Supplementary Material of "MKG-FENN: A Multimodal Knowledge Graph Fused End-to-end Neural Network for Accurate Drug–Drug Interaction Prediction"

Outline. This document supplements the manuscript in the following aspects. Section 1 will present the symbols utilized in the paper. Section 2 presents the algorithm of our proposed MKG-FENN approach (to promoto readability and re-producibility). Section 3 will present the formulas of the evaluation metrics used in the paper. Section 4 will present some details of the baseline model. Section 5 is a complement to hyperparameter sensitivity experiments. Section 6 will present implementation details.

1 Symbols and Notations

For clarification, we summarize the complete symbols and notations employed in Table 1.

2 Algorithm Design

By analyzing the model, we designed the MKG-FENN algorithm. In below, we'll describe the flow of the algorithm 1.

To get started, we will build the DDI matrix y and 4 drug knowledge graphs $\mathcal{G}^1$,$\mathcal{G}^2$,$\mathcal{G}^3$ and $\mathcal{G}^4$. Then we initialize the drug knowledge graph. In steps 5-6, we randomly sample a fixed-size sample $\{\mathcal{N}_l\}_{l=1}^{L}$ from the drug knowledge graph. In steps 7-11, we used GNN to calculate the higher-order structure and semantic relationship of drugs. In step 14-17, we will connect four drug representations obtained according to the four drug knowledge graphss in series to obtain the final drug representation. Then, we combine the representations of the pair of drugs that are expected to interact in the DDI matrix. Finally, these combined representations serve as the inputs for predicting our DDIs.

Algorithm 1: MKG-FENN Algorithm

input : DDI matrix y,4 drug knowledge graphs $\mathcal{G}^1$,$\mathcal{G}^2$,$\mathcal{G}^3$ and $\mathcal{G}^4$.
output: $\Gamma(d_i, d_j|y, \mathcal{G}^1, \mathcal{G}^2, \mathcal{G}^3, \mathcal{G}^4)$

1 *Initialization* $\mathcal{G}^1, \mathcal{G}^2, \mathcal{G}^3, \mathcal{G}^4$;
2 **while** *MKG-FENN not converge* **do**
3 **for** (d_i, d_j) *in* y **do**
4 **for** *j,* $\mathcal{G}$ *in enumerate (*$[\mathcal{G}^1, \mathcal{G}^2, \mathcal{G}^3, \mathcal{G}^4$ *])* **do**
5 $\{\mathcal{N}_l\}_{l=1}^{L} \leftarrow$ *Neighborhood Sampling(entity e)*;
6 $e^0 \leftarrow e, \forall e \in \mathcal{N}_0$;
7 **for** $l = 1, ..., L$ **do**
8 **for** $e \in \mathcal{N}_l$ **do**
9 $e_{\mathcal{N}_l}^{(l)} \leftarrow \sum_{t_n \in \mathcal{N}_l(e)} \pi_{(e,r_{in})}^{(l)} e_{t_n}^{(l-1)}$;
10 **end**
11 **end**
12 $E_{d_i}^{j+1} \leftarrow e_{d_i}^{(l)}, E_{d_j}^{j+1} \leftarrow e_{d_j}^{(l)}$;
13 **end**
14 $\hat{E}_{d_i} \leftarrow {E^1}_{d_i} \oplus E_{d_i}^2 \oplus {E^3}_{d_i} \oplus E_{d_i}^4$;
15 $\hat{E}_{d_j} \leftarrow {E^1}_{d_j} \oplus E_{d_j}^2 \oplus {E^3}_{d_j} \oplus E_{d_j}^4$;
16 $\hat{E}_{d_{i,j}} \leftarrow \hat{E}_{d_j} \oplus \hat{E}_{d_i}$;
17 *Calculate* $y_{ij} = f\left(\hat{E}_{d_{i,j}}\right)$;
18 *Update parameters* Θ;
19 **end**
20 **end**

3 Evaluation Metric

Regarding the evaluation metrics for model assessment, we utilize multi-class classification evaluation metrics, which encompass accuracy (ACC), area under the precision-recall curve (AUPR), area under the ROC curve (AUC), F1 score, precision, and recall (Lyu et al. 2021). The formulas for these metrics are as follows:

$$\text{ACC} = \frac{\sum_{i=1}^{n=65} TP_{\text{i}}}{\sum_{i=1}^{n=65} TP_{\text{i}} + \sum_{i=1}^{n=65} FN_{\text{i}}} \tag{1}$$

$$\text{Precision} = \left(\sum_{i=1}^{n=65} \frac{TP_i}{TP_i + FP_i} \right) / n \tag{2}$$

Symbol	Explanation
$\mathcal{Y}$	The DDI matrix.
N_d	The number of drugs in the matrix.
d_i, d_j	Drug i and drug j.
y_{ij}	An interaction event between drug i and drug j.
$\mathcal{L}$	The set of interaction event .
y_{N_l}	A event type, $N_l \in [1, 2, 3...65]$.
$\mathcal{D}, \mathcal{R}, \mathcal{T}$	Respectively representing a series of drug entities, relationships, and tail entities.
d, r_{dt}, t	Respectively representing drug entity, relationship, and tail entity.
$\mathcal{G}^i$	Drug knowledge graph, $i \in [1, 2, 3, 4]$
$\Gamma(\cdot), \Theta$	The objective function and its model parameters, respectively.
$E^i{}_{\mathcal{G}^i}$	The i-th knowledge graph's initial representation matrix
N_d, N_r, N_k	The number of drugs, relationships, and tail entities, respectively.
${e_d}^{(0)}, {e_r}^{(0)}, {e_t}^{(0)}$	Initial embeddings for drugs, relationships, and tail entities.
R^d	The embedding dimension of the drug knowledge graph.
$N_s\left(d_i\right)$	The fixed-size neighborhoods of drug i.
r_{in}, t_n	t_n represents the neighborhood of drug d_i and r_{in} represents the semantic relationship.
$e_{r_{in}}^{(l-1)}, e_{d_i}^{(l-1)}, e_{t_n}^{(l-1)}$	The embedding of the relationship, drug i and the neighborhood of drug i in the $(l-1)^{th}$ layer, respectively.
$W_1^{(p)}, b_1^{(p)}$	The trainable weight matrix and the bias vector in the p layer, respectively.
$\pi_{(d_i, r_{in})}^{(l)}$	The semantic feature score of drug d_i and the relationship in the $(l)^{th}$ layer.
$W_1^{(p)}, b_1^{(p)}$	The trainable weight matrix and the bias vector in the p layer, respectively.
$\odot, \oplus$	Element-wise multiplication and the concatenation operation, respectively.
σ	The activation function.
$e_{\mathcal{N}_s(d_i)}^{(l)}$	A vector of neighborhood representation
W_2, b_2	The trainable weight matrix and a bias vector, respectively.
E_{d_i}	The embedding of drug i.
${E^1}_{d_i}, {E^2}_{d_i}, {E^3}_{d_i}, {E^4}_{d_i}$	Drug embeddings corresponding to the four parts of the dataset.
$\hat{E}_{d_i}, \hat{E}_{d_j}$	The final representations of the drug i and drug j.
$W_3^{(q)}, b_3^{(q)}$	The trainable weight matrix and the bias vector, q is the number of fully connected layers.
$\hat{y}_{ij}$	The value of prediction.

Table 1: Symbols and Notations

$$\mathrm{F1} = 2 * \frac{\mathrm{P}ecision * \mathrm{R}ecall}{\mathrm{P}ecision + \mathrm{R}ecall} \tag{3}$$

TP_i represents that the actual disease is the i-th disease, and the prediction is also the i-th disease.FN_{i} represents that the actual disease is the i-th disease, but it is predicted to be other diseases.FP_i means that the real disease is different from the i-th disease, but it is predicted to be the i-th disease. TN_i indicates that the real disease is a different disease, and it is correctly predicted as such.

$$\mathrm{TPR} = \mathrm{R}ecall = \left(\sum_{i=1}^{n=65} \frac{TP_i}{TP_i + FN_i} \right) / n \tag{4}$$

$$\mathrm{FPR} = \left(\sum_{i=1}^{n=65} \frac{FP_i}{FP_i + TN_i} \right) / n \tag{5}$$

When FPR (False Positive Rate) and TPR (True Positive Rate) are used as the x-axis and y-axis, AUC represents the area under the FPR-TPR curve. On the other hand, when $\mathrm{R}ecall$ and $Precision$ are used as the x-axis and y-axis, AUPR is the area under the $Precision$-$\mathrm{R}ecall$ curve.

4 Baseline Model

Due to the length of the text, we will introduce some details of the baseline model here. We will introduce six state-of-the-art models: MDDI-SCL, MDF-SA-DDI, DDIMDL, MDNN, Lee et al.'s methods, and DeepDDI and several traditional classification methods: DNN, RF, KNN, and LR. Details of the comparison model are shown in Table 2.

5 Additional Experiments

We identified three crucial parameters: the size of the sampling neighborhood $\mathcal{N}_s$, the dimension of the drug embedding d in the drug knowledge graph, and the clas- sification loss weight (CLW). Hyper-parameter sensitivity experiments are presented in Figure 1.

Effect of neighborhood size. We investigated the impact of neighborhood size on model performance by altering the size of the sampled neighborhood. Figure 1 demonstrated that the model achieved optimal performance when $\mathcal{N}_s$ = 6. When the neighborhood size was too small, the model struggled to effectively structure the information. Conversely, when Ns is too large, the model becomes susceptible to being misled by noise.

Effect of embedding dimension. The performance of the model can be influenced by varying the embedding dimensions, and we examined the impact of changing the value of d on model performance. Selecting an appropriate value for d allows the model to capture an adequate amount of drug and entity information, leading to improve performance.

Effect of classification loss weight. Fine-tuning the CLW can have a significant impact on the model's performance. After conducting several experiments, we have discovered that the optimal CLW value for achieving the best model performance is $1e-8$.

Model	Description
DeepDDI (Jae Yong, Kim, and Lee 2018)	It is the representative matrix factorization model that factorizes data of the user-item matrix for the recommender systems, *Proc. Natl. Acad. Sci. U.S.A. 2018.*
Lee et al.'s methods (Lee, Park, and Ahn 2019)	It is a new deep learning-based model to improve classification accuracy, *BMC Bioinform. 2019.*
DDIMDL (Yifan et al. 2020)	It is an method that combines multiple drug profiles with deep learning, *Bioinformatics 2020.*
MDNN (Lyu et al. 2021)	It is a method for solving DDI events based on multimodal deep neural networks, *IJCAI 2021.*
MDF-SA-DDI (Shenggeng et al. 2022b)	It is a method based on multi-source drug fusion, multi-source feature and transformer self-attention mechanism, *BRIEF BIOINFORM 2021.*
MDDI-SCL (Shenggeng et al. 2022a)	It is a method based on supervised contrastive learning, *J. Cheminformatics 2022.*
DNN (Yifan et al. 2020)	It is a traditional classification method deep neural network.
RF (Yifan et al. 2020)	It is a traditional classification method random forest.
KNN (Yifan et al. 2020)	It is a traditional classification method k-nearest neighbour.
LR (Yifan et al. 2020)	It is a traditional classification method logistic regression.
MKG-FENN	Our proposed model is a end-to-end model based multi-channel.

Table 2: Descriptions of all the contrasting models.

6 Implementation details

Learning rate, batch size, and regularization rate are fine-tuned for all the models to achieve their own optimal prediction accuracy. Drug embedding size and the sampling size were searched using an arithmetic progression to attain the optimal performance of the model. We reproduced the experimental results of the comparative models through two approaches: 1. Replicating them using their open-source code with parameter settings as recommended in the original papers, and 2. Referring to the experimental outcomes reported in the original papers. All experiments were conducted on a host machine running the Windows 11 operating system that has an Intel Core i5-10500 CPU operating at 3.10 GHz with 8.00 GB RAM.

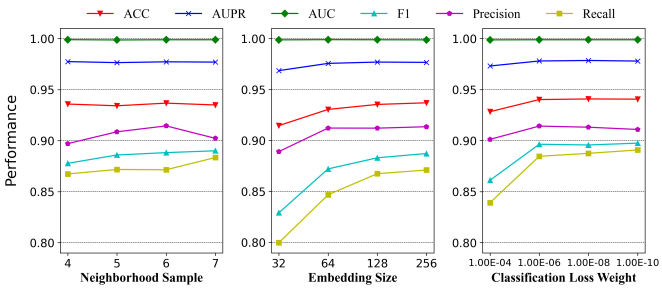

Figure 1: Results of MKGNN with varying values of neighborhood sample NS, embedding size ES in DKG and classification loss weight CLW.

References

Jae Yong, R.; Kim, H. U.; and Lee, S. Y. 2018. Deep learning improves prediction of drug-drug and drug-food interactions. *Proceedings of the National Academy of Sciences*, 115: E4304–E4311. Doi: 10.1073/pnas.1803294115.

Lee, G.; Park, C.; and Ahn, J. 2019. Novel deep learning model for more accurate prediction of drug-drug interaction effects. *BMC Bioinformatics*, 20: 415. Doi: 10.1186/s12859-019-3013-0.

Lyu, T.; Gao, J.; Tian, L.; Li, Z.; Zhang, P.; and Zhang, J. 2021. MDNN: A Multimodal Deep Neural Network for Predicting Drug-Drug Interaction Events. *In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence,*, 3536–3542.

Shenggeng, L.; Weizhi, C.; Gengwang, C.; Songchi, Z.; Dong-Qing, W.; and Yi, X. 2022a. Mddi-scl: predicting multi-type drug-drug interactions via supervised contrastive learning. *Journal of Cheminformatics,*, 14: 1–12.

Shenggeng, L.; Yanjing, W.; Lingfeng, Z.; Yanyi, C.; Yatong, L.; Yitian, F.; Mingming, J.; Qiankun, W.; Bowen, Z.; Yi, X.; and Dong-Qing, W. 2022b. MDF-SA-DDI: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Briefings in Bioinformatics,*, bbab421.

Yifan, D.; Xinran, X.; Yang, Q.; Jingbo, X.; Wen, Z.; and Shichao, L. 2020. A multimodal deep learning framework for predicting drug-drug interaction events. *Bioinformatics,*, 36: 4316–4322.