Bioinformatics
doi.10.1093/bioinformatics/xxxxxx
Advance Access Publication Date: Day Month Year
Original Paper

OXFORD

Data and text mining

DPSP: A multimodal deep learning framework for polypharmacy side effects prediction

Raziyeh Masumshah [1], and Changiz Eslahchi [1,2,*]

[1]Department of Computer and Data Sciences, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran and [2]School of Biological Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Polypharmacy is a significant concern in the field of pharmaceutical research. Since unanticipated Drug-Drug interactions (DDIs) can result in severe bodily harm, identifying the side effects of polypharmacy is one of the most critical tasks in human health. In recent decades, computational methods for predicting the adverse effects of polypharmacy have been developed.
Results: This article introduces DPSP, a framework for the prediction of polypharmacy side effects based on two steps; the construction of novel drug features and the application of a deep neural network to predict DDIs. In the first step, a variety of drug information, such as mono side effects, targets, enzymes, chemical substructures, and pathway information are considered. A method of feature extraction and the Jaccard similarity is used to calculate similarities between two drugs. By integrating these similarities, novel feature for each drug is generated. In the second step, the method predicts DDIs for specific DDI events using a multimodal framework and drug feature vectors. On two benchmark datasets, the performance of DPSP is evaluated by comparing its results to those of a number of well-known methods, including GNN-DDI, MSTE, MDF-SA-DDI, NNPS, DDIMDL, DNN, DeepDDI, KNN, LR, and RF. Based on distinct classification criteria, DPSP is superior to these methods. The results indicate that the use of diverse drug information and their integration is effective and efficient and assists in identifying DDI adverse effects.
Availability: The source code and datasets are available at https://github.com/raziyehmasumshah/DPSP.
Contact: ch-eslahchi@sbu.ac.ir
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Polypharmacy (the concurrent use of multiple drugs) has become a successful strategy for combating complex or co-existing diseases (Tanvir *et al*., 2022; Yao *et al*., 2022; Masumshah *et al*., 2021; Lin *et al*., 2022; Guillot *et al*., 2020; Masnoon *et al*., 2017). However, in some cases, drug combinations can cause Adverse Drug Reactions (ADRs) (i.e., side effects) (Novacek *et al*., 2020; Zitnik *et al*., 2018; Shah *et al*., 2012; Bumgardner *et al*., 2021; Lin *et al*., 2022). DDIs may alter the activity of the drugs with increasing or decreasing the effect of drugs, and in the worst case may lead to death (Tanvir *et al*., 2022; Masumshah *et al*., 2021).Therefore, the effective detection of adverse DDIs is an urgent problem in human health (Yan *et al*., 2018; LinX *et al*., 2020; Molokhia *et al*., 2017; Liu *et al*., 2017). A general source for found DDIs is through the experiments (Tari *et al*., 2010; Kim *et al*., 2022; Chen *et al*., 2016). However, identifying polypharmacy side effects in vitro and in vivo is impractical in terms of cost and time (Kim *et al*., 2022; Rohani *et al*., 2019; Feng *et al*., 2020; Bansal *et al*., 2014; Deng *et al*., 2020; Al-Rabeah *et al*., 2022). Thus, many computational models are developed to detect DDIs. The existing methods are roughly divided into two types: classification-based methods and similarity-based methods (Zitnik *et al*., 2018; Yan *et al*., 2018; Han *et al*., 2021). Classification-based methods consider DDI prediction as binary classification task (Yao *et al*., 2022; Yan *et al*., 2018; Han *et al*., 2021; cami *et al*., 2013; Cheng *et al*., 2014; Huang *et al*., 2014). In these techniques, the presence or absence of interactions are considered as positive and negative samples, respectively, and used to train classification models like KNearest-Neighbor (KNN), Logistic Regression (LR), and Random Forest (RF) (Cheng *et al*., 2014; Luo *et al*., 2021; Breiman *et al*., 2001; Denisko *et al*., 2018; Mitchell

et al., 2005; Peterson *et al*., 2009). In the recent classification-based method, Yao et al. proposed a new knowledge graph embedding method, denoted by MSTE which converts the side effects prediction task into the link prediction problem (Yao *et al*., 2022). In contrast, similarity-based methods assume that similar drugs may interact with the same drug (Zitnik *et al*., 2018; Mitchell *et al*., 2005; Peterson *et al*., 2009; Zhang *et al*., 2017). The GNN-DDI method generated a drug embedding vector and used a graph neural network to predict polypharmacy side effects prediction (Al-Rabeah *et al*., 2022). Lin proposed the MDF-SA-DDI method based on multi-source drug fusion,multi-source feature fusion, and transformer self-attention mechanism to perform latent feature fusion (Wang *et al*., 2022). Masumshah et al. presented the NNPS method based on a dimension reduction technique and used new features to predict drug-drug interactions (Masumshah *et al*., 2021). The DDIMDL method proposed by Deng et al. constructs sub-models based on diverse drug features and by adopting a deep neural network framework, combines the sub-models and feeds to fully connected neural networks (Deng *et al*., 2020). Hou et al. built a deep neural network (DNN) model and predicted DDI events by using SMILES features (Hou *et al*., 2019). Ryu et al. presented a deep learning approach, named DeepDDI, based on chemical substructures to predict crucial DDIs (Ryu *et al*., 2018). Although previous methods were effective in predicting polypharmacy side effects, but the methods like GNN-DDI, DDIMDL, and MSTE have some problems, like not being able to use data from more than one source and needing a lot of data that has been labeled. Also, NNPS only uses a small number of features and may not capture the full complexity of DDI, and the quantity and the caliber of training data in the DeepDDI method might be a restriction on its performance. Furthermore, older methods like MDF-SA-DDI, DDIMDL, and MSTE, have very complicated architectures and need a lot of computing power. So, there is a need for a more thorough and reliable method that can use more than one property of a drug and take into account the complex factors that lead to DDIs. Using five different drug features, including mono side effects, target proteins, enzymes, chemical structure, and pathways, and feeding them into a neural network architecture optimized for DDI prediction, our proposed method addresses some of these limitations. Other methods are different from ours in a number of ways, but the predictive neural network architecture is the most important difference. Our method uses a neural network architecture with only three hidden layers that are fairly simple. This simple architecture reduces overfitting and improves interpretability while maintaining high prediction accuracy. In addition, our neural network is able to rapidly learn due to the use of five distinct drug characteristics to identify similarities between drugs, which are then fed into the network. This speed of learning and prediction is especially remarkable when compared to the longer training and inference times required by more complex architectures employed by earlier methods. In this study, we develop a Deep learning framework for Polypharmacy Side effects Prediction (DPSP) based on multiple features. DPSP achieves better results in comparison with 10 well-known methods in terms of accuracy, complexity, and running time speed. The required datasets and the details of the DPSP method are described in the next section. In the Experiments and Results section, six criteria are used to compare the results of DPSP with those of other methods. Finally, the conclusion and future works are provided in the Discussion and Conclusion section.

2 Materials and Methodology

In this section, we formulate the problem and present the details of the method, including the input format and all of its steps. The overall framework of DPSP is illustrated in Figure 1.

2.1 Problem Formulation

Given a set of drugs $D = \{d_1, ..., d_m\}$ and a set of event types $E = \{e_1, ..., e_n\}$, where m and n are the total number of drugs and events, respectively. A DDI prediction task can be regarded as a probability function $f : D \times D \rightarrow E$, which predicts the probability of each type of event that will occur between the two drugs.

2.2 Datasets Description

DrugBank is a comprehensive database that provides information about 12,151 drugs, including experimental and FDA-approved drugs (Deng *et al*., 2020; Al-Rabeah *et al*., 2022; Lakizadeh *et al*., 2022; Knox *et al*., 2010). KEGG is an integrated database for biological interpretation (Ryu *et al*., 2018; Hou *et al*., 2019; Knox *et al*., 2010; Kanehisa *et al*., 2017, 2000). Pubchem consists of diverse types of substructures with values of 0 or 1 which denotes the absence or presence of a certain substructure (Kim *et al*., 2016). The OFFSIDES and Side Effect Resource (SIDER) databases were integrated to derive the side effects of individual drugs (mono side effects) (Tatonetti *et al*., 2012; Kuhn *et al*., 2016). In the current study, to verify the robustness of DPSP, we validate it on different benchmark datasets. We used the benchmark datasets that used in previous studies (Deng *et al*., 2020; Al-Rabeah *et al*., 2022; Lakizadeh *et al*., 2022; Lin *et al*., 2022). The first benchmark (DS1) contains 572 drugs with 37,264 drug-drug interactions, which are associated with 65 types of events. This benchmark includes five types of features which the four drug features, including targets, enzymes, chemical substructures, and pathways are selected from the DrugBank, KEGG, and Pubchem databases, and the mono side effects are collected from the OFFSIDES and SIDER databases. The second benchmark (DS2) contains 1258 drugs with 161,770 drug-drug interactions, which are associated with 100 types of events. This benchmark include three types of features includes targets, enzymes, and chemical substructures which are collected from the DrugBank. The representation of DDI events in both benchmark datasets, are defined as a four-tuple structure: (drug A, drug B, mechanism, action). The "mechanism" in this tuple refers to the effects of drugs in terms of the metabolism, the severity of adverse effects, the serum concentration, the therapeutic efficacy, etc. Reducing or increasing the effects are represented in the "action" term. Further details about these datasets are available in Table 1.

2.3 The DPSP Method

The architecture of DPSP is depicted in Figure 1, which is composed of two main steps: Constructing drug features and DDI prediction.

2.3.1 Constructing Drug Features

In general, the feature matrix construction section consists of feature extraction module, and aggregation module.

- **Feature extraction module**: As shown in Figure 1a, DPSP considers a set of features $F = \{f_1, ..., f_p\}$ and a feature matrix was constructed based on each feature, where p is the total number of features. In DS1 there are five feature matrices of the drugs, including the mono side effects matrix with dimension 572×9991, targets matrix with dimension 572×1162, enzymes matrix with dimension 572×202, chemical substructures matrix with dimension 572×881, and pathways matrix with dimension 572×957. In another dataset (DS2), there are three feature matrices of the drugs, containing the targets matrix with dimension 1258×1651, enzymes matrix with dimension 1258×316, and chemical substructures matrix with dimension 1258×2040. In each matrix, each row represents a drug and each column represents a feature. Two types of entries, 0 or 1, indicate the absence or presence of the corresponding feature for each

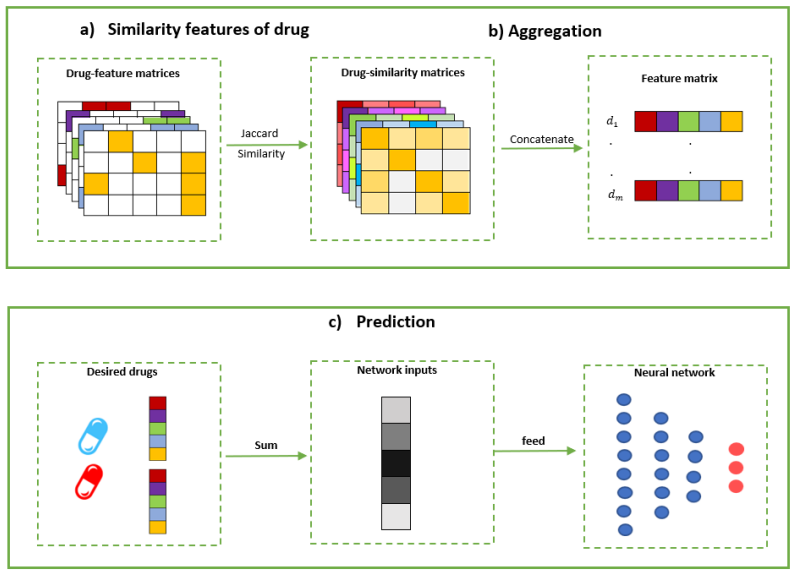

Fig. 1. The scheme of DPSP workflow. a) Constructing similarity feature matrix by using multiple information of drugs. b) Concatenating similarity matrices to obtain the feature matrix. c) Sum of the i-th and j-th rows of the drug feature matrix and feed to the neural network to predict the interaction between drug d_i and d_j.

Table 1. Details of Benchmarks.

Benchmark name	Benchmark Details	Benchmark Features	Reference
DS1	No. Drugs = 572 No.Interactions = 37,264 No.Events = 65	No. Side effects = 9991 No.Targets = 1162 No.Enzymes = 202 No.Chemical substructure = 881 No.Pathways = 957	(Deng *et al*., 2020; Al-Rabeah *et al*., 2022; Lakizadeh *et al*., 2022)
DS2	No. Drugs = 1258 No.Interactions = 161,770 No.Events = 100	No.Targets = 1651 No.Enzymes = 316 No.Chemical substructure = 2040	(Lin *et al*., 2022)

drug. For example in the case of the mono side effects matrix we have 9991 mono side effects, and so 9991 columns corresponding to these side effects. The entry correspond to the row d_i and column c_j has a value of one if drug d_i has side effect c_j and zero otherwise.
After obtaining the feature matrices, the similarity matrices for each feature based on Jaccard similarity are constructed (Figure 1a). The Jaccard similarity is defined as below:

$$J(X_{d_i}, X_{d_j}) = \frac{\mid X_{d_i} \cap X_{d_j} \mid}{\mid X_{d_i} \cup X_{d_j} \mid} = \frac{\mid X_{d_i} \cap X_{d_j} \mid}{\mid X_{d_i} \mid + \mid X_{d_j} \mid - \mid X_{d_i} \cap X_{d_j} \mid} \quad (1)$$

X_{d_i} presents, for a certain feature matrix, the set of columns whose entries in the row d_i of that feature matrix have the value 1. In equation 1, $\mid X_{d_i} \cap X_{d_j} \mid$ denote the number of common elements of X_{d_i} and X_{d_j}, and $\mid X_{d_i} \cup X_{d_j} \mid$ denote the number of union elements of X_{d_i} and X_{d_j}, respectively. Therefore, for each pair of drugs in each feature, the similarity between the two drugs will be calculated based on the above equation. So, in the DS1, we will have five similarity matrices of size 572×572, and in the DS2, we will have three matrices of size 1258×1258.

- **Aggregation module**: At this step, by concatenating the similarity matrices, the main drug feature matrix has calculated (Figure 1b). Therefore, the drug feature matrix in DS1 will be of dimension 572×2860, and in DS2 will be of dimension 1258×3774. The rows of the resulting matrix represent the IDs of the drugs, while the columns contain information about the features.

2.3.2 DDI prediction

In the Constructing Drug Features section, a unique feature vector for each drug was constructed. The purpose of this section is to express the architecture of the model, which is a deep neural network, and finally, predict the DDIs events. A deep neural network is an Artificial Neural Network (ANN) with an input layer, multiple hidden layers, and an output layer (Schmidhuber , 2015; Emmert-Streib *et al*., 2020; Nair *et al*., 2010). Parameters can affect the performance of our DPSP model. Therefore, we investigate the number of hidden layers, the number of neurons, the

Table 2. Results of comparison of the DPSP with some of the machine learning methods on DS1.

Method	ACC	AUPR	AUROC	F_ score	Precision	Recall
DPSP	**0.9344**	**0.9773**	**0.9990**	**0.8796**	**0.9103**	**0.8517**
GNN-DDI	0.9180	0.9709	0.9985	0.8440	0.9000	0.8165
MSTE	0.8584	0.9343	0.9981	0.7420	0.8471	0.6979
MDF-SA-DDI	0.9121	0.9657	0.9989	0.8659	0.8833	0.8501
NNPS	0.9113	0.9699	0.9989	0.8524	0.8816	0.8420
DDIMDL	0.8757	0.9353	0.9977	0.7624	0.8278	0.7287
DNN	0.8797	0.9134	0.9963	0.7223	0.8047	0.7027
DeepDDI	0.8371	0.8899	0.9961	0.6848	0.7275	0.6611
KNN	0.7201	0.7854	0.9821	0.5323	0.7334	0.4600
LR	0.7315	0.7988	0.9939	0.3372	0.5341	0.2836
RF	0.7756	0.8515	0.9955	0.5162	0.7215	0.4394

Note: Bold numbers show the best performance for each criterion

activation function, and the dropout rate based on the following steps to achieve the best performance.

- The number of hidden layers: $\{1, 2, 3\}$
- The number of neurons in hidden layers: $\{128, 256, 512\}$
- Activation functions: {Rectified Linear Unit (ReLU), Sigmoid, and Softmax}
- The dropout rate: $\{0.1, 0.2, 0.3, 0.4, 0.5\}$

We trained several networks based on these parameters and represented them in Tables S1-S4 in the Additional file. As shown in these tables, three hidden layers with 512, 256, and 128 neurons, respectively with a dropout rate of 0.3, utilizing the ReLU (Srivastava *et al*., 2014; Nair *et al*., 2010; Nwankpa *et al*., 2018) as the activation function for the first hidden layer, and Sigmoid activation function for other hidden layers, and the Softmax for output layer has the best results (Kouretas *et al*., 2019; Narayan , 1997). Now for a given drug pair (d_i, d_j), i-th and j-th rows of the drug feature matrix are summed and fed to the considered fully connected neural network (Figure 1c). Then, we adopt binary-cross-entropy (Zhang2018 *et al*., 2018) as the loss function and the training iteration (epoch) set to 100 with a batch size of 128. Adam optimizer (Kingma *et al*., 2014) was used as the optimization algorithm to train the network. Finally, the network returns the probability of occurrence of each DDIs event. The pseudocode of DPSP is shown in Algorithm 1.

Algorithm 1 The pseudocode of DPSP algorithm

Require: All features of drugs $F = \{f_1, ..., f_p\}$ including mono side effects, targets, enzymes, chemical substructure, and pathways;

Ensure: Prediction results.

1: **for** all $i \in m$ **do**
2: Calculate Jaccard Similarity of drug features to measure drug similarity;
3: Concatenate drug similarities;
4: Put the sum of i-th and j-th feature vectors as the input of the deep neural network model;
5: The model outputs the probability of occurrence of each DDI's event;
6: The event with maximum probability will be selected;

3 Experiments and Results

There are two main tasks in DDI prediction, first is identifying the absence or presence the interaction between the drugs. The second is determining the type of interaction among the drugs. In this paper, we deal with the second case, and split the drug pairs into training and test sets, and employ 5-fold cross-validation to evaluate the DDI prediction task. In each fold, one subset is considered as the test set and the rest as training. For validating the performance of DPSP, we investigated it on two datasets (DS1 and DS2). We compared our method with 10 well-known methods GNN-DDI, MSTE, MDF-SA-DDI, NNPS, DDIMDL, DNN, DeepDDI, KNN, LR, and RF. We adopt the 5-fold cross-validation for 100 iterations and use the average of the results to assess the performance of all methods. The average of accuracy (ACC), area under the precision-recall curve (AUPR), area under the receiver-operating characteristic curve (AUROC), F_ score, Precision, and Recall values of all methods on DS1 and DS2 is shown in Tables 2 and S5, respectively. According to Tables 2 and S5, DPSP achieves better results than other well-known methods in terms of all criteria. DPSP outperforms both the similarity-based methods and the classification-based methods. Improvement in performance in each metric score of our method DPSP with respect to the highest score of the baseline methods are equal to 1.64% in ACC, 0.64% in AUPR, 0.05% in AUROC, 3.56% in F_ score, 1.03% in Precision, 3.52% in Recall in DS1, and 0.16% in ACC, 0.14% in AUPR, 0.02% in AUROC, 0.04% in F_ score, 0.16% in Precision, and 0.43% in Recall in DS2. The difference between DPSP performance with other methods in terms of F_ score, Precision, and Recall is outstanding. To compare the results of DPSP more precisely, we compare it to the results of the best method between all methods (GNN-DDI) with more details. Figures S1-S2 in Additional file, illustrate the boxplots of the F_ score, Precision, and Recall criteria in all events by DPSP and GNN-DDI methods on DS1 and DS2 in 100 iterations, respectively. As shown in Fig S1 and S2, the range of variation of the F_ score, Precision, and Recall criteria in both datasets for DPSP method are less than the range of variation of the F_ score, Precision, and Recall criteria for the GNN-DDI method which is the evidence of good performance and robustness of DPSP.

3.1 Feature Evaluation

In this section, we do two different evaluations based on the feature sets. In the first step of, we assessed the importance of the features utilized in our drug-drug interaction prediction model. To achieve this, the results of the DPSP method are compared for DS1 after removing one of the features including mono side effect, target, enzyme, chemical substructure, and pathway. The investigation's findings are presented in Table 3. All of these features were discovered to be crucial for achieving

Table 3. This table demonstrates the significance of DS1 data set features by removing one feature and evaluating the results of the DPSP method using the remaining features based on all evaluation criteria.

Not consider	ACC	AUPR	AUROC	F_ score	Precision	Recall
Mono Side Effect	0.9283	0.9744	0.9989	0.8547	0.8957	0.8339
Target	0.9319	0.9752	0.9985	0.8435	0.8778	0.8272
Enzyme	0.9244	0.9727	0.9978	0.8443	0.8800	0.8323
Chemical Substructure	0.9200	0.9767	0.9982	0.8573	0.8742	0.8516
Pathway	0.8070	0.8906	0.9969	0.6443	0.7582	0.7027

Table 4. The results of dimension reduction technique in DS1.

PCA method	ACC	AUPR	AUROC	F_ score	Precision	Recall
scenario 1	0.9264	0.9734	0.9989	0.8493	0.8919	0.8332
scenario 2	0.8942	0.9573	0.9987	0.7925	0.8279	0.7782

high model performance. When each feature was eliminated and the DPSP method was re-executed with the four remaining features, we observed a significant decrease in model performance across all evaluation criteria. In addition, our results demonstrated that the pathway-defining feature played an important role in our model. Therefore, we execute the DPSP method utilizing only the pathway feature and present the results in Table S6. The results indicate that pathway information on its own was insufficient to achieve high performance. We have done similar experiments for DS2. Again, it is evident that all three characteristics of this data set, targets, enzymes, and chemical substructures are essential for achieving high model performance. When each feature was removed and the DPSP method was re-executed with the remaining two features, we observed a significant decrease in model performance across all evaluation criteria (see Table S7). Moreover, our results demonstrated that the chemical substructure property played an important role in our model. Consequently, we execute the DPSP method using only the chemical substructure and present the results in Table S8. The results indicate that chemical substructure data alone was insufficient for achieving high performance. The second criterion for evaluation is based on the size of the features. We are employing dimension reduction techniques to reduce the size of the feature matrix due to its large size. To accomplish this, we use the principal components analysis (PCA) to reduce the size of the DPSP algorithm's input feature matrix. In this section, we examined two distinct scenarios: 1) applying the PCA method to each of the similarity matrices and then concatenating them, and 2) concatenating the similarity matrices first and then applying the PCA method to the concatenated matrix (the one that we directly used in method). In both instances, the minimum number of principal components is selected to retain 95% of variance. In the first case, and for the DS1 data set, the size of each Jaccard similarity matrix mono side effects, targets, enzymes, chemical substructures, and pathways, is reduced to 572×481, 572×186, 572×12, 572×11, and 572×173, respectively and in the second case, the feature matrix with dimension 572×2860 is reduced to 572×107. Observing the results in Table 4 reveals that information is lost in both instances when the dimension size of feature matrices is decreased. On DS2, we employ similar scenarios. The results of these cases are displayed in Table S9 of the Additional File, which demonstrates that we obtained similar results in this data set as in the previous data set and that the dimension reduction technique causes information loss. Therefore, not only do we need all of the features, but employing the dimension reduction technique causes us to lose the efficiency of the method.

4 Discussion and Conclusion

Due to the side effects caused by the simultaneous use of medicinal compounds, it is impossible to screen all pairs of drugs in terms of time and cost. So, computational methods are developed to predict DDI events. In this paper, we obtain the DDI data from the DrugBank, KEGG, PubChem, SIDER, and OFFSIDES, and consider two types of datasets (DS1 and DS2) which classifies DDI-associated events into 65 and 100 events, respectively. For the prediction of polypharmacy side effects, we utilize DPSP, a deep neural network framework that employs multiple drug features. On both DS1 and DS2 data sets, the model achieves excellent performance across all criteria. Table 5 of Additional file1 displays the execution time of each method on the DS1 and DS2 data sets. This table demonstrates that the DPSP method has the shortest execution time, 20 minutes on DS1 and 2 hours on DS2. These execution times demonstrate the superior performance of this method compared to other methods. Therefore, we conclude that the method can produce extremely reliable results in a very precise amount of time. Several reasons account for the high performance of DPSP:

- It gets diverse information from different aspects of drugs and takes advantage of different types of similarity matrices that yield an inclusive insight into them.
- DPSP uses the summation operator to aggregate the feature vectors of two drugs into one vector for representing the drug-drug pairs in neural network.
- It utilizes a deep neural network that can extract high-level features from inputs for better prediction and also is easy to implement.

In both the DS1 and DS2 data sets, five events occur most frequently. To further demonstrate DPSP's ability to predict unknown DDIs, we analyzed DPSP's predictions for each of these events separately. The results are shown in Tables 6 and S10. In these tables, the six utilized criteria, including F_ score, Precision, and Recall, demonstrate the effectiveness of DPSP in analyzing events from both datasets. For the final evaluation of the presented method, the five False-Positive drug pairs for which the DPSP method gave the highest probability on DS1 and DS2 were considered in Table 7. It should be noted that the version of the Drug Bank used in this study was 5.1.9, but these interactions were added to the most recent version of the Drug Bank, which validated the accuracy of these 10 predictions and the performance of the method. As a result, we propose that the method's other False Positive with a high probability be considered for further experimental investigations.

According to the findings of this study, using a combination of drug features, a rigorous aggregation schema, and a simple neural network

Table 5. Comparison results between the execution time of the DPSP method and other machine learning methods on both datasets.

Dataset	DPSP	GNN-DDI	MSTE	MDF-SA-DDI	NNPS	DDIMDL	DNN	DeepDDI	KNN	LR	RF
DS1	**20 min**	90 min	1440 min	120 min	45 min	70 min	40 min	60 min	31 min	40 min	50 min
DS2	**120 min**	240 min	4320 min	276 min	141 min	289 min	110 min	140 min	153 min	205 min	260 min

Table 6. This table displays the performance of the DPSP method on the five most common drug combinations in DS1 according to all evaluation criteria.

Rank	Event Name	ACC	AUPR	AUROC	F_ score	Precision	Recall
1	Decreasing the metabolism	0.9680	0.9466	0.9617	0.9398	0.9312	0.9485
2	Increasing the risk or severity of adverse effects	0.9792	0.9644	0.9729	0.9593	0.9586	0.9600
3	Increasing the serum concentration	0.9718	0.9140	0.9422	0.9064	0.9131	0.8997
4	Decreasing the serum concentration	0.9894	0.9198	0.9475	0.9163	0.9338	0.8994
5	Decreasin the therapeutic efficacy	0.9957	0.9407	0.9577	0.9387	0.9616	0.9169

Table 7. This table displays, for each of the five most frequent events on DS1 and DS2, the new interactions predicted by the DPSP method with the highest probabilities.

	DS1		DS2	
Rank	Drug Name1	Drug Name2	Drug Name1	Drug Name2
1	Amiodarone	Glimepiride	Capsaicin	Fluvoxamine
2	Amiloride	Nabilone	Cyanocobalamin	Aclidinium
3	Agomelatine	Fluvoxamine	Betamethasonephosphate	Cyclosporine
4	Buspirone	Dabrafenib	Tacrolimus	Vitamine
5	Alogliptin	Etacrynicacid	Ranolazine	Crizotinib

architecture can be an effective method for predicting polypharmacy side effects. For possible future work, we propose utilizing additional characteristics, such as off-label adverse effects and transporter features, as well as various similarity measurements, such as Dice similarity and Cosine similarity.

Acknowledgements

The authors thank the anonymous reviewers for their valuable suggestions.

References

Tanvir, F., Saifuddin, K.M. and Akbas, E (2022) DDI Prediction via Heterogeneous Graph Attention Networks, *arXiv preprint arXiv:2207.05672*

Yao, J., Sun, W., Jian, Z., Wu, Q. and Wang, X (2022) Effective knowledge graph embeddings based on multidirectional semantics relations for polypharmacy side effects prediction, *Bioinformatics*, 38(8), pp.2315-2322.

Masumshah, R., Aghdam, R. and Eslahchi, C (2021) A neural network-based method for polypharmacy side effects prediction, *BMC bioinformatics*, 22(1), pp.1-17, 2021.

Lin, S., Zhang, G., Wei, D.Q. and Xiong, Y (2022) DeepPSE: Prediction of polypharmacy side effects by fusing deep representation of drug pairs and attention mechanism, *Computers in Biology and Medicine*, 149, p.105984.

Guillot, J., Maumus-Robert, S. and Bezin, J (2020) Polypharmacy: a general review of definitions, descriptions and determinants, *Therapies*, 75(5), pp.407-416.

Masnoon, N., Shakib, S., Kalisch-Ellett, L. and Caughey, G.E (2017) What is polypharmacy? A systematic review of definitions, *BMC geriatrics*, 17(1), pp.1-10.

Nováek, V. and Mohamed, S.K (2020) Predicting polypharmacy side-effects using knowledge graph embeddings, *AMIA Summits on Translational Science Proceedings*, p.449.

Zitnik, M., Agrawal, M. and Leskovec, J (2018) Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics*, 34(13), pp.i457-i466.

Shah, B.M. and Hajjar, E.R (2012) Polypharmacy, adverse drug reactions, and geriatric syndromes, *Clinics in geriatric medicine*, 28(2), pp.173-186.

B. Bumgardner, F. Tanvir, K. M. Saifuddin and E. Akbas (2021) Drug-Drug Interaction Prediction: a Purely SMILES Based Approach, *2021 IEEE International Conference on Big Data (Big Data)*, pp. 5571-5579.

Lin, S., Chen, W., Chen, G., Zhou, S., Wei, D.Q. and Xiong, Y (2022) MDDI-SCL: predicting multi-type drug-drug interactions via supervised contrastive learning, *Journal of Cheminformatics*, 14(1), pp.1-12.

Yan, Z., Zhao, L., Wei, X. and Zhang, Q (2018) mproved Label Propagation Model to Predict Drug-drug Interactions, *MATEC Web of Conferences*, 173, p. 02007.

Lin, X., Quan, Z., Wang, Z.J., Ma, T. and Zeng, X (2020) KGNN: Knowledge Graph Neural Network for Drug-Drug Interaction Prediction, *IJCAI*, 380, pp. 2739-2745.

Molokhia, M. and Majeed, A (2017) Current and future perspectives on the management of polypharmacy, *BMC Family Practice*, 18(1), pp.1-9.

Liu, R., AbdulHameed, M.D.M., Kumar, K., Yu, X., Wallqvist, A. and Reifman, J (2017) Data-driven prediction of adverse drug reactions induced by drug-drug interactions, *BMC Pharmacology and Toxicology*, 18, pp.1-18.

Tari, L., Anwar, S., Liang, S., Cai, J. and Baral, C (2010) Discovering drugdrug interactions: a text-mining and reasoning approach based on properties of drug metabolism, *Bioinformatics*, 26(18), pp.i547-i553.

Kim, E. and Nam, H (2022) DeSIDE-DDI: interpretable prediction of drug-drug interactions using drug-induced gene expressions, *Journal of cheminformatics*, 14(1), pp.1-12.

Chen, C.M., Kuo, L.N., Cheng, K.J., Shen, W.C., Bai, K.J., Wang, C.C., Chiang, Y.C. and Chen, H.Y (2016) The effect of medication therapy management service combined with a national PharmaCloud system for polypharmacy patients, *Computer Methods and Programs in Biomedicine*, 134, pp.109-119.

Rohani, N. and Eslahchi, C (2019) Drug-drug interaction predicting by neural network using integrated similarity, *Scientific reports*, 9(1), pp.1-11.

Feng, Y.H., Zhang, S.W. and Shi, J.Y (2020) DPDDI: a deep predictor for drug-drug interactions, *BMC bioinformatics*, 21(1), pp.1-15.

Bansal, M., Yang, J., Karan, C., Menden, M.P., Costello, J.C., Tang, H., Xiao, G., Li, Y., Allen, J., Zhong, R. and Chen, B (2014) A community computational challenge to predict the activity of pairs of compounds, *Nature biotechnology*, 32(12), pp.1213-1222.

Deng, Y., Xu, X., Qiu, Y., Xia, J., Zhang, W. and Liu, S (2020) A multimodal deep learning framework for predicting drugdrug interaction events, *Bioinformatics*, 12(1), pp.1-13.

Al-Rabeah, M.H. and Lakizadeh, A (2022) Prediction of drug-drug interaction events using graph neural networks based feature extraction, *Scientific Reports*, 134, pp.109-119.

Lakizadeh, A. and Babaei, M (2022) Detection of polypharmacy side effects by integrating multiple data sources and convolutional neural networks, *Molecular Diversity*, pp.1-11.

Han, K., Cao, P., Wang, Y., Xie, F., Ma, J., Yu, M., Wang, J., Xu, Y., Zhang, Y. and Wan, J (2021) A Review of Approaches for Predicting Drug-Drug Interactions Based on Machine Learning, *Frontiers in Pharmacology*, 12, pp.814858-814858.

Cami, A., Manzi, S., Arnold, A. and Reis, B.Y (2013) Pharmacointeraction network models predict unknown drug-drug interactions, *PloS one*, 8(4), p.e61468.

Cheng, F. and Zhao, Z (2014) Machine learning-based prediction of drugdrug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties, *Journal of the American Medical Informatics Association*, 21(e2), pp.e278-e286.

Huang, H., Zhang, P., Qu, X.A., Sanseau, P. and Yang, L (2014) Systematic prediction of drug combinations based on clinical side-effects, *Scientific reports*, 4(1), pp.1-7.

Luo, H., Li, M., Yang, M., Wu, F.X., Li, Y. and Wang, J (2021) Biomedical data and computational models for drug repositioning: a comprehensive review, *Briefings in bioinformatics*, 22(2), pp.1604-1619.

Breiman, L (2001) Random forests, *Mach. learning*, 45, 532.

Denisko, D. and Hoffman, M.M (2018) Classification and interaction in random forests, *Proceedings of the National Academy of Sciences*, 15(8), pp.1690-1692.

Mitchell, T. M (2005) Logistic regression, *Mach. learning*, 10, 701.

Peterson, L. E (2009) K-nearest neighbor, *Scholarpedia*, 4, 1883.

Zhang, W., Chen, Y., Liu, F., Luo, F., Tian, G. and Li, X (2017) Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data, *BMC bioinformatics*, 18(1), pp.1-12.

Lin, S., Wang, Y., Zhang, L., Chu, Y., Liu, Y., Fang, Y., Jiang, M., Wang, Q., Zhao, B., Xiong, Y. and Wei, D.Q (2022) MDF-SA-DDI: predicting drugdrug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism, *Briefings in Bioinformatics*, 23(1), p.bbab421.

Ryu, J.Y., Kim, H.U. and Lee, S.Y (2018) Deep learning improves prediction of drugdrug and drugfood interactions, *Proceedings of the National Academy of Sciences*, 115(18), pp.E4304-E4311.

Hou, X., You, J. and Hu, P (2019) Predicting drug-drug interactions using deep neural network, *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, pp. 168-172.

Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V. and Djoumbou, Y (2010) DrugBank 3.0: a comprehensive resource for omics research on drugs, *Nucleic acids research*, 39, pp.D1035-D1041.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs, *Nucleic acids research*, 45(D1), pp.D353-D361.

Kanehisa, M. and Goto, S (2000) KEGG: kyoto encyclopedia of genes and genomes, *Nucleic acids research*, 28(1), pp.27-30.

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A. and Wang, J (2016) PubChem substance and compound databases, *Nucleic acids research*, 44(D1), pp.D1202-D1213.

Tatonetti, N.P., Ye, P.P., Daneshjou, R. and Altman, R.B (2012) Data-driven prediction of drug effects and interactions, *Science translational medicine*, 4(125), pp.125ra31-125ra31.

Kuhn, M., Letunic, I., Jensen, L.J. and Bork, P (2016) The SIDER database of drugs and side effects, *Nucleic acids research*, 44(D1), pp.D1075-D1079.

Schmidhuber, J (2015) Deep learning in neural networks: An overview, *Neural networks*, 61, pp.85-117.

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S. and Dehmer, M (2020) An introductory review of deep learning for prediction models with big data, *Frontiers in Artificial Intelligence*, 3, p.4.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R (2014) Dropout: a simple way to prevent neural networks from overfitting, *The journal of machine learning research*, 15(1), pp.1929-1958.

Nair, V. and Hinton, G.E (2010) Rectified linear units improve restricted boltzmann machines, *Icml*.

Nwankpa, C., Ijomah, W., Gachagan, A. and Marshall, S (2018) Activation functions: Comparison of trends in practice and research for deep learning, *arXiv preprint arXiv:1811.03378*.

Narayan, S (1997) The generalized sigmoid activation function: Competitive supervised learning, *Information sciences*, 99(1-2), pp.69-82.

Kouretas, I. and Paliouras, V (2019) Simplified hardware implementation of the softmax activation function, *2019 8th international conference on modern circuits and systems technologies (MOCAST)*, pp. 1-4.

Zhang, Z. and Sabuncu, M (2018) Generalized cross entropy loss for training deep neural networks with noisy labels, *Advances in neural information processing systems*, 31.

Kingma, D.P. and Ba, J (2014) Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*.