Supplementary Material of "DHENN: A Deeper Hybrid End-to-end Neural Network for Highly Accurate Drug-Drug Interaction Events Prediction"

ACM Reference Format:
. 2024. Supplementary Material of "DHENN: A Deeper Hybrid End-to-end Neural Network for Highly Accurate Drug-Drug Interaction Events Prediction". In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Outline. This document serves as a supplementary material to the manuscript, providing additional insights and details. Section 1 outlines the algorithm underlying our proposed DHENN approach, aimed at enhancing readability and reproducibility. Section 2 presents the mathematical formulas of the evaluation metrics employed in the paper. Section 3 delves into the specifics of the baseline model. Section 4 complements the hyperparameter sensitivity experiments conducted in the main manuscript. Finally, Section 5 details the implementation procedures and considerations.

1 ALGORITHM DESIGN AND TIME COMPLEXITY ANALYSIS

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

First, we construct the DDI matrix y and the multi-modal knowledge graph $\mathcal{G}$. Then we initialize the multi-modal knowledge graph $\mathcal{G}$.In steps 4-5, we randomly sample a fixed-size sample $\{\mathcal{N}_l\}_{l=1}^{L}$ from the drug knowledge graph. In steps 6-12, We employed GNN to compute the higher-order structure and semantic relationships among drugs, and concatenated the representations of drug pairs. In step 14-19, We employ a cascaded deep structure to predict drug representations in order to enhance predictive performance.

From the overall algorithmic perspective, DHENN is divided into two parts: GNN extracts drug representations, and the cascaded deep structure predicts DDI events. The time complexity of the GNN part is $O(N \times D \times \mathcal{N}_l \times L)$, where N represents the number of DDIs, D represents the dimension of drug encodings, $\mathcal{N}_l$ represents the neighborhood size, and L represents the number of layers in the GNN. In the cascaded deep structure part, the corresponding time complexity is $O(N \times m \times (D + C))$, where m represents the number of cascaded layers and C represents the number of DDI prediction categories. Therefore, the overall time complexity of the final model is $O(N \times ((D \times \mathcal{N}_l \times L) + (m \times (D + C))))$.

Algorithm 1: DHENN Algorithm

input : DDI matrix y,multi-modal knowledge graph $\mathcal{G}$.
output: $\Gamma(d_i, d_j|y, \mathcal{G})$

1 *Initialization* $\mathcal{G}$;
2 **while** *not converge* **do**
3 **for** (d_i, d_j) *in* y **do**
4 $\{\mathcal{N}_l\}_{l=1}^{L} \leftarrow$ *Neighborhood Sampling(entity e)*;
5 $e^0 \leftarrow e, \forall e \in N_0$;
6 **for** $l = 1, ..., L$ **do**
7 **for** $e \in \mathcal{N}_l$ **do**
8 $e_{\mathcal{N}_l}^{(l)} \leftarrow \sum_{t_n \in \mathcal{N}_l(e)} \pi_{(e,r_{in})}^{(l)} e_{t_n}^{(l-1)}$;
9 **end**
10 **end**
11 $E_{d_i}^{j+1} \leftarrow e_{d_i}^{(l)}, E_{d_j}^{j+1} \leftarrow e_{d_j}^{(l)}$;
12 $\hat{E}_{d_{i,j}} \leftarrow \hat{E}_{d_j} \oplus \hat{E}_{d_i}$;
13 $y_{ij} \leftarrow 0$;
14 **for** $l = 1, ..., m$ **do**
15 $\hat{E}_{d_{i,j}} \leftarrow \hat{E}_{d_{i,j}} \oplus y_{ij}$;
16 *Calculate* $y_{ij} = f\left(\hat{E}_{d_{i,j}}\right)$;
17 *Update parameters* Θ;
18 **end**
19 **end**
20 **end**

2 EVALUATION METRIC

Regarding the evaluation metrics for model assessment, we utilize a diverse array of multi-class classification evaluation metrics to ensure a comprehensive understanding of the model's performance. These metrics, which include accuracy (ACC), area under the precision-recall curve (AUPR), area under the receiver operating characteristic curve (AUC), F1 score, precision, and recall [1].The formulas for these metrics are as follows:

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

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

$$\mathrm{F1} = 2 * \frac{\mathrm{Pr}\,ecision * \mathrm{Re}call}{\mathrm{Pr}\,ecision + \mathrm{Re}call} \tag{3}$$

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Model	Description
MDDI-SCL [3]	It is a method that leverages supervised contrastive learning as its foundation, *J. Cheminformatics 2022.*
MDF-SA-DDI [4]	It is grounded in multi-source drug fusion, incorporating multi-source features and the transformer self-attention mechanism, *BRIEF BIOINFORM 2021.*
DDIMDL [1]	It combines multiple drug profiles using deep learning techniques, *Bioinformatics 2020.*
Lee et al.'s methods [2]	It is a novel deep learning model aimed at enhancing classification accuracy., *BMC Bioinform. 2019.*
DeepDDI [5]	It is a representative matrix factorization model that decomposes the user-item matrix data for use in recommender systems, *Proc. Natl. Acad. Sci. U.S.A. 2018.*
DNN [1]	It is a traditional classification method deep neural network.
RF [1]	It is a traditional classification method random forest.
KNN [1]	It is a traditional classification method k-nearest neighbour.
LR [1]	It is a traditional classification method logistic regression.
DHENN	Our model is a multimodal, deep learning-based predictive system with a cascade structure for accurate predictions.

Table 1: Descriptions of all the contrasting models.

TP_i denotes the situation where both the actual disease and the predicted disease are the i-th type. Conversely, FN_{i} signifies a scenario where the actual disease is the i-th type, but the prediction erroneously indicates a different disease. On the other hand, FP_i occurs when the actual disease differs from the i-th type, yet the prediction incorrectly identifies it as the i-th disease. Lastly, TN_i represents a correct prediction where the actual disease is not the i-th type, and the prediction accurately reflects this. It is worth noting that n represents the types of events that will occur.

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

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

When plotting the False Positive Rate (FPR) on the x-axis and the True Positive Rate (TPR) on the y-axis, the AUC (Area Under the Curve) represents the total area enclosed by the FPR-TPR curve. Conversely, when using Re$call$ as the x-axis and $Precision$ as the y-axis, the AUPR (Area Under the $Precision$-Re$call$ Curve) denotes the enclosed area beneath the $Precision$-Re$call$ curve.

3 BASELINE MODEL

Owing to the extensive nature of the text, we shall focus on presenting an overview of the baseline model in this context. Specifically, we will introduce six cutting-edge models: MDDI-SCL, MDF-SA-DDI, DDIMDL, Lee et al.'s methods, and DeepDDI. Additionally, we will also consider several traditional classification methods, namely DNN, RF, KNN, and LR, for comparison. A comprehensive breakdown of the comparison models is detailed in Table 1.

4 ADDITIONAL EXPERIMENTS

In this study, we have identified four pivotal parameters: the dimensionality of drug embeddings within the drug knowledge graph (d), the extent of the sampling neighborhood ($\mathcal{N}S$), the weighting factor for the regularization controlling (RCW), and the coefficient regulating the cascaded loss function (CCL).Hyper-parameter sensitivity exper- iments are presented in Figure 1

Effect of embedding dimension. The performance of the model can be affected by changing the embedding dimensions, and we investigated the influence of varying the value of d on model performance. Choosing an appropriate value for d enables the model to capture a sufficient amount of drug and entity information, resulting in improved performance. From Figure 1, we can see that Dataset 1 utilized an embedding dimension of $d = 128$, and Dataset 2 also employed $d = 128$.

Effect of neighborhood size. We examined how the performance of the model is affected by varying the size of the sampled neighborhood. Figure 1 demonstrates the optimal values of the neighborhood sample ($\mathcal{N}S$) for the three datasets. In Dataset 1, the optimal $\mathcal{N}S$ value is 10. Similarly, in Dataset 2, the optimal $\mathcal{N}S$ value is also 10. When the neighborhood size was too small, the model faced difficulties in effectively organizing the information. On the other hand, when $\mathcal{N}S$ was too large, the model became more susceptible to being influenced by noise.

Effect of regularization controlling weight. The impact RCW on the model's performance is substantial. After conducting several experiments, we have determined that fine-tuning the RCW can significantly improve the model's performance. Figure 1 reveals that Dataset 1 achieved the best model performance with an optimal RCW value of 1e-8, and Dataset 2 had the optimal RCW value of 1e-10.

Effect of coefficient of cascaded loss. Figure 1 discusses the impact of CCL on the model. By studying the performance of the model with varying CCL values across two datasets, it was observed that the model achieved the best results when the CCL was in an increasing state.

REFERENCES

[1] Yifan Deng, Xinran Xu, Yang Qiu, Jingbo Xia, Wen Zhang, and Shichao Liu. 2020. A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics* 36, 15 (2020), 4316–4322.
[2] Geonhee Lee, Chihyun Park, and Jaegyoon Ahn. 2019. Novel deep learning model for more accurate prediction of drug-drug interaction effects. *BMC bioinformatics* 20 (2019), 1–8.
[3] Shenggeng Lin, Weizhi Chen, Gengwang Chen, Songchi Zhou, Dong-Qing Wei, and Yi Xiong. 2022. MDDI-SCL: predicting multi-type drug-drug interactions via

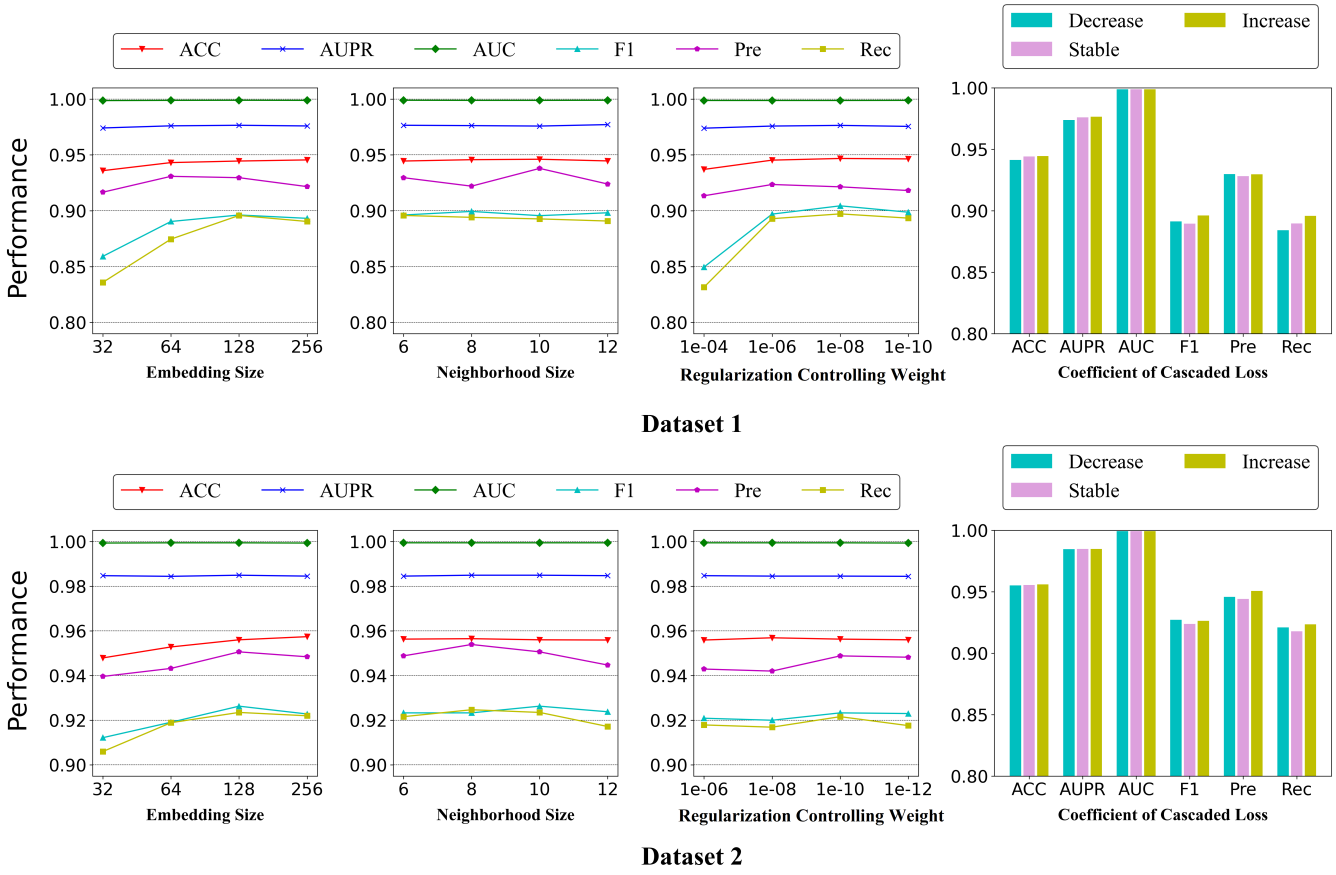

Figure 1: Sensitivity analysis of parameters in two datasets for exploring the impact of parameter variations on results.

supervised contrastive learning. *Journal of Cheminformatics* 14, 1 (2022), 1–12.

[4] Shenggeng Lin, Yanjing Wang, Lingfeng Zhang, Yanyi Chu, Yatong Liu, Yitian Fang, Mingming Jiang, Qiankun Wang, Bowen Zhao, Yi Xiong, et al. 2022. 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* 23, 1 (2022), bbab421.

[5] Jae Yong Ryu, Hyun Uk Kim, and Sang Yup Lee. 2018. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the national academy of sciences* 115, 18 (2018), E4304–E4311.