Methods 220 (2023) 1–10

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

MATT-DDI: Predicting multi-type drug-drug interactions via heterogeneous attention mechanisms

Shenggeng Lin [a], Xueying Mao [a], Liang Hong [b,c], Shuangjun Lin [a], Dong-Qing Wei [a,d,e], Yi Xiong [a,b,*]

[a] State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China
[b] Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China
[c] School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China
[d] Zhongjing Research and Industrialization Institute of Chinese Medicine, Nanyang 473006, China
[e] Peng Cheng National Laboratory, Shenzhen 518055, China

ARTICLE INFO

Keywords:
Drug-drug interaction
Attention mechanism
Similarity feature
Information leakage

ABSTRACT

The joint use of multiple drugs can result in adverse drug-drug interactions (DDIs) and side effects that harm the body. Accurate identification of DDIs is crucial for avoiding accidental drug side effects and understanding potential mechanisms underlying DDIs. Several computational methods have been proposed for multi-type DDI prediction, but most rely on the similarity profiles of drugs as the drug feature vectors, which may result in information leakage and overoptimistic performance when predicting interactions between new drugs. To address this issue, we propose a novel method, MATT-DDI, for predicting multi-type DDIs based on the original feature vectors of drugs and multiple attention mechanisms. MATT-DDI consists of three main modules: the top k most similar drug pair selection module, heterogeneous attention mechanism module and multi-type DDI prediction module. Firstly, based on the feature vector of the input drug pair (IDP), k drug pairs that are most similar to the input drug pair from the training dataset are selected according to cosine similarity between drug pairs. Then, the vectors of k selected drug pairs are averaged to obtain a new drug pair (NDP). Next, IDP and NDP are fed into heterogeneous attention modules, including scaled dot product attention and bilinear attention, to extract latent feature vectors. Finally, these latent feature vectors are taken as input of the classification module to predict DDI types. We evaluated MATT-DDI on three different tasks. The experimental results show that MATT-DDI provides better or comparable performance compared to several state-of-the-art methods, and its feasibility is supported by case studies. MATT-DDI is a robust model for predicting multi-type DDIs with excellent performance and no information leakage.

1. Introduction

Polypharmacy, the practice of administering multiple medications, is necessary in certain situations, either due to the complexity of a particular disease [1–3], or because one patient with co-occurrence of multiple diseases requires treatment with multiple medications [4–6]. However, drug-drug interactions (DDIs) frequently occur with polypharmacy, where the combination of multiple drugs alters the effects of one or more drugs [4,7,8]. Ideally, treatment involves a synergistic action that results in a therapeutic benefit. However, it is common to encounter adverse DDIs for one patient with polypharmacy, which can lead to toxic side effects or reduce efficacy of the treatment [7,9,10].

Adverse DDIs have become a major healthcare concern in recent years [11,12]. According to a study investigating the correlation between potential DDIs and mortality rates among elderly hospitalized patients, more than 62.77 % of patients experience at least one DDI, with severe cases resulting in fatalities [13]. Adverse DDIs are also a major cause of drug withdrawal from the market [7]. For instance, Terfenadine

Abbreviations: DDIs, drug-drug interactions; IDP, input drug pair; NDP, new drug pair; DNN, deep neural network; KGs, knowledge graphs; GNNs, graph neural networks; MATT, multi-attention; RF, random forest; KNN, k-nearest neighbor; LR, logistic regression.

* Corresponding author at: State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China.
E-mail address: xiongyi@sjtu.edu.cn (Y. Xiong).

https://doi.org/10.1016/j.ymeth.2023.10.007
Received 22 September 2023; Received in revised form 13 October 2023; Accepted 17 October 2023
Available online 18 October 2023

was withdrawn due to its cardiotoxic side effects resulting from its interaction with CYP3A4 isoenzyme inhibitors [14], while Cerivastatin was withdrawn due to drug-related rhabdomyolysis caused by its combination with Gemfibrozil [15].

Given the consequences of DDIs, it is crucial to accurately identify them and gain insight into their underlying mechanisms. The gained information can aid pharmaceutical companies in drug development and assist clinicians and patients in making informative decisions regarding polypharmacy prescriptions [16,17]. However, traditional wet-lab experiments for validation of DDIs are time-consuming and resource-intensive, making frequent and large-scale adoption challenging [18]. In recent years, there has been a growing availability of scientific literature, electronic medical records, population-based reports of adverse events, drug labels, and other related sources [19–21]. As a result, researchers have developed computational methods to predict DDIs [16,22–24].

More recent studies have also focused on developing multi-type DDI prediction methods that offer greater insight into the causal mechanisms of DDIs [16]. For example, Ryu et al. constructed a gold standard DDI dataset using DrugBank [25], which included 192,284 DDI samples associated with 86 DDIs from 191,878 drug pairs [26]. Then, they formulated multi-type DDI prediction as a multi-class classification task and proposed DeepDDI, a deep neural network (DNN) that leverages the structural information of chemical compounds of a drug pair. This approach achieved state-of-the-art performance in predicting multi-type DDIs and provided insight into the underlying mechanisms of these interactions. Several state-of-the-art multi-type DDI prediction methods have been developed based on the DeepDDI architecture. For example, some methods use auto-encoders or the encoder module of Transformer to learn low-dimensional latent features of drugs [27–30]. DNN algorithms are then used for classification [16,18,31–39]. These methods have improved multi-type DDI prediction by incorporating various types of biological information, such as drug targets and enzymes, to represent a drug pair in addition to the structural information of drugs [16,18,31,40–42]. It is noteworthy that those methods represent the feature vector of a drug by the similarity profile, which is generated by the similarity (i.e., structural similarity) of a given drug against each one in the rest of drugs across the entire dataset. However, relying on similarity profiles as feature vectors for drugs may result in information leakage when calculating the similarity profiles of drugs against the entire dataset which consists of the training and test set. Moreover, on some DDI prediction tasks, drugs in the test set are absent in the training set. In this case, it is not feasible to use the similarity profiles as feature vectors of drugs. Specifically, most methods calculate the similarity of each of two drugs across the entire dataset and use the similarity as the feature vectors of the drugs. Then, these drugs are divided into the training set and the test set. However, since the similarity is obtained in the entire dataset, the feature vectors of the drugs in the training set already contain the drug information in the test set, so this leads to information leakage. The DDI prediction task consists of three tasks: predicting unobserved interaction types between known drugs (Task 1), predicting interaction types between known drugs and new drugs (Task 2), and predicting interaction types between new drugs (Task 3). Known drugs refer to drugs in the training set, and new drugs refer to drugs in the test set. Therefore, for Task 1, the similarity feature vectors do not cause information leakage because all the drugs are in the training set. However, for Task 2 and Task 3, the similarity feature vectors in the training set contain the information of the drugs in the test set, which leads to information leakage.

Around the same time DeepDDI appeared, Zitnik et al. proposed a novel approach for predicting polypharmacy side effects by using a graph neural network model [43,44]. Zitnik et al. formulated the polypharmacy side effect identification task as a multi-relational link prediction problem using a two-layer multimodal graph/network consisting of two types of nodes, including drugs and proteins. The construction of the two-layer multimodal network is accomplished by building a protein–protein interaction network to depict interactions among proteins, a drug-drug interaction network containing 964 different types of edges to illustrate which drug pairs lead to specific side effects, and drug-protein links that show the proteins targeted by a given drug. As the amount of large biomedical knowledge graphs (KGs) grows, some studies have begun to explore the integration of KGs with other data sources, such as drug molecular structures, for multi-type DDI predictions using graph neural networks (GNNs) [5,32,45–57]. Nonetheless, large KGs often contain redundant and noisy data, and only a small sub-graph may be relevant to the prediction targets [47,58]. Therefore, KG-based prediction methods for DDIs are still in their early stages of development [59].

In more recent work, Deng et al. employed few-shot learning based on latent features derived from a pair of drug structures to enhance the prediction performance for rare types of DDIs that have limited samples [60]. Lin et al. proposed a supervised contrastive learning-based method, MDDI-SCL, to predict multiple type DDIs [16]. Liu et al. proposed the CSMDDI method, which generated embeddings of drugs and DDIs and then learned a mapping function to link drug attributes to their embedding to predict multi-type DDIs [37].

Indeed, despite the progress made in multi-type DDI prediction, some limitations still exist. Firstly, many existing methods use similarity profiles as drug feature vectors, which can lead to information leakage and inaccurate model evaluation, particularly when predicting interactions for new drugs [61–69]. Secondly, while most methods perform well in predicting unknown DDIs between known drugs, they often fail to do so for new drugs [68,70]. In addition, while these models have achieved satisfactory performance in some DDI prediction tasks, they are often difficult to interpret. Interpretable deep learning methods are needed to further help researchers understand DDIs [71]. Therefore, it is necessary to develop new approaches to address these issues and further improve prediction performance.

To overcome these limitations, we propose a new method named MATT-DDI (Fig. 1) for multi-type DDI prediction, which is based on Multi-ATTention (MATT) mechanisms. Firstly, based on the feature vectors of the input drug pairs (IDP) and the cosine similarity between drug pairs, k drug pairs that are most similar to the input drug pairs are selected from the training dataset. Then, the vectors of k selected drug pairs are averaged to obtain a new drug pair (NDP). Next, IDP and NDP are input into multiple attention modules, which include the scaled dot product attention and bilinear attention, to extract latent feature vectors. Finally, the latent feature vectors are taken as input into the classification module to predict the DDIs.

Experimental results demonstrate that MATT-DDI achieves better or comparable performance than several state-of-the-art methods on all three tasks. Additionally, we also demonstrate that employing similarity as the drug feature vector may result in information leakage, leading to an overestimation of the model's performance. Furthermore, the results of case studies confirm the practical feasibility of our proposed method.

2. Material and methods

2.1. Datasets and drug feature representation

In this study, we utilized a benchmark dataset collected by Deng et al. [31]. The dataset comprises of 572 drugs with 74,528 pairwise DDI samples that are associated with 65 types of DDIs. Detailed descriptions of the 65 types of DDIs are shown in Supporting information Table S1. Each drug in dataset has four types of features: substructures, targets, pathways and enzymes. According to Deng et al.'s experiments, the combination of substructures, targets and enzymes performs best in all combinations of features. Therefore, our model also uses only these three types of drug features.

Each feature type of a drug corresponds to a specific set of descriptors. Therefore, a drug can be represented by using a binary feature vector where the value (1 or 0) indicates the presence or absence of the

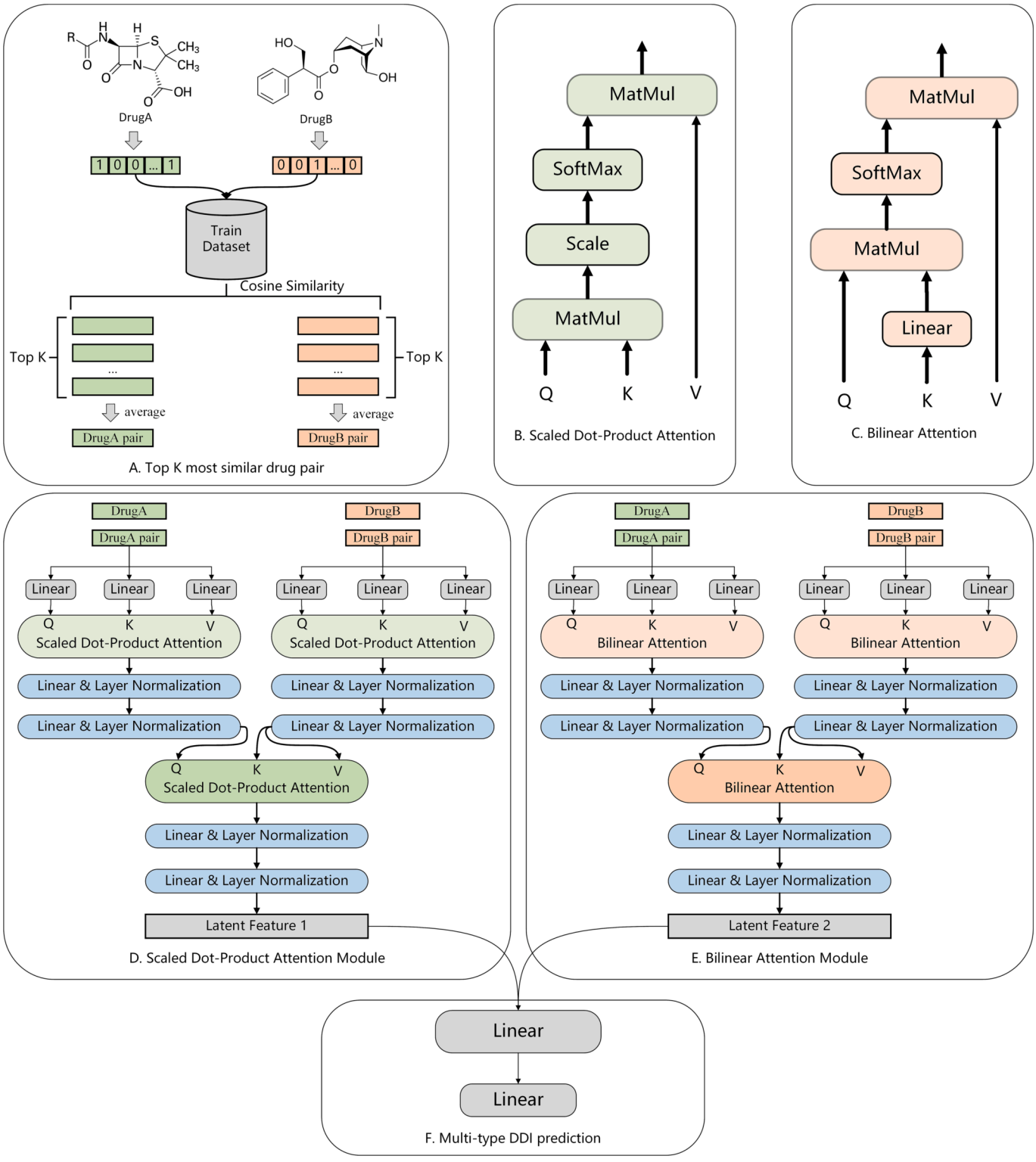

Fig. 1. Overview of the proposed MATT-DDI method. (A) The module for selection of top k most similar drug pairs. (B) Scaled dot-product attention mechanism. (C) Bilinear attention mechanism. (D) Scaled dot-product attention module. (E) Bilinear attention module. (F) Multi-type DDI prediction module.

corresponding element. Unlike previous studies [18,31], we did not calculate the similarity profiles of each drug, but directly used the original binary vector as the feature vector of each drug.

2.2. The module for selection of top k most similar drug pairs

We represent each drug pair as a binary feature vector. During the model training process, we identify the k most similar drug pairs from the training dataset for each drug pair according to the cosine similarity between drug pairs, and these pairs are jointly used as input to train the model. It is worth noting that we continue to use binary feature vectors as the feature representations of drug pairs, not similarity features, which distinguishes our method from previous methods [18,31]. And original binary vectors in the training set don't contain the information

of the drugs in the test set, so this can avoid information leakage caused by similarity profiles. During the model validation process, we retrieve the k most similar drug pairs from the training dataset for each drug pair in the test dataset according to the cosine similarity between drug pairs, and these pairs are used together in the model for prediction.

According to the assumption that similar drugs are more likely to have similar functions, the use of k most similar drug pairs not only enables the model to utilize the information contained in the feature vectors of the input drug pairs, but also refer to the information contained in the feature vectors of the similar drug pairs, thus improving the performance of the model. In addition, the use of similar drug pairs also introduces noise while increasing the information, which can alleviate the over-fitting problem of the model, and this is also verified in the experimental section below. To our knowledge, we are the first method to use similar drug pairs in the training set to predict multi-type DDIs.

2.3. Heterogeneous attention mechanism module

The attention mechanism has gained widespread use in the domain of deep learning, particularly in natural language processing [27]. Notably, the method of calculating attention scores stands out as the key differentiator among the different attention mechanisms. Our model primarily relies on two attention mechanisms, namely the scaled dot-product attention and the bilinear attention. The scaled dot-product attention was initially introduced in Transformer [27]. The bilinear attention was first proposed in BAN [72]. Both attention scores can be calculated as Formula 1 and Formula 2, respectively.

$$\mathrm{Attention}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \mathrm{softmax}\left(\frac{\mathrm{QK}^{\mathrm{T}}}{\sqrt{\mathrm{d_k}}}\right)\mathrm{V} \tag{1}$$

$$\mathrm{Attention}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \mathrm{softmax}(\mathrm{Q}^{*}(\mathrm{W}^{*}\mathrm{K})^{\mathrm{T}})\mathrm{V} \tag{2}$$

where Q (Query), K (Key) and V (Value) matrices are derived from the linear transformation of drug feature vectors, respectively. W is a learnable parameter matrix and d_k is the dimension of K.

The scaled dot-product attention mechanism and bilinear attention mechanism differ in how they calculate the distribution of attention. The former uses a dot-product operation, while the latter uses a bilinear function. The calculation of dot-product is faster, but gradient explosion may occur. In contrast, the bilinear function is slower but can avoid this problem. Additionally, input vectors in scaled dot-product attention mechanisms require normalization, while bilinear attention mechanisms do not. These mechanisms are commonly used in different applications, with the scaled dot-product attention mechanism in sequence-to-sequence models such as machine translation and text summarization, and the bilinear attention mechanism in image-to-text models such as image description generation and visual question answering.

Both mechanisms have their own advantages and limitations. The scaled dot-product attention mechanism is faster to calculate and has high accuracy in capturing similarity between input vectors. However, it may be sensitive to input vector length, leading to decreased accuracy when the lengths of vectors are long. The bilinear attention mechanism has stronger expressiveness, can model nonlinear interactions between input vectors with higher accuracy, but is computationally complex. High input vector length and dimension can cause a sharp increase in computation and result in the problem of gradient vanishing [72]. Therefore, in order to combine the advantages of scaled dot-product attention mechanism and bilinear attention mechanism, our model uses both attention mechanisms.

2.4. Multi-type DDI prediction module

The module employs two fully connected layers to predict multi-type DDIs, and the number of neurons in the second fully connected layer is the number of DDI types. We utilize Gaussian error linear unit activation function and Adam optimizer. The dropout layer and batch normalization layer are placed between the fully connected layers [73].

3. Results and discussion

3.1. Experimental settings of prediction tasks

This study conducts an evaluation of multi-type DDI prediction tasks using three experimental settings. The first setting involves predicting unobserved interaction types between known drugs (Task1). The second setting involves predicting interaction types between known drugs and new drugs (Task2). In this scenario, the new drugs from the test set are not included in the training set. The third setting involves predicting interaction types between new drugs (Task3), where both drugs from the test set are not observed in the training set.

For Task1, we apply five-fold cross-validation (5-CV) test to DDIs, where all DDI samples are split into five subsets. Models are trained on the DDI samples in four of these subsets, and then used to make predictions for the DDI samples in the remaining subset. For Task2 and Task3, we apply 5-CV test to drugs instead of DDIs. The drugs are randomly split into five subsets, with four subsets used as training drugs and one subset used as test drugs. For Task2, prediction models are trained on the known DDIs between two drugs in the training set, and then used to predict DDIs between training set and test set. For Task3, prediction models are built on the known DDIs between two drugs in the training set, which are then used to predict DDIs between two drugs in the test set.

Several evaluation metrics are used to evaluate the performance of the multi-type DDI prediction models, including accuracy (ACC), area under the precision-recall-curve (AUPR), area under the ROC curve (AUC), F1 score, precision, and recall.

All experiments are performed on our own server. The operating system of the server is ubuntu 20.04.3, the memory size of the server is 256 GB, the processor is Intel(R) Xeon(R) Gold 6240, and the GPU is GeForce RTX 3090.

3.2. Hyper-parameters setting

Hyper-parameters play a crucial role in the performance of machine learning models. In this study, four hyper-parameters (including pair number, batch size, learning rate and training epochs) were tuned to determine their effects on the performance of MATT-DDI for Task2.

Pair number determines the number of k most similar drug pairs. Batch size determines the number of samples used in each iteration of model training. A larger batch size can lead to faster convergence. The learning rate determines the step size at each iteration during model training to update the weights. A higher learning rate can help the model converge faster, but it may also cause the model to overshoot the optimal weights and converge to a sub-optimal solution. Training epochs determine the number of times the entire training dataset is passed through the model during training. Increasing the number of epochs can improve the model's accuracy, but too many epochs can lead to over-fitting. The metric scores under different configurations are shown in Fig. 2.

According to Fig. 2, the performance of the model does not change drastically as the hyper-parameters change. Almost all metrics vary within the range of 0.01. These results illustrate the stability of our model. Finally, we chose 50 for pair number, 128 for batch size, 1e-5 for learning rate and 100 for training epoch. For Task 1 and Task 3, we also choose the same hyper-parameters.

3.3. The effect of similar drug pairs

In order to verify whether the use of similar drug pairs in the training set contributes to multi-type DDIs prediction, we conducted the ablation

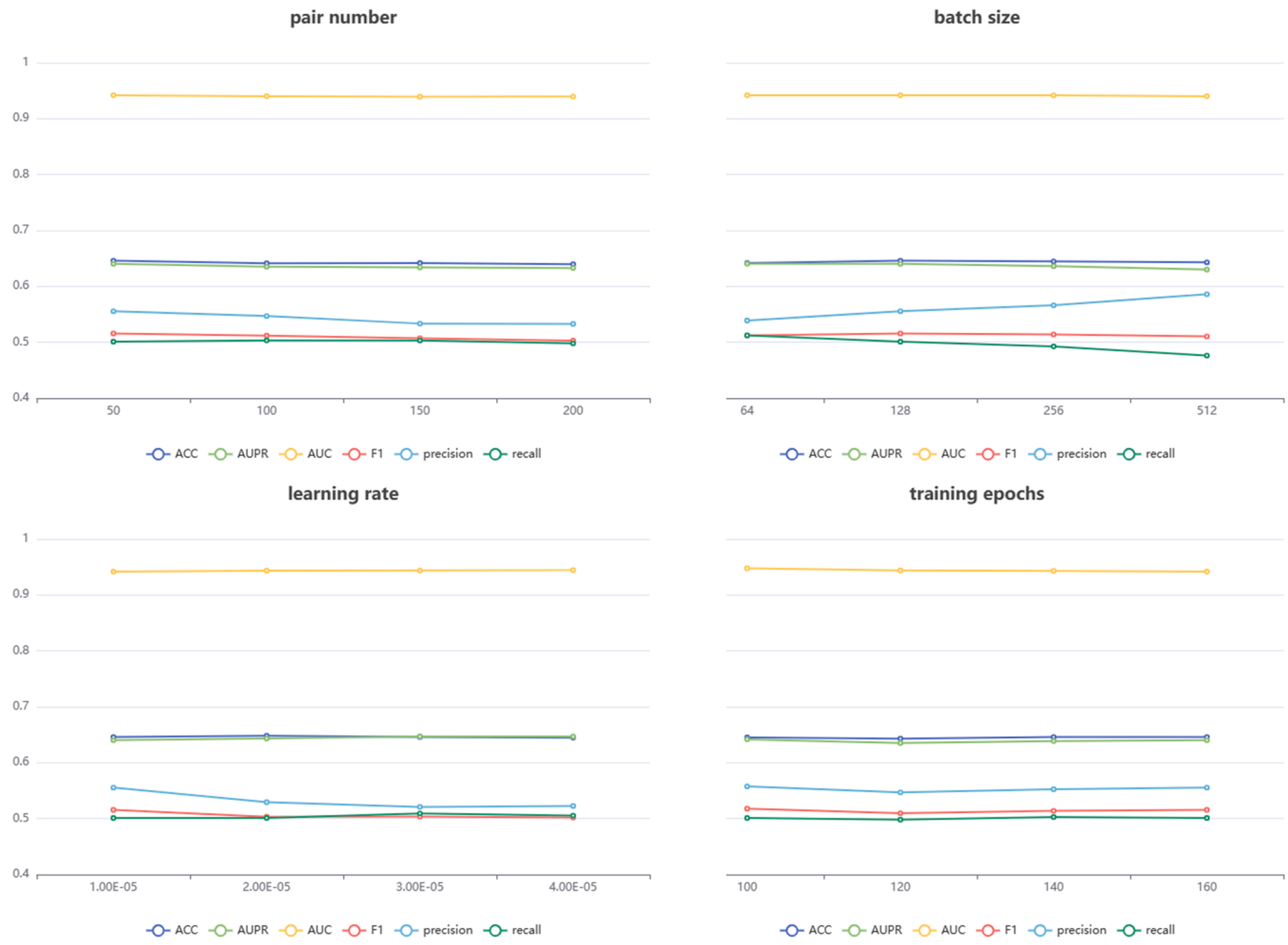

Fig. 2. The prediction performance of four hyper-parameters settings on Task2.

experiment. According to the experimental results shown in Fig. 3, using or not using similar drug pairs have slight effect on the model performance, the performance difference of models with or without similar drug pairs is within 0.01. Specifically, on Task 1, models with similar drug pairs performer slightly better than models without similar drug pairs. On Task 2 and Task 3, models without similar drug pairs performed slightly better than those with similar drug pairs. This may be because Task 1 is a relatively simple task, so the model may be prone to over-fitting. The use of similar drug pairs is equivalent to adding noise to the original drug pair, which can alleviate the problem of over-fitting. Task 2 and Task 3 are relatively difficult tasks. By using similar drug pairs as inputs, the difficulty of prediction task is increased, resulting in slightly worse performance of the model. But either way, using or not using similar drug pairs as inputs didn't make much difference to the

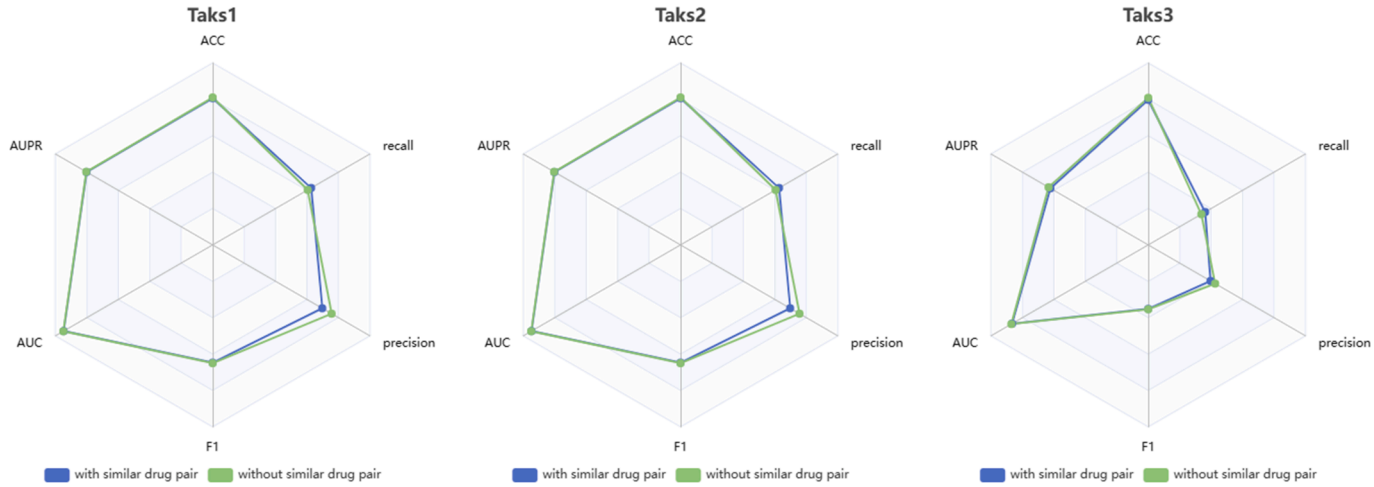

Fig. 3. The prediction effect of similar drug pairs on three tasks.

model's performance.

3.4. The effect of heterogeneous attention mechanisms

In order to verify whether the heterogeneous attention mechanism works in our model, we verified the effectiveness of the heterogeneous attention mechanisms on all three tasks. The specific evaluation method is to compare the performance of the model with only scaled dot-product attention (MATT-DDI-SDP), the model with only bilinear attention (MATT-DDI-BA), and the model with both scaled dot-product attention and bilinear attention on three tasks. The effectiveness of heterogeneous attention mechanisms is determined by comparing the metric scores. The results of all prediction models are shown in Table 1.

On all three tasks, the F1 score and recall score of MATT-DDI are higher than that of MATT-DDI-SDP and MATT-DDI-BA. On task1, the F1 score of MATT-DDI is about 0.02 higher than that of MATT-DDI-SDP. On task2, the F1 score of MATT-DDI is 0.02 higher than that of MATT-DDI-BA. More importantly, according to the experimental results, MATT-DDI-BA performs better on all metric scores than MATT-DDI-SDP on Task 1. But On task 2 and task 3, MATT-DDI-SDP performs better on all metric scores than MATT-DDI-BA. Therefore, the experimental results show that bilinear attention is better at solving task 1, and scaled dot-product attention is better at solving task 2. All three tasks are better solved by using both scaled dot-product attention and bilinear attention. This also proves the effectiveness of heterogeneous attention mechanisms.

3.5. The effect of similarity feature vectors

In order to explore the effect of using similarity feature vectors and binary feature vectors on prediction performance. We compared the performance of three models, MATT-DDI, MDF-SA-DDI [18] and MDDI-SCL [16], using similarity feature vectors and original feature vectors on the benchmark dataset collected by Deng et al.. The details of the dataset are described in Section 2.1 in detail. MDF-SA-DDI is one of the state-of-the-art methods, and it predicts multi-type DDIs based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. MDDI-SCL predicts multi-type DDIs based on supervised contrastive learning. The experimental results of two models are shown in Table 2.

According to Table 2, on Task 1, the model using the binary feature vectors outperforms the model using the similarity feature vectors. On Task 2, the performance of the model using similarity feature vectors is slightly better than that using binary feature vectors. On Task 3, the performance of the model using similarity feature vectors is obviously better than that using binary feature vectors. On Task 1, drugs in the test set can appear in the training set, while on Task 3, drugs in the test set cannot appear in the training set. However, on Task 3, the similarity feature vectors leak the drug information in the test set, so the performance of the model using the similarity feature vectors is significantly better than that using the binary feature vectors. However, this does not mean that the similarity feature vectors are superior to the binary feature vectors. On Task 1, the performance of the model using the similarity feature vectors is not better than that using the original feature vectors, because the similarity feature vectors lose the bias of information leakage on Task 1.

3.6. Comparison with several state-of-the-art methods

Due to the fact that most of the previous models use similarity feature vectors as drug feature vectors, we first compare MATT-DDI with other five state-of-the-art methods for prediction of multi-type DDIs: DeepDDI [26], Lee et al.'s methods [40], DDIMDL [31], MDDI-SCL [16] and MDF-SA-DDI [18], and also several baseline classification methods, such as fully connected DNN, random forest (RF), k-nearest neighbor (KNN) and logistic regression (LR) using similarity feature vectors. The performance comparison of all prediction models is shown in Table 3.

According to Table 3, MATT-DDI is significantly superior to all methods except MDF-SA-DDI and MDDI-SCL on all three tasks. And MATT-DDI achieves comparable performance to MDF-SA-DDI and MDDI-SCL. On Task1, MATT-DDI achieves the ACC 0.9301 and the AUPR 0.9755. The F1 score of MATT-DDI is 0.8647, which is 0.0231 worse than that of MDF-SA-DDI. On Task 2 and 3, the performance of MATT-DDI is slightly worse than MDF-SA-DDI and MDDI-SCL, in which similarity features can cause information leakage. We compared the model performance of MATT-DDI, MDF-SA-DDI and MDDI-SCL using original feature vector, (results are shown in Table 4). According to Table 4, MATT-DDI achieves comparable performance to MDF-SA-DDI and MDDI-SCL on Task 1, and the performance of MATT-DDI is significantly better than MDF-SA-DDI and MDDI-SCL on Task 2 and 3. On Task 2, MATT-DDI achieves the highest ACC 0.6452, AUPR 0.6418, AUC 0.9477, and recall 0.5011. The AUPR score of MATT-DDI is around 0.02 and 0.05 higher than that of MDF-SA-DDI and MDDI-SCL on Task 2, respectively. On Task 3, MATT-DDI also achieves the highest AUC and recall. The F1 score of MATT-DDI is around 0.05 higher than that of MDDI-SCL on Task 3.

In general, MATT-DDI outperforms all methods except for MDF-SA-DDI and MDDI-SCL on all three tasks when using similarity features as drug feature vectors. However, MATT-DDI is significantly superior to the MDF-SA-DDI and MDDO-SCL when using binary feature vectors as model input.

3.7. Case studies

In this section, we conducted case studies to further validate the effectiveness of MATT-DDI in practice. We used all the DDI samples on dataset collected by Deng et al. to train the prediction model, and then predicted the drug-drug pairs that do not exist on the dataset. We focused on the five most frequent DDI types and checked up the top 20 predictions related to each type. The most frequent DDI type refers to the DDI type with the largest sample size. We used the interactions checker tool provided by https://go.drugbank.com to verify these predictions.

Among the total of 100 samples, 35 DDI samples were confirmed, which are shown in Supporting information Table S2. For example, the interaction between Enasidenib and Nateglinide is predicted to cause the DDI type #0, which means that metabolism of Enasidenib can be

Table 1
The prediction effect of heterogeneous attention mechanisms on three tasks.

		ACC	AUPR	AUC	F1	Precision	Recall
Task 1	MATT-DDI	0.938	0.977	0.9987	0.8851	0.9229	0.8661
	MATT-DDI-BA	0.9377	0.9786	0.9989	0.8810	0.9115	0.8628
	MATT-DDI-SDP	0.9375	0.9780	0.9989	0.8658	0.9019	0.8497
Task 2	MATT-DDI	0.6452	0.6418	0.9477	0.5179	0.5576	0.5011
	MATT-DDI-BA	0.6347	0.6291	0.9508	0.4981	0.5475	0.4783
	MATT-DDI-SDP	0.6435	0.6406	0.9524	0.5082	0.5638	0.4847
Task 3	MATT-DDI	0.3989	0.3112	0.8657	0.1759	0.198	0.1809
	MATT-DDI-BA	0.3890	0.3060	0.8718	0.1663	0.1952	0.1675
	MATT-DDI-SDP	0.3993	0.3155	0.8745	0.1696	0.2058	0.1708

Table 2
The prediction effect of similarity feature vectors on three tasks.

		ACC	AUPR	AUC	F1	Precision	Recall
Task 1	MATT-DDI similarity feature	0.9301	0.9755	0.9988	0.8647	0.8771	0.8613
	MATT-DDI original feature	0.938	0.977	0.9987	0.8851	0.9229	0.8661
	MDF-SA-DDI similarity feature	0.9301	0.9737	0.9989	0.8878	0.9085	0.876
	MDF-SA-DDI original feature	0.937	0.9767	0.999	0.8904	0.9207	0.8765
	MDDI-SCL similarity feature	0.9378	0.9782	0.9983	0.8755	0.8804	0.8767
	MDDI-SCL original feature	0.9256	0.9742	0.9985	0.8605	0.8701	0.8615
Task 2	MATT-DDI similarity feature	0.6557	0.6581	0.9535	0.5315	0.5546	0.532
	MATT-DDI original feature	0.6452	0.6418	0.9477	0.5179	0.5576	0.5011
	MDF-SA-DDI similarity feature	0.6633	0.6776	0.9497	0.5584	0.6547	0.5078
	MDF-SA-DDI original feature	0.6254	0.6175	0.9267	0.5283	0.6198	0.4861
	MDDI-SCL similarity feature	0.6767	0.6947	0.9634	0.5304	0.6254	0.4814
	MDDI-SCL original feature	0.5953	0.5904	0.9433	0.2934	0.3518	0.2822
Task 3	MATT-DDI similarity feature	0.4202	0.3413	0.8776	0.2015	0.214	0.2132
	MATT-DDI original feature	0.3989	0.3112	0.8657	0.1759	0.198	0.1809
	MDF-SA-DDI similarity feature	0.4338	0.3873	0.863	0.2329	0.2715	0.2226
	MDF-SA-DDI original feature	0.3921	0.3142	0.8088	0.1761	0.2125	0.1752
	MDDI-SCL similarity feature	0.4589	0.3938	0.9053	0.1919	0.2585	0.1678
	MDDI-SCL original feature	0.4045	0.3229	0.8597	0.1285	0.1553	0.1332

Table 3
Prediction performance comparison with the state-of-the-art methods.

		ACC	AUPR	AUC	F1	Precision	Recall
Task1	MATT-DDI	0.9301	0.9755	0.9988	0.8647	0.8771	0.8613
	MDF-SA-DDI	0.9301	0.9737	0.9989	0.8878	0.9085	0.8760
	MDDI-SCL	0.9378	0.9782	0.9983	0.8755	0.8804	0.8767
	DDIMDL	0.8852	0.9208	0.9976	0.7585	0.8471	0.7182
	Lee et al.'s methods	0.9094	0.9562	0.9961	0.8391	0.8509	0.8339
	DeepDDI	0.8371	0.8899	0.9961	0.6848	0.7275	0.6611
	DNN	0.8797	0.9134	0.9963	0.7223	0.8047	0.7027
	RF	0.7775	0.8349	0.9956	0.5936	0.7893	0.5161
	KNN	0.7214	0.7716	0.9813	0.4831	0.7174	0.4081
	LR	0.7920	0.8400	0.9960	0.5948	0.7437	0.5236
Task2	MATT-DDI	0.6557	0.6581	0.9535	0.5315	0.5546	0.532
	MDF-SA-DDI	0.6633	0.6776	0.9497	0.5584	0.6547	0.5078
	MDDI-SCL	0.6767	0.6947	0.9634	0.5304	0.6254	0.4814
	DDIMDL	0.6415	0.6558	0.9799	0.4460	0.5607	0.4319
	Lee et al.'s methods	0.6405	0.6244	0.9247	0.5039	0.5388	0.4891
	DeepDDI	0.5774	0.5594	0.9575	0.3416	0.3630	0.3890
	DNN	0.6239	0.6361	0.9796	0.2997	0.4237	0.2840
Task3	MATT-DDI	0.4202	0.3413	0.8776	0.2015	0.214	0.2132
	MDF-SA-DDI	0.4338	0.3873	0.8630	0.2329	0.2715	0.2226
	MDDI-SCL	0.4589	0.3938	0.9053	0.1919	0.2585	0.1678
	DDIMDL	0.4075	0.3635	0.9512	0.1590	0.2408	0.1452
	Lee et al.'s methods	0.4097	0.3184	0.8302	0.2022	0.2216	0.2027
	DeepDDI	0.3602	0.2781	0.9059	0.1373	0.1586	0.1450
	DNN	0.4087	0.3776	0.9550	0.1152	0.1836	0.1093

decreased when combined with Nateglinide. The interaction between Meclizine and Netupitant is predicted to cause the DDI type #2, which means that metabolism of Netupitant can be decreased when combined with Meclizine. In addition, we found that some drugs were highly correlated with certain types of drug interactions. For example, Netupitant is associated with all 20 samples of DDI type #3. This means that Netupitant is likely to be associated with a decrease in the serum concentration of the drugs. In Supporting information Table S3, we list the other 65 drug pairs among the 100 DDI samples. The interactions of these drug pairs are not reported in the DrugBank, but these DDIs are

Table 4
Comparison of prediction between MATT-DDI, MDF-SA-DDI and MDDI-SCL using original feature vectors.

		ACC	AUPR	AUC	F1	Precision	Recall
Task 1	MATT-DDI	0.938	0.977	0.9987	0.8851	0.9229	0.8661
	MDF-SA-DDI	0.937	0.9767	0.999	0.8904	0.9207	0.8765
	MDDI-SCL	0.9256	0.9742	0.9985	0.8605	0.8701	0.8615
Task 2	MATT-DDI	0.6452	0.6418	0.9477	0.5179	0.5576	0.5011
	MDF-SA-DDI	0.6254	0.6175	0.9267	0.5283	0.6198	0.4861
	MDDI-SCL	0.5953	0.5904	0.9433	0.2934	0.3518	0.2822
Task 3	MATT-DDI	0.3989	0.3112	0.8657	0.1759	0.198	0.1809
	MDF-SA-DDI	0.3921	0.3142	0.8088	0.1761	0.2125	0.1752
	MDDI-SCL	0.4045	0.3229	0.8597	0.1285	0.1553	0.1332

likely to occur when taken together, which may be helpful for pharmaceutical research.

To make a fair comparison with other state-of-the-art models, we only used the DrugBank database to verify the model's predictions in case studies. But in fact, some of the predictions not confirmed by DrugBank can be confirmed by other databases or the latest literature. For example, the interaction between Folic acid and Nicardipine is not confirmed by the DrugBank database, but can be confirmed by the TWOSIDES [4] database. The interaction between Midodrine and Lubiprostone is also not confirmed by DrugBank, but the combination of Middrine and Lubisprostone can treat Lewy body dementia [74].

3.8. Limitations of the method and future work

Although MATT-DDI has achieved satisfactory performance, there are still some limitations to be solved. For example, the time complexity of the scaled dot-product attention mechanism used in the model is O (N^2), where N is the dimensionality of the input vector. When the vector dimensionality increases, this time complexity is unacceptable. Therefore, in future work, we could also explore attention mechanisms with lower time complexity, such as Efficient Attention [75] and External Attention [76].

In addition, the interpretability of the model is significant for the application of the model, and we will further study the interpretability of the model. The structure of the drug determines the function of the drug, so we will try to find out the relationship between the drug substructures and DDIs by the attention model. A feasible approach is to use attention models to evaluate and visualize drug substructures with attention scores, and then interpret the model's prediction results based on attention scores and expert knowledge.

4. Conclusions

We proposed a multi-type DDI prediction model based on heterogeneous attention mechanisms. Evaluated by 5-CV, the performance of MATT-DDI model is better than that of state-of-the-art models when not using similarity features as drug feature vectors. In addition, we also proved that the use of similarity feature vectors may lead to information leakage, so that the model can get more optimistic performance than the actual performance. Last but not least, the effectiveness of our model is supported by case studies in practice.

Data and software availability

The source code and the dataset are freely available at https://github.com/ShenggengLin. The data used in the case studies section of this manuscript can be found in the supporting information.

Funding

This work is supported by grants from the National Science Foundation of China (Grant Nos. 62172274, 32070662).

CRediT authorship contribution statement

Shenggeng Lin: Conceptualization, Resources, Methodology. **Xueying Mao:** Investigation. **Liang Hong:** Funding acquisition. **Shuangjun Lin:** Funding acquisition. **Dong-Qing Wei:** Funding acquisition. **Yi Xiong:** Writing – review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymeth.2023.10.007.

References

[1] K.A. Ryall, A.C. Tan, Systems biology approaches for advancing the discovery of effective drug combinations, J. Cheminform. 7 (2015) 7.
[2] Y. Qiu, Y. Zhang, Y. Deng, S. Liu, W. Zhang, A Comprehensive Review of Computational Methods For Drug-Drug Interaction Detection, IEEE/ACM Trans. Comput. Biol. Bioinf. 19 (4) (2022) 1968–1985.
[3] Y. Cheng, Y. Gong, Y. Liu, B. Song, Q. Zou, Brief. Bioinform. 22 (2021) bbab344.
[4] N.P. Tatonetti, P.P. Ye, R. Daneshjou, R.B. Altman, Data-driven prediction of drug effects and interactions, Sci. Transl. Med. 4 (2012) 125r–131r.
[5] K. Nyamabo Arnold, Yu Hui, Zun Liu, Jian-Yu Shi, Drug–drug interaction prediction with learnable size-adaptive molecular substructures, Brief. Bioinform. 23 (2022) bbab441.
[6] J. Zhu, Y. Liu, Y. Zhang, D. Li, Attribute Supervised Probabilistic Dependent Matrix Tri-Factorization Model for the Prediction of Adverse Drug-Drug Interaction, IEEE J. Biomed. Health Inform. 25 (7) (2021) 2820–2832.
[7] T.H. Vo, N.T.K. Nguyen, Q.H. Kha, N.Q.K. Le, On the road to explainable AI in drug-drug interactions prediction: A systematic review, Comput. Struct. Biotechnol. J. 20 (2022) 2112–2123.
[8] J. Wang, J. Mao, M. Wang, X. Le, Y. Wang, Explore drug-like space with deep generative models, Methods 210 (2023) 52–59.
[9] R. Su, X. Liu, L. Wei, Q. Zou, Deep-Resp-Forest: A deep forest model to predict anti-cancer drug response, Methods 166 (2019) 91–102.
[10] H. Qian, C. Lin, D. Zhao, S. Tu, L. Xu, AlphaDrug: protein target specific de novo molecular generation, PNAS Nexus 1 (2022) c227.
[11] E.D. Kantor, C.D. Rehm, J.S. Haas, A.T. Chan, E.L. Giovannucci, Trends in Prescription Drug Use Among Adults in the United States From 1999–2012, J. Am. Med. Assoc. 314 (17) (2015) 1818.
[12] C. Yan, G. Duan, Y.i. Pan, F.-X. Wu, J. Wang, DDIGIP: predicting drug-drug interactions based on Gaussian interaction profile kernels, BMC Bioinf. 20 (S15) (2019).
[13] O. Rosas-Carrasco, C. García-Peña, S. Sánchez-García, G. Vargas-Alarcón, L. M. Gutiérrez-Robledo, T. Juárez-Cedillo, The relationship between potential drug-drug interactions and mortality rate of elderly hospitalized patients, Revista De Investigacion Clinica-Clinical and Translational Investigation 63 (2011) 564–573.
[14] A. Olasinska-Wisniewska, J. Olasinski, S. Grajek, Cardiovascular safety of antihistamines, Postepy Dermatol. Alergol. 31 (2014) 182–186.
[15] C.D. Furberg, B. Pitt, Withdrawal of cerivastatin from the world market, Curr. Control. Trials Cardiovasc. Med. 2 (2001) 205–207.

[16] S. Lin, W. Chen, G. Chen, S. Zhou, D.-Q. Wei, Y.i. Xiong, MDDI-SCL: predicting multi-type drug-drug interactions via supervised contrastive learning, J. Cheminform. 14 (1) (2022).

[17] N. Liu, C.-B. Chen, S. Kumara, Semi-Supervised Learning Algorithm for Identifying High-Priority Drug-Drug Interactions Through Adverse Event Reports, IEEE J. Biomed. Health Inform. 24 (1) (2020) 57–68.

[18] S. Lin, Y. Wang, L. Zhang, Y. Chu, Y. Liu, Y. Fang, M. Jiang, Q. Wang, B. Zhao, Y. Xiong, D.-Q. Wei, MDF-SA-DDI: predicting drug-drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism, Brief. Bioinform. 23 (2022) bbab421.

[19] S. Liu, B. Tang, Q. Chen, X. Wang, Drug-Drug Interaction Extraction via Convolutional Neural Networks, Comput. Math. Methods Med. 2016 (2016) 1–8.

[20] Y. Zhang, W. Zheng, H. Lin, J. Wang, Z. Yang, M. Dumontier, Drug-drug interaction extraction via hierarchical RNNs on sequence and shortest dependency paths, Bioinformatics 34 (2018) 828–835.

[21] T. Song, X. Zhang, M. Ding, A. Rodriguez-Paton, S. Wang, G. Wang, DeepFusion: A deep learning based multi-scale feature fusion method for predicting drug-target interactions, Methods 204 (2022) 269–277.

[22] S. Vilar, C. Friedman, G. Hripcsak, Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media, Brief. Bioinform. 19 (2018) bbx010.

[23] Z. Zhao, Z. Yang, L. Luo, H. Lin, J. Wang, Drug drug interaction extraction from biomedical literature using syntax convolutional neural network, Bioinformatics 32 (2016) 3444–3453.

[24] M. Fatehifar, H. Karshenas, Drug-Drug interaction extraction using a position and similarity fusion-based attention mechanism, J. Biomed. Inform. 115 (2021) 103707.

[25] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R;. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Research 46 (2018) D1074–D1082.

[26] J.Y. Ryu, H.U. Kim, S.Y. Lee, Deep learning improves prediction of drug–drug and drug–food interactions, Proc. Natl. Acad. Sci. 115 (2018) E4304–E4311.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, Adv. Neural Inf. Proces. Syst. (2017) 6000–6010.

[28] T. Wang, J. Sun, Q.i. Zhao, Investigating cardiotoxicity related with hERG channel blockers using molecular fingerprints and graph attention mechanism, Comput. Biol. Med. 153 (2023) 106464.

[29] H. Gao, J. Sun, Y. Wang, Y. Lu, L. Liu, Q. Zhao, J. Shuai, Predicting metabolite-disease associations based on auto-encoder and non-negative matrix factorization, Brief. Bioinform. 24 (2023) bbad259.

[30] R. Meng, S. Yin, J. Sun, H. Hu, Q.i. Zhao, scAAGA: Single cell data analysis framework using asymmetric autoencoder with gene attention, Comput. Biol. Med. 165 (2023) 107414.

[31] Y. Deng, X. Xu, Y. Qiu, J. Xia, W. Zhang, S. Liu, A multimodal deep learning framework for predicting drug–drug interaction events, Bioinformatics 36 (2020) 4316–4322.

[32] Y. Chen, T. Ma, X. Yang, J. Wang, B. Song, X. Zeng, MUFFIN: multi-scale feature fusion for drug-drug interaction prediction, Bioinformatics 2021;37:2651-2658.

[33] S.S. Deepika, T.V. Geetha, A meta-learning framework using representation learning to predict drug-drug interaction, J. Biomed. Inform. 84 (2018) 136–147.

[34] W. Zheng, H. Lin, L. Luo, Z. Zhao, Z. Li, Y. Zhang, Z. Yang, J. Wang, An attention-based effective neural model for drug-drug interactions extraction, BMC Bioinf 18 (1) (2017) 445.

[35] P. Kumar Shukla, P. Kumar Shukla, P. Sharma, P. Rawat, J. Samar, R. Moriwal, M. Kaur, Efficient prediction of drug-drug interaction using deep learning models, IET Syst. Biol. 14 (4) (2020) 211–216.

[36] A. Gottlieb, G.Y. Stein, Y. Oron, E. Ruppin, R. Sharan, INDI: a computational framework for inferring drug interactions and their associated recommendations, Mol. Syst. Biol. 8 (2012) 592.

[37] Z. Liu, X.-N. Wang, H. Yu, J.-Y. Shi, W.-M. Dong, Predict multi-type drug-drug interactions in cold start scenario, BMC Bioinf 23 (2022) 75.

[38] J. Wang, Y. Chu, J. Mao, H.-N. Jeon, H. Jin, A. Zeb, Y. Jang, K.-H. Cho, T. Song, K. T. No, De novo molecular design with deep molecular generative models for PPI inhibitors, Brief. Bioinform. 23 (2022) bbac285.

[39] X. Zhang, G. Wang, X. Meng, S. Wang, Y. Zhang, A. Rodriguez-Paton, J. Wang, X. Wang, Molormer: a lightweight self-attention-based method focused on spatial structure of molecular graph for drug-drug interactions prediction, Brief. in Bioinform. 23 (2022) bbac296.

[40] G. Lee, C. Park, J. Ahn, Novel deep learning model for more accurate prediction of drug-drug interaction effects, BMC Bioinf. (20) (2019) 415.

[41] S. Qian, S. Liang, H. Yu, A. Schlessinger, Leveraging genetic interactions for adverse drug-drug interaction prediction, PLoS Comput. Biol. 15 (5) (2019) e1007068.

[42] F. Cheng, Z. Zhao, Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties, J. Am. Med. Inform. Assoc. 21 (2014) e278–e286.

[43] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks, Bioinformatics 2018;34:i457-i466.

[44] A. Cami, S. Manzi, A. Arnold, B.Y. Reis, M.A. Medina, Pharmacointeraction network models predict unknown drug-drug interactions, PLoS One 8 (4) (2013) e61468.

[45] Y. Yu, K. Huang, C. Zhang, L. M Glass, J. Sun, C. Xiao, SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization, Bioinformatics 37 (2021) 2988–2995.

[46] R. Guimerà, M. Sales-Pardo, N.D. Price, A network inference method for large-scale unsupervised identification of novel drug-drug interactions, PLoS Comput. Biol. 9 (12) (2013) e1003374.

[47] X. Su, L. Hu, Z. You, P. Hu, B. Zhao, Attention-based Knowledge Graph Representation Learning for Predicting Drug-drug Interactions, Brief. Bioinform. 23 (2022) bbac140.

[48] Feng YH, Zhang SW, Zhang QQ et al. deepMDDI: A deep graph convolutional network framework for multi-label prediction of drug-drug interactions, Analytical Biochemistry 2022;646:114631.

[49] Shi JY, Mao KT, Yu H et al. Detecting drug communities and predicting comprehensive drug-drug interactions via balance regularized semi-nonnegative matrix factorization, J Cheminform 2019;11:28.

[50] Y. Dai, C. Guo, W. Guo, C. Eickhoff, Drug-drug interaction prediction with Wasserstein Adversarial Autoencoder-based knowledge graph embeddings, Brief. Bioinform. 22 (2020) bbaa256.

[51] R. Celebi, H. Uyar, E. Yasar, O. Gumus, O. Dikenelli, M. Dumontier, Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings, BMC Bioinformatics 20 (2019) 726.

[52] P. Zhang, F. Wang, J. Hu, R. Sorrentino, Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects, Sci. Rep. 5 (2015) 12339.

[53] Z. Yang, W. Zhong, Q. Lv, C. Yu-Chian Chen, Learning size-adaptive molecular substructures for explainable drug-drug interaction prediction by substructure-aware graph neural network, Chem. Sci. 13 (29) (2022) 8693–8703.

[54] M. Takarabe, D. Shigemizu, M. Kotera, S. Goto, M. Kanehisa, Network-based analysis and characterization of adverse drug-drug interactions, J. Chem. Inf. Model. 51 (11) (2011) 2977–2985.

[55] Asada M, Miwa M, Sasaki Y. Using drug descriptions and molecular structures for drug-drug interaction extraction from literature, Bioinformatics 2021;37:1739-1746.

[56] P. Zhang, S. Tu, N. Ben-Tal, MGAE-DC: Predicting the synergistic effects of drug combinations through multi-channel graph autoencoders, PLoS Comput. Biol. 19 (3) (2023) e1010951.

[57] P. Zhang, S. Tu, W. Zhang, L. Xu, Predicting cell line-specific synergistic drug combinations through a relational graph convolutional network with attention mechanism, Brief. Bioinform. 23 (2022) bbac403.

[58] L. Hu, J. Zhang, X. Pan, H. Yan, Z.-H. You, HiSCF: leveraging higher-order structures for clustering analysis in biological networks, Bioinformatics 37 (2021) 542–550.

[59] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, X. Li, Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data, BMC Bioinf 18 (2017) 18.

[60] Y. Deng, Y. Qiu, X. Xu, S. Liu, Z. Zhang, S. Zhu, W. Zhang, META-DDIE: predicting drug-drug interaction events with few-shot learning, Brief. Bioinform. 23 (2022) bbab514.

[61] Sridhar D, Fakhraei S, Getoor L. A probabilistic approach for collective similarity-based drug-drug interaction prediction, Bioinformatics 2016;32:3175-3182.

[62] Y.i. Zheng, H. Peng, X. Zhang, Z. Zhao, X. Gao, J. Li, DDI-PULearn: a positive-unlabeled learning method for large-scale prediction of drug-drug interactions, BMC Bioinf. 20 (S19) (2019).

[63] S. Vilar, E. Uriarte, L. Santana, N.P. Tatonetti, C. Friedman, P. Taylor, Detection of drug-drug interactions by modeling interaction profile fingerprints, PLoS One 8 (3) (2013) e58321.

[64] Y.H. Feng, S.W. Zhang, J.Y. Shi, DPDDI: a deep predictor for drug-drug interactions, BMC Bioinf. 21 (2020) 419.

[65] N. Rohani, C. Eslahchi, Drug-Drug Interaction Predicting by Neural Network Using Integrated Similarity, Sci. Rep. 9 (2019) 13645.

[66] C. Yan, G. Duan, Y. Zhang, F.-X. Wu, Y.i. Pan, J. Wang, Predicting Drug-Drug Interactions Based on Integrated Similarity and Semi-Supervised Learning, IEEE/ACM Trans. Comput. Biol. Bioinf. 19 (1) (2022) 168–179.

[67] T. Takeda, M. Hao, T. Cheng, S.H. Bryant, Y. Wang, Predicting drug-drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge, J Cheminform 9 (2017) 16.

[68] A. Kastrin, P. Ferk, B. Leskošek, J.-M. Yang, Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning, PLoS One 13 (5) (2018) e0196865.

[69] J.-Y. Shi, H. Huang, J.-X. Li, P. Lei, Y.-N. Zhang, K. Dong, S.-M. Yiu, TMFUF: a triple matrix factorization-based unified framework for predicting comprehensive drug-drug interactions of new drugs, BMC Bioinf 19 (2018) 411.

[70] AK Nyamabo, H. Yu, J.-Y. Shi, SSI-DDI: substructure-substructure interactions for drug-drug interaction prediction, Briefings in Bioinformatics 22 (2021) bbab133.

[71] J. Jiménez-Luna, F. Grisoni, G. Schneider, Drug discovery with explainable artificial intelligence, Nature Machine Intelligence 2 (2020) 573–584.

[72] K. Jin-Hwa, J. Jun, B. Zhang, Bilinear Attention Networks, in: Advances in Neural Information Processing Systems, Cornell University Library, arXiv.org, Ithaca, 2018, pp. 1571–1581.

[73] Ioffe S, Szegedy C (2015), 'Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift', International conference on machine learning, pp. 448-456.

[74] S. MacDonald, A.S. Shah, B. Tousi, Current Therapies and Drug Development Pipeline in Lewy Body Dementia: An Update, Drugs Aging 39 (7) (2022) 505–522.
[75] Z. Shen , M. Zhang , H. Zhao , S. Yi , H. Li , Efficient Attention: Attention with Linear Complexities', 2021 IEEE Winter Conference on Applications of Computer Vision (WACV) , pp. 3530 – 3538 .
[76] M.-H Guo, Z.-N. Liu, T.-J Mu, S.-M. Hu, Beyond Self-Attention: External Attention Using Two Linear Layers for Visual Tasks, IEEE Trans. Pattern Anal. Mach. Intell. 45 (2023) 5436–5447.