Zhang, Y., Qin, L., & Lin, X. (2020). GoGNN: Graph of graphs neural network for predicting structured entity interactions. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. https://doi.org/10.24963/ijcai.2020/183

Drug-Drug Interaction. TDC. https://tdcommons.ai/multi_pred_tasks/ddi/

Sklearn.preprocessing.onehotencoder.scikit. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html