Causal Intervention for Measuring Confidence in Drug-Target Interaction Prediction

Wenting Ye[a], Chen Li[b], Yang Xie[a], Wen Zhang[a,d,e], Hong-Yu Zhang[a,d,e], Bowen Wang[a],
Debo Cheng[c*], Zaiwen Feng[a,d,e,f,g*]

[a] College of Informatics, Huazhong Agric. University, Wuhan, China
[b] Graduate School of Informatics, Nagoya University, Chikusa, Nagoya, 464-8602, Japan
[c] UniSA STEM, University of South Australia, Adelaide, 5095, Australia
[d] Hubei Key Laboratory of Agricultural Bioinformatics, Huazhong Agricultural University, Wuhan 430070, China
[e] Key Laboratory of Smart Farming for Agricultural Animals, Huazhong Agricultural University, Wuhan 430070, China
[f] Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China
[g] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

Abstract—Identifying and discovering drug-target interactions (DTIs) are vital steps in drug discovery and development. They play a crucial role in assisting scientists in finding new drugs and accelerating the drug development process. Recently, knowledge graph and knowledge graph embedding (KGE) models have made rapid advancements and demonstrated impressive performance in drug discovery. However, such models lack authenticity and accuracy in drug target identification, leading to an increased misjudgment rate and reduced drug development efficiency. To address these issues, we focus on the problem of drug-target interactions, with knowledge mapping as the core technology. Specifically, a causal intervention-based confidence measure is employed to assess the triplet score to improve the accuracy of the drug-target interaction prediction model. Experimental results demonstrate that the developed confidence measurement method based on causal intervention can significantly enhance the accuracy of DTI link prediction, particularly for high-precision models. The predicted results are more valuable in guiding the design and development of subsequent drug development experiments, thereby significantly improving the efficiency of drug development.

Index Terms—Drug-Target Prediction, Knowledge Graph, Causal Intervention, Confidence Measurement, Probability Calibration

I. INTRODUCTION

A drug target is a biomacromolecule or biomolecular structure in the body, usually including a protein or nucleic acid, intrinsically associated with a specific disease process. Drugs function by binding to these specific targets and modifying the gene function of the target to achieve the desired therapeutic effect [1], [2]. Computational prediction of DTI is critical to advancing our understanding of drug mechanisms of action, disease pathology, and side effects [3]. The traditional drug development process takes 12-15 years, and costs more than $1 billion. However, with the advancement of computer technology and bioinformatics, the identification of drug targets has increasingly turned to information technology. Sequence

and structure alignments [4], [5], homology modelling [6], and deep learning-based approaches [7], which have been widely used in drug-target research.

Recently, many models have emerged in the field of drug-target interactions with significant breakthroughs and effects. However, most previous studies have problems with existing confidence measurement methods, which are usually based on the original confidence score or score ranking of the model output, and the stability cannot be guaranteed. Additionally, the output scores of such models are often uninterpretable. To solve this problem, we apply the confidence measure method of causal intervention to drug-target link prediction. The method first applies the idea of causal intervention to change the embedding representation. Then, the KGE model was used to complete a new triplet composed of drug intervention entities to score again. Finally, a new confidence score was obtained by consistency calculation. Unlike the previous approach, which only focused on the optimal score, this approach measures the robustness of the embedded representation results by actively intervening in the input of the entity vector. Our motivation is to more truly reflect the confidence probability of the model's predicted results and improve the performance of the KGE model in drug-target interaction tasks.

The implementation of domain intervention consistency [8] proved that the intervention approach can improve the model confidence score, but it has not been tried in the field of drug development. This study aims to solve the problem that the model confidence is not accurate enough in the drug-target prediction problem dominated by the knowledge graph embedding model, so as to improve the accuracy and stability of drug-target prediction. The main contributions are summarized as follows.

- The embedding representation effects of different KGE models are investigated in two datasets applicable to the drug-target domain class.
- The model performance is evaluated by comparing the prediction results before and after intervening in the neighborhood of the embedded vectors.

- Probabilistic calibration is used to comprehensively evaluate the methods, and then to verify whether the confidence measure based on causal intervention can improve the confidence score of drug-target link prediction.

II. Related Work

Drug-target prediction is a very important field, and advanced techniques such as machine learning, deep learning and knowledge graphs have also made many achievements in this field. [9] used multi-core learning and clustering methods for drug-target measurement modeling, which demonstrated its certain advantages in predictive performance. Mohamed et al 2019 proposed the Complex [10] model to transform the drug-target prediction problem into a link prediction problem on the knowledge graph. Experiments show that the Complex model achieves the best results compared with the traditional KGE model. Bonner et al 2022 [11], [12]discussed the application and performance of different knowledge graph embedding models in drug discovery and compared their performance in drug target prediction, drug side effect prediction, and drug interaction prediction.

Although previous studies have made progress in using KGE technology for drug-target prediction, there are still some shortcomings and challenges. For example, TransH [13], RotatE, DisMult and other models have achieved good performance in drug-target prediction, but lack interpretation of predicted results. [14], [15]also has the problem of poor model interpretation. Researchers in related fields have made some efforts. [16] Integrated information from the chemical and genomic space, [17] explored the chemical properties, biological pathways, phenotypic effects, and network interactions between drugs of binding drugs, [18] learned and integrated characteristic representations of drugs and diseases at different convolution layers.

In order to solve the problem of the unexplainability of model output scores, we propose a new method to use causal intervention confidence measurement in drug-target link prediction. Instead of focusing only on the optimal score, this method measures the robustness of the embedded representation results by actively interfering with the input of the entity vector. Our motivation is to improve the performance of the KGE model in drug-target interaction tasks, making it a more reliable and interpretable predictive tool.

III. The Proposed Causal Intervention Method

In this section, we first overview our proposed causal intervention. Next, we describe the embedded intervention [19]. The third part will introduce the domain score sequence acquisition, and the fourth part will discuss the causal intervention confidence calculation. Finally, we will discuss the drug-target link prediction calibration.

A. Overview

The core idea of the proposed causal intervention method is to intervene in the dimension value of a drug entity embedding vector, and then the robustness of the score is calculated as the criterion to determine whether the original triplet is real enough by comparing the triplet score sequence composed of the original drug entity and the triplet score sequence composed of the entity after the intervention.

We visualize the overview of our proposed causal intervention method in Fig. 1. It describes the main steps and processes of the proposed causal intervention method for confidence measurement on drug-target interaction prediction. Firstly, target entities such as drugs and proteins in the data, along with their biological relationships, are embedded using the KGE model for learning latent representation. Subsequently, the learned representation is employed to predict the links between drugs and targets and their relationships in the dataset. For each triplet, the model generates a score, forming a score sequence. Next, the intervention entity is embedded to assess the confidence of the causal intervention. Finally, the results are calibrated using two common probabilistic calibration methods. The causal intervention method is then compared with the confidence method in traditional link prediction tasks to verify the effectiveness of the proposed causal intervention method in drug-target link prediction. In the following sections, we provide a description of the detailed steps of the proposed causal intervention method for the confidence measurement method.

B. Embedded Intervention

In this paper, our main focus is on studying the impact of causal intervention on the representation characteristics of embedded vectors. By intervening in the embedded vectors, we can manipulate the representation eigenvalues to observe how such changes affect the output results of the model [20]. For a drug-target entity represented by a D-dimensional embedding vector with d features, we explore different intervention strategies to interfere with the eigenvalues, resulting in d different representation vectors. The only difference between these d vectors and the initial drug-target entity's embedding representation lies in the variation of each dimension value. However, this change may cause some loss of the original semantic information. To assess the accuracy and robustness of the semantic representation, we compare the influence of the vector representation after causal intervention with the original vector representation on the link prediction results. The core idea of the proposed causal intervention method is to determine whether the original entity embedding representation possesses sufficient anti-interference and representativeness, analyzed from the perspective of stability measurement.

The current design of the intervention function has several limitations. One such limitation is its tendency to intervene more on low-dimensional vectors, as high-dimensional data vectors are typically denser, making intervention and interpretation more challenging. In practical applications, dimension reduction is often performed on embedded vectors to enhance their feature representation interpretability. Furthermore, careful consideration should be given to the selection of the number of intervention dimensions. If the number is too large, it may lead to the vector transitioning from one semantic
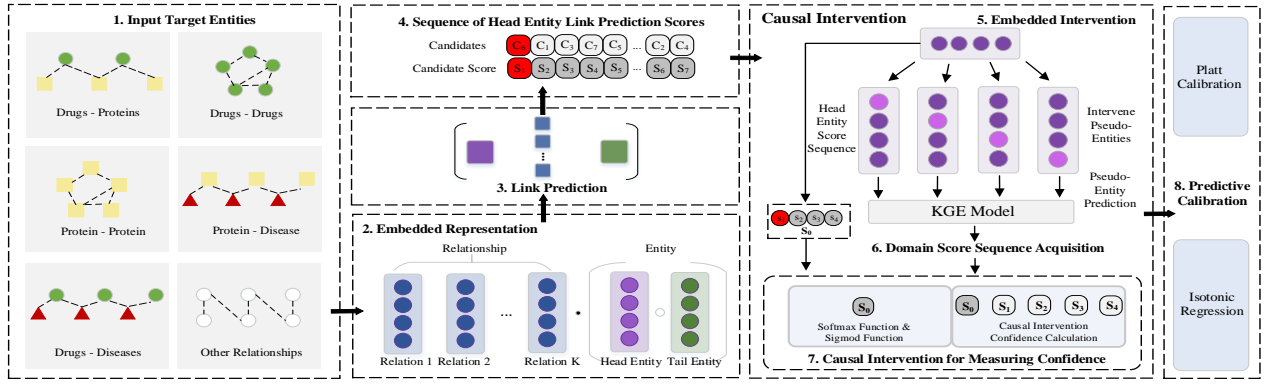
Fig. 1. The schematic diagram of the proposed causal intervention method for measuring confidence in drug-target causal intervention prediction.

category to another, thus adversely affecting the performance of the model and hindering the achievement of the desired intervention outcomes.

C. Domain Score Sequence Acquisition

Drug-target link prediction tasks can be divided into three categories: head entity prediction, relationship prediction, and tail entity prediction. In this paper, the tail entity is adopted as the prediction method for the link prediction task. Specifically, each triplet $< h, r, t >$ is modified by replacing the tail entity t, and the trained KGE model is used to compute a new score for each prediction triplet. The scores obtained for each replacement tail entity form a set of original score sequences, denoted as S. For the KGE model, the optimal score does not necessarily imply that the triplet is the most realistic triplet. Therefore, in confidence scoring, unlike the traditional focus on the optimal score, this work evaluates the confidence score by predicting the robustness of the embedded entity.

As shown in Fig. 4, the drug entity undergoes a causal intervention operation, leading to the generation of a group of domain intervention entities by altering the embedded representation values in each dimension. During the link prediction process, the KGE model scores each triplet formed by the domain pseudo-entities and the corresponding relationships between the original entity and each target prediction entity. Suppose a 4-dimensional embedded entity e, a relation r, and four candidate target tail entities $T = \{t_1, t_2, t_3, t_4\}$ are predicted by the $e - r$ query. After the intervention, four domain pseudo-entities $E' = \{e'_1, e'_2, e'_3, e'_4\}$ are generated. The model then scores the prediction triplet formed by each pseudo-entity e', relation r, and four tail entities T, generating the score sequence $S' = \{S'_1, S'_2, S'_3, S'_4\}$ corresponding to each pseudo-entity e'. As a result, there are d score sequences, and the total number of scores is $d \times m$, represented as S_i. Next, the scoring sequence S'_i corresponding to each pseudo-entity e'_i and the scoring sequence S corresponding to the initial entity e are used to calculate the consistency difference, ultimately yielding a new consistency score.

D. Causal Intervention for Measuring Confidence

The confidence score, also known as the consistency score, measured by the causal intervention method, indicates the stability of an entity's embedding. As depicted in Fig. 5, the primary calculation involves assessing the consistency of each field's score sequence in the pseudo-entity score sequence with the original entity score sequence, followed by normalization of the consistency values. The underlying assumption is that when an entity is true in the triplet of link prediction, its embedded characteristics align with the model's score prediction. Additionally, if a dimension feature in the entity embedding vector is altered, its effect on the entity is relatively small. Consequently, comparing the consistency of the two scoring sequences becomes crucial.

For the domain sequence, the index ordering of the first K candidate entities is denoted as $\rho(S^{N_i})$, and the index robustness scores of the first K candidate entities in the initial score sequence are calculated as follows:

$$\text{Consis}(S, S^{N_i}) = \sum_{j=0}^{J} \text{Softmax}(S)_j \cdot (1 - \text{Sgn}(\left| \rho(S)_j - \rho(S^{N_i})_j \right|)), \quad (1)$$

where J is a hyperparameter acting as a threshold to select the first J indexes that fit. A robustness score value of 1 indicates that the index of the domain score sequence matches that of the initial sequence. A lower robustness score implies a more mismatched index ranking based on the first J positions, indicating a larger degree of difference and worse robustness. After the intervention of a feature, the robustness score is obtained as $\text{Consis}(S, S^{N_i})$. Then, for the entity embedding vector with dimension d, the normalized confidence score can be obtained as follows:

$$\text{P}_{ci}(S) = \frac{1}{d} \sum_{i=0}^{d} \text{Consis}(S, S^{N_i}). \quad (2)$$

E. Drug-target Link Prediction Calibration

In the link prediction task, model calibration is of great importance. Because the machine learning model may produce uncertain prediction results, leading to overestimated

or underestimated probability values. This reduction in the reliability of predictions increases the cost of research and development and raises the risk of misdiagnosis. Therefore, proper model calibration is crucial to improve the accuracy and dependability of the predicted outcomes.

The specific operation step of calibration involves normalizing the confidence score of a prediction result to the interval [0, 1], and dividing this interval into n bin. For each bin, the difference between the average confidence score and the average true score for all of the pretests is calculated. A weighted average of all bin scores is then computed to obtain an ECE (Expected Calibration Error) score [21].

In a word, the purpose of model calibration is to increase the reliability of model prediction results, allowing it to better reflect the actual prediction outcomes of samples. A smaller ECE score indicates that the confidence score of the model is closer to the real score, implying higher reliability. In drug-target link prediction, the introduction of the causal intervention method aims to more accurately evaluate the confidence score of the model and reduce the error of probabilistic calibration. This holds significant importance for drug research and development, as it helps enhance the accuracy and trustworthiness of the model's predictions.

$$ \text{ECE} = \sum_{b=1}^{B} \frac{n_b}{N} \left| acc(b) - conf(b) \right|, \tag{3} $$

where b represents the number of samples in this bin, $acc(b)$ represents the true and accurate average value of samples in this bin, and $conf(b)$ represents the average predicted value of samples in this bin.

IV. EXPERIMENTS

A. Experimental Setup

Two public knowledge graphs applicable to the field of drug-target discovery, Hetionet and BioKG, were used in this experiment. Additionally, six KGE models were trained according to the original setting. The dimension of the high-dimensional model is set to 200. For the low-dimensional model, the embedded dimension chosen in this paper is 50. For the hyperparameters of causal intervention, we use the entity of the former TopK with parameters 3, 5, 10, 100, 200, 300, and for the first J relative positions in the consistency calculation, the parameters are 3, 5, 10, 100. All models are implemented based on the Pykeen framework [22], and all algorithms are implemented using the PyTorch framework based on Python. All experiments were run on NVIDIA RTX A400 GPU and Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz on Centos.

B. Evaluation Measures

In order to fully reflect the effect of the link prediction task, we introduce three evaluation indicators:

- **ECE**: Expected Calibration Error is commonly utilized to evaluate the calibration effect, the lower the ECE, the better the calibration effect of the model.

TABLE I
ECE OF DIFFERENT CONFIDENCE MEASURES ON DRUG-TARGET
INTERACTION PREDICTION MODELS

model	Hetionet			BioKG		
	ECE ↓			ECE ↓		
	SIG	TOP	CI	SIG	TOP	CI
TransE	.058	.032	**.016**	.150	.106	**.063**
Complex	.094	.042	**.035**	.154	.088	**.027**
DisMult	.051	.037	**.024**	**.026**	.062	.041
RotatE	.035	.022	**.020**	.083	.093	**.063**
TransH	.081	.027	**.013**	.034	.035	**.027**
CrossE	.067	.037	**.021**	.018	.031	**.016**

* The values in bold indicate optimal (minimum) values.

TABLE II
T10ACC OF DIFFERENT CONFIDENCE MEASUREMENT METHODS FOR
DRUG-TARGET INTERACTIONS

model	Hetionet				BioKG			
	T10ACC ↑			ACC ↑	T10ACC ↑			ACC ↑
	SIG	TOP	CI	H10	SIG	TOP	CI	H10
TransE	.146	**.237**	.121	.117	.453	.398	**.692**	.141
Complex	.022	.027	**.132**	.061	.015	.036	**.064**	.013
DisMult	.419	.301	**.428**	.168	.195	**.322**	.232	.053
RotatE	.299	.453	**.475**	.243	.397	.621	**.650**	.238
TransH	.278	.139	**.380**	.192	.455	.407	**.588**	.183
CrossE	.442	.249	**.521**	.154	.158	.227	**.394**	.066

* The values in bold indicate optimal (maximum) values.

- **ACC**: Triple accuracy refers to the average accuracy of triples in the same confidence level.
- **T10ACC**: the average accuracy of the top 10% high-confidence triples. The higher the T10ACC, the better the confidence measurement effect of the confidence measurement method.

C. Comparative Analysis of ECE and T10ACC

The accuracy effects of 3 confidence measures on 6 different KGE prediction models were compared, and the causal intervention confidence measures were compared with the traditional SigmodidMax(SIG) and TopKSoftmax(TOP) confidence measures, which have been applied in very recent literature [23], [24]. ECE and T10ACC scores for each model on different data sets were calculated separately. The two typical calibration methods used in this experiment are Platt Scaling(Platt) [25] and Isotonic Regression(Isoto) [26].

Table I shows the ECE of different confidence measures for drug-target interaction prediction models. For ECE indicators, the SIG is generally higher than the TOP method in the Hetionet data set, and the CI has more obvious effects than the other two methods, and obtains lower ECE error values, indicating that the causal intervention method has achieved results in the stability of embedded representation, making the semantic representation of each entity and relationship more realistic.

For Hetionet in Table II, the CrossE model improves the low-precision prediction results to a relatively high confidence accuracy of 0.521 by the CI, indicating that the accuracy of prediction results of this model is better than other baseline models. However, the overall confidence in the complex model was low due to low ACC scores. In the Hetionet dataset, the ACC of Complex is higher than that in the BioKG dataset. The T10ACC scores based on the SIG and the TOP were similar, but the CI was nearly twice as high, indicating that the causal
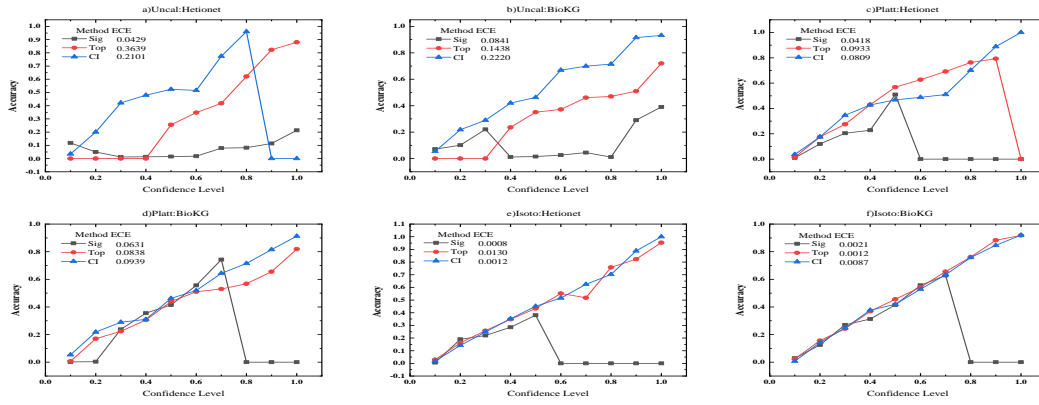
Fig. 2. Reliability diagrams of the RotatE model using different confidence measurement and calibration methods for two datasets: Hetionet(a) and BioKG(b).

intervention method would show more obvious effects with high ACC scores, while the performance was relatively poor with low accuracy.

D. Comparative Analysis of Confidence Measures

This experiment focuses on the differences in probabilistic calibration of different confidence measurement methods for a single KGE model. In this experiment, the RotatE model was selected as the reference model, and three calibration modes, non-calibration, the Platt and the Isoto, were used to analyze the experimental results, where the $X-axis$ represents the average prediction probability and the $Y-axis$ represents the actual accuracy.

As shown in Fig. 2a)–b), the SIG function without calibration performs more clearly for high and low confidence scores, while the Top mainly targets high confidence scores. Compared with the SIG, the TOP is more uniform, and the CI can better predict the drug-target interactions.

The Platt for calibrating the model is shown in Fig.2c)–d). The ECE value has a significant decline, in which the decline of the SIG and the TOP is more obvious, while that of the CI is relatively small. Compared with Fig. 2a)–b), it can be seen that the uncalibrated model is not realistic enough to predict drug-target interactions, and the error is relatively large. Especially for the traditional confidence measurement method, it is often untrue in the high confidence score range. Using the CI, the confidence score is relatively uniform in the overall probability interval distribution, and the ECE decline is relatively low before and after calibration, which also reflects the robustness of the method based on causal intervention compared with the traditional method.

The Isoto for calibrating the model is shown in Fig.2e)–f). The overall calibration fit of the Isoto is better, especially in the TOP and the CI, and ECE indexes are also lower. Among them, the CI is smoother than the TOP, which is close to the optimal calibration baseline. In general, the CI and the Isoto for model calibration show significant effects in both biological data sets.
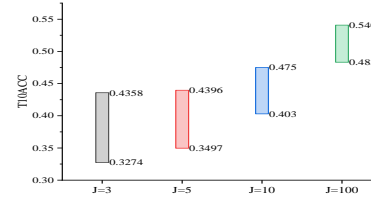

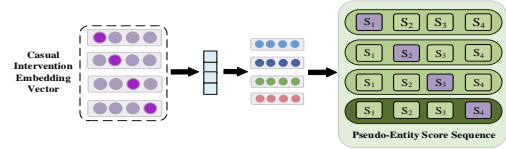Fig. 3. Effect of hyperparameters K and J on T10ACC

Fig. 4. Schematic diagram of pseudo entity score sequence

E. Hyperparameter Experimental Analysis

In the causal intervention, the replacement of each target entity in the predicted triplet will generate corresponding scores. Due to the large number of tail entities, it is necessary to select the tail entity with a high TopK score, and then intervene in the embedded representation of the head entity of the triplet, and combine the TopK tail entity into a triplet to score again. The first J intervention entities are selected as the number of consistent entity measures. Therefore, this experiment analyzed the influence of parameter changes on the results of causal intervention. In this experiment, the DisMult model was selected for testing on the Hetionet dataset to explore the influence of hyperparameters K and J on the causal intervention methods, including the values of ECE and T10ACC. Where, the selected value of J is $\{3, 5, 10, 100\}$, where the value of K should be greater than or equal to $J \leq K$, so the initial value of K changes with the change of J, in this experiment $J \leq K \leq 300$.

As shown in Fig.3, it is the interval distribution diagram of hyperparameters K and J for T10ACC indicators. The
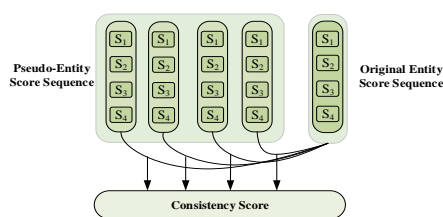
Fig. 5. Schematic diagram of consistency score calculation

bottom of each column interval indicates the lowest T10ACC value, and the hyperparameter $K = J$. The value above each column interval represents the highest T10ACC value, but it does not increase with the increase of K. Due to computational power limitations, there is a significant decrease when K is set to 300, which also indicates that T10ACC can only improve the accuracy of relatively high confidence triples. By comparing the choices of different J, it can be found that T10ACC increases with the increase of parameter J and when $K = J$, which proves that there are more values in the score sequence of the selected intervention entity and the original score sequence, and the consistency comparison is more accurate.

V. Conclusion

In this paper, a causal intervention confidence measure was used to measure drug-target interactions prediction. Experimental results show that the proposed method improves the reliability and confidence of the model compared with the traditional method. However, it is more suitable for low-dimensional embedding representation. In the future, it is necessary to further verify the embedding choice of different dimensions and determine the appropriate dimension vector that takes into account accuracy and efficiency. All in all, this method is a beneficial supplement to the improvement of model performance in the field of drug-target, and has important significance in accelerating the process of new drug research and development, reducing research and development costs, and improving drug safety.

Acknowledgment

References

[1] D. J. Triggle and J. B. Taylor, *Comprehensive Medicinal Chemistry II*. Elsevier, 2006, vol. 8.

[2] H. Wang, J. Wang *et al.*, "A novel approach for drug-target interactions prediction based on multimodal deep autoencoder," *Frontiers in Pharmacology*, vol. 10, p. 1592, 2020.

[3] A. Ezzat, M. Wu *et al.*, "Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey," *Briefings in Bioinformatics*, vol. 20, no. 4, pp. 1337–1357, 2019.

[4] L. Zhang, L. Zhang *et al.*, "Mcdb: A comprehensive curated mitotic catastrophe database for retrieval, protein sequence alignment, and target prediction," *Acta Pharmaceutica Sinica B*, vol. 11, no. 10, pp. 3092–3104, 2021.

[5] K. Ahmadi, F. Zahedifard *et al.*, "Active site-based analysis of structural proteins for drug targets in different human coronaviruses," *Chemical Biology & Drug Design*, vol. 99, no. 4, pp. 585–602, 2022.

[6] L. Hegazy, L. E. Gill *et al.*, "Identification of novel mitochondrial pyruvate carrier inhibitors by homology modeling and pharmacophore-based virtual screening," *Biomedicines*, vol. 10, no. 2, p. 365, 2022.

[7] H. Öztürk, A. Özgür, and E. Ozkirimli, "Deepdta: deep drug–target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, 2018.

[8] K. Wang, Y. Liu, and Q. Z. Sheng, "Neighborhood intervention consistency: Measuring confidence for knowledge graph link prediction." in *IJCAI*, 2021, pp. 2090–2096.

[9] A. C. Nascimento *et al.*, "A multiple kernel learning algorithm for drug-target interaction prediction," *BMC Bioinformatics*, vol. 17, pp. 1–16, 2016.

[10] T. Trouillon, J. Welbl *et al.*, "Complex embeddings for simple link prediction," in *International Conference on Machine Learning*. PMLR, 2016, pp. 2071–2080.

[11] S. Bonner, I. P. Barrett *et al.*, "A review of biomedical datasets relating to drug discovery: a knowledge graph perspective," *Briefings in Bioinformatics*, vol. 23, no. 6, p. bbac404, 2022.

[12] S. Bonner, I. P. Barrett, C. Ye, R. Swiers, O. Engkvist, C. T. Hoyt, and W. L. Hamilton, "Understanding the performance of knowledge graph embeddings in drug discovery," *Artificial Intelligence in the Life Sciences*, vol. 2, p. 100036, 2022.

[13] Z. Wang, J. Zhang *et al.*, "Knowledge graph embedding by translating on hyperplanes," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28, no. 1, 2014.

[14] H. Wang, G. Zhou *et al.*, "Drug-target interaction prediction with graph attention networks," *ArXiv Preprint ArXiv:2107.06099*, 2021.

[15] F. Wang, X. Lei *et al.*, "Predicting drug–drug interactions by graph convolutional network with multi-kernel," *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab511, 2022.

[16] Y. Yamanishi, M. Araki *et al.*, "Prediction of drug-target interaction networks from the integration of chemical and genomic spaces," *Bioinformatics*, vol. 24, no. 13, pp. i232–i240, 2008.

[17] W. Zhang, Y. Chen *et al.*, "Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data," *BMC Bioinformatics*, vol. 18, pp. 1–12, 2017.

[18] Z. Yu, F. Huang *et al.*, "Predicting drug–disease associations through layer attention graph convolutional network," *Briefings in Bioinformatics*, vol. 22, no. 4, p. bbaa243, 2021.

[19] D. Cheng, J. Li *et al.*, "Data-driven causal effect estimation based on graphical causal modelling: A survey," *arXiv preprint arXiv:2208.09590*, 2022.

[20] J. Pearl, *Causality*. Cambridge university press, 2009.

[21] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd International Conference on Machine Learning*, 2005, pp. 625–632.

[22] M. Ali, M. Berrendorf *et al.*, "Pykeen 1.0: a python library for training and evaluating knowledge graph embeddings," *the Journal of Machine Learning Research*, vol. 22, no. 1, pp. 3723–3728, 2021.

[23] P. Tabacof and L. Costabello, "Probability calibration for knowledge graph embedding models," *ArXiv Preprint ArXiv:1912.10000*, 2019.

[24] T. Safavi, D. Koutra, and E. Meij, "Evaluating the calibration of knowledge graph embeddings for trustworthy link prediction," *ArXiv Preprint ArXiv:2004.01168*, 2020.

[25] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in Large Margin Classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

[26] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1321–1330.