

A PU-Learning-Based Framework for Polypharmacy Adverse Effect Detection Using Feature Extraction and Multi-Layer Perceptron Models

Krzysztof Wiśniewski^{1*}, Małgorzata Nowak¹, Tomasz Adamczyk²

¹ Department of Clinical Pharmacology and ML-based Pharmacovigilance, Faculty of Medicine, Warsaw University of Life Sciences, Warsaw, Poland.

² Department of Polypharmacy and Adverse Drug Reaction Prediction, Faculty of Medicine, Jagiellonian University, Krakow, Poland.

*E-mail ✉ krzysztof.wisniewski@sggw.edu.pl

Abstract

Polypharmacy—the simultaneous administration of several drugs—heightens the likelihood of harmful reactions caused by interactions between medications. With the rising occurrence of polypharmacy, accurately predicting these interactions has become crucial in drug development and healthcare. Because clinical trials struggle to identify uncommon side effects linked to multiple drug use, researchers are creating computational models to forecast such adverse outcomes. This paper presents PU-MLP, a Multi-Layer Perceptron-driven framework for predicting the side effects of drug combinations. The study applies state-of-the-art machine learning approaches to uncover associations between pharmaceuticals and their adverse reactions. The framework operates in three main phases: first, it generates high-quality representations for individual drugs by combining a random forest classifier, Graph Neural Networks (GNNs), and dimensionality reduction techniques. Next, it uses Positive Unlabeled learning to handle data uncertainty. Finally, a Multi-Layer Perceptron forecasts polypharmacy side effects. Evaluation via 5-fold cross-validation demonstrates that the proposed method outperforms existing techniques, achieving AUPR, AUC, and F1 scores of 0.99, 0.99, and 0.98, respectively.

Keywords: Positive-unlabeled learning, Polypharmacy side effect, Multi-layer perceptron, Drug-drug interaction

Introduction

Polypharmacy refers to the concurrent intake of multiple medications and has grown more widespread, especially among older adults and patients with multifaceted medical needs [1]. As the aging population expands, complex treatment regimens involving several drugs are becoming increasingly necessary for effective disease management [2, 3]. Data from a 2010–2011 survey indicated that 67% of elderly Americans were taking five or more medications [4, 5]. While combining drugs can improve treatment efficacy [6], it also raises serious

concerns regarding adverse side effects resulting from drug-drug interactions [7, 8]. From a scientific standpoint, comprehensively testing every possible drug-drug interaction is difficult because such events are relatively rare and hard to detect [9, 10]. Although clinical trials may catch some side effects, these studies are typically costly and time-consuming [11]. Adverse drug reactions remain a leading cause of emergency medical visits [12, 13].

To overcome these obstacles, researchers have developed computational techniques to detect drug-drug interactions. These methods generally belong to two categories. The first category determines whether interactions exist without specifying their type, drawing on experimental data and medical records [14, 15]. The second category aims to pinpoint the particular side effects that may result from specific drug pairs [16, 17]. Approaches in the first group often rely on potential drug combinations and techniques such as cell viability assays

Access this article online

<https://smerpub.com/>

Received: 28 March 2025; Accepted: 14 June 2025

Copyright CC BY-NC-SA 4.0

How to cite this article: Wiśniewski K, Nowak M, Adamczyk T. A PU-Learning-Based Framework for Polypharmacy Adverse Effect Detection Using Feature Extraction and Multi-Layer Perceptron Models. J Med Sci Interdiscip Res. 2025;5(1):189-203. <https://doi.org/10.51847/xS3nx9UXDp>

[18, 19]. For instance, DEDICOM applies tensor decomposition [20], while Decagon uses a multimodal interaction graph [11]. KBLRN predicts various drug side effects via graph embedding methods [21], and TriVec uses tensor factorization for link prediction within a 3D tensor that represents entities and relations [22]. TIP adopts a Decagon-inspired strategy using three subgraphs (protein-protein, protein-drug, and drug-drug) to build a heterogeneous graph [23]. PSECNN forecasts polypharmacy side effects through a similarity-based technique that derives five feature vectors per drug using the Jaccard index [24]. DANN-DDI integrates multiple drug feature sources via a deep neural network to learn representations of drug pairs and predict unknown interactions [25]. CNN-DDI incorporates features related to drug categories, targets, enzymes, and pathways, using Jaccard similarity within a deep convolutional neural network to predict drug-drug interactions [26]. GNN-DDI constructs an attributed heterogeneous network from diverse data sources and uses a deep multimodal framework to predict drug-drug interactions and associated events [27]. DeepPSE is an end-to-end deep learning model for polypharmacy side-effect prediction that includes a drug-pair representation module and a transformer encoder for latent feature extraction [28]. SimVec employs a three-stage learning procedure that iteratively refines side-effect embeddings along edges [29]. DeSIDE-DDI offers a deep learning architecture for drug-drug interaction prediction based on gene expression profiles, improving interpretability through a gating mechanism that models combined drug effects [30]. VGAE reports results on three multimodal networks involving drug and protein nodes, DrugBank data, and drug-cell nodes [31]. DRSE predicts drug-disease associations by fusing multiple data sources while incorporating drug side effect information [32]. LHGCE introduces an end-to-end heterogeneous graph convolutional encoder for forecasting potential drug-disease links [33].

The NNPS method creates distinctive feature vectors from mono-drug side effects and drug-protein interaction (DPI) data to boost prediction performance. It generates two feature matrices—one for mono-drug side effects and another for DPIs. To manage the high dimensionality, Principal Component Analysis (PCA) is applied, preserving 95% of the variance and yielding reduced matrices F1 (645×503) and F2 (645×22). These are concatenated into a unified drug feature matrix (645×525). The model uses a feedforward neural

network with an input layer, three hidden layers, and an output layer. It calculates the probability (π) for each side effect and applies a threshold (θ_i) between 0 and 1 to decide whether the side effect is likely to occur for a given drug pair [34].

The DPSP method first applies Jaccard similarity to evaluate drug relationships for detecting drug-drug interactions (DDIs), followed by a multimodal framework to predict specific DDI events [35]. GADNN, which relies on attention-driven Graph Neural Networks, forecasts DDIs by assessing the importance of different drug-related features [36]. A common challenge in these tasks is the limited availability of confirmed positive examples, which leads most data points to be treated as negative and creates labeling uncertainty. Positive Unlabeled (PU) learning addresses this by making optimal use of available data to improve model reliability [37]. PU learning includes three primary strategies: (1) the direct method, which performs classification using only positive and unlabeled samples without needing confirmed negatives; (2) the two-step method, which first identifies reliable negative samples from unlabeled data and then iteratively classifies all instances; and (3) bagging-based PU methods, subdivided into inductive and transfer learning variants. Liu *et al.* [38] introduced a direct approach using biased SVM, while Elkan and Noto [39] developed the LPU method for learning from positive and unlabeled data. Two-step techniques iteratively refine negative sample selection, and bagging-based methods further differentiate into inductive and transfer categories [40]. These PU learning strategies have demonstrated success in bioinformatics tasks such as drug-target interaction prediction and disease gene identification [41, 42].

The PU-GNN method, built on graph neural networks, predicts drug side effects through three core stages: (1) feature extraction via a new biclustering algorithm, (2) uncertainty reduction using positive-unlabeled learning, and (3) prediction of polypharmacy effects with a graph neural network [43]. The Msff-ma-ddi method focuses on multi-source feature integration and attention mechanisms for predicting drug interaction events, though it does not specifically address interactions within defined subgroups [44]. PPAEDTI presents a self-adjusting model for drug-target interaction prediction that performs well but has limited predictive power [45]. The approach by He *et al.* [46] merges graph transformers and GCNs for drug-disease relationship prediction and supports in-depth analysis. In contrast, the

proposed PU-MLP method focuses specifically on adverse-effect prediction. By incorporating graph neural networks and clustering techniques, it achieves a more streamlined and effective architecture for identifying harmful drug interactions.

According to the reviewed studies, forecasting adverse reactions from the simultaneous administration of multiple drugs involves several notable difficulties. The primary issues are: (1) selecting the most suitable features for each medication that truly capture its relevant properties, (2) managing uncertainty within the input data stemming from many undocumented interactions that are frequently presumed insignificant due to insufficient experimental validation, which in turn lowers overall predictive performance, and (3) even with numerous existing techniques, creating a robust and effective solution for polypharmacy side effect prediction continues to be a major hurdle in computational pharmacology.

To overcome these obstacles, we introduce PU-MLP, a multi-class, multi-label framework grounded in a Multi-Layer Perceptron (MLP) architecture. The PU-MLP approach consists of three primary stages.

In the initial phase, the method derives high-quality features for each drug that faithfully represent its characteristics. This begins with a Random Forest Classifier to identify the most informative drug attributes associated with different side effects, producing a compact feature vector that encapsulates critical information.

In addition to random forests, this work incorporates several biclustering techniques, including Similar Patterns of Target Proteins (SPP) [43], LAS (Large Average Submatrices) [47], Canonical Correlation Analysis (CCA) [48], and Principal Component Analysis (PCA) [49].

SPP biclustering is particularly effective at handling the complexity of large-scale datasets. It proceeds through similarity computation, structural decomposition, cluster identification, and feature extraction, which collectively simplify the data and reveal meaningful groupings. This technique enhances data quality by minimizing noise, improves the performance of subsequent machine learning models, and produces more interpretable results that support pattern discovery in polypharmacy side-effect analysis [43].

In contrast to conventional clustering approaches, the LAS biclustering method simultaneously groups data along two dimensions (e.g., features and samples) and

focuses on coherent submatrices. It applies locality-sensitive criteria to detect correlations between these dimensions. A notable strength is its dynamic adaptability to the underlying data structure, leading to more precise cluster detection within intricate patterns. This dual-perspective analysis enables deeper insights into areas such as biotechnology and complex data interpretation [47].

Canonical Correlation Analysis (CCA) is a sophisticated statistical method that identifies shared underlying patterns between two sets of variables. It begins with preprocessing to emphasize important features, followed by data segmentation into clusters that expose hidden relationships. CCA is useful across numerous domains, including data science, psychology, and the social sciences, as its outcomes support better decision-making and deeper understanding. The resulting clusters serve as valuable inputs for downstream predictive tasks or pattern recognition [48].

Each biclustering algorithm generates 'n' biclusters. These clusters are organized as columns within a feature matrix, enabling clearer visualization and categorization of the data. For every drug, a binary feature vector of length 'n' is constructed, where a value of '1' indicates membership in a specific bicluster and '0' indicates otherwise. This binary encoding enables straightforward comparison of drugs based on their cluster affiliations. The resulting matrix provides a solid basis for machine learning algorithms and additional downstream analyses. Overall, these complementary strategies not only improve clustering accuracy but also streamline subsequent data processing and decision-making, leading to better pattern recognition and more reliable analytical outcomes.

Following feature extraction, the study models the relational structure of drug-side effect networks using Graph Neural Networks (GNNs) [43]. Uncertainty present in the input data is further mitigated by applying a Positive Unlabeled learning algorithm. Finally, the proposed method employs a multi-layer perceptron to predict polypharmacy-related side effects.

Compared with the Msff-ma-ddi method [44], PU-MLP places greater emphasis on capturing independent drug relationships through dimensionality reduction and clustering, thereby improving the precision of polypharmacy adverse-effect predictions. Compared with the PPAEDTI method [45], PU-MLP directly addresses data uncertainty via positive-unlabeled

learning, yielding higher accuracy in forecasting drug-related side effects.

Materials and Methods

Dataset

This research frames polypharmacy side effect prediction as a multi-class, multi-label classification task. We employed established benchmark datasets cited in prior works [24, 27]. The TWOSIDES dataset (denoted as DS) contains two main components. The first provides drug characteristics used to build feature vectors for individual drugs. The second includes drug-drug pairs along with their associated side effects. The dataset integrates protein-protein interaction networks, side effect profiles, and drug-target interactions from multiple sources to form a comprehensive Knowledge Graph (KG). This KG

consists of drug and protein nodes connected by edges representing protein-protein, drug-protein, and drug-drug interactions. Each drug-drug interaction is annotated with a specific side-effect type. TWOSIDES is a widely recognized benchmark in side-effect studies and was used in this research. After preprocessing, the final dataset comprises 645 drugs.

The OFFSIDES dataset supplies additional drug combinations and their corresponding side effects [24]. We also incorporated the bio-targets dataset, which includes roughly 131,000 interactions involving the same 645 drugs and 3,648 proteins. This benchmark encompasses four categories of drug features: targets, enzymes, SMILES strings, and pathways, obtained from DrugBank and KEGG. **Table 1** summarizes the detailed characteristics of the datasets used.

Table 1. The details of the datasets used.

Details	DS1 dataset	DS2 dataset
Drugs	645	572
SideEffect	963	47
DrugDrugInteractions	63473	37264
DrugDrugSideEffect	4649441	37264
Targets	1167	1162
Pathway	959	957
SMILES	592	881
Enzyme	206	202

Methods

This section outlines the three primary difficulties encountered in predicting polypharmacy side effects. To tackle these issues, we introduce PU-MLP, a multi-class, multi-label approach built upon a Multi-Layer Perceptron. The framework is structured into three essential phases, each targeting a core challenge.

In the first phase, the method derives the most informative and representative features for every individual drug. In the second phase, a Positive

Unlabeled (PU) learning algorithm is applied to handle uncertainty in the input data effectively. This PU learning procedure serves as a preprocessing step that helps differentiate positive from negative samples among unlabeled instances.

In the final phase, a Multi-Layer Perceptron model is employed to forecast potential side effects when two drugs are taken together. **Figure 1** illustrates the overall architecture of the proposed PU-MLP method.

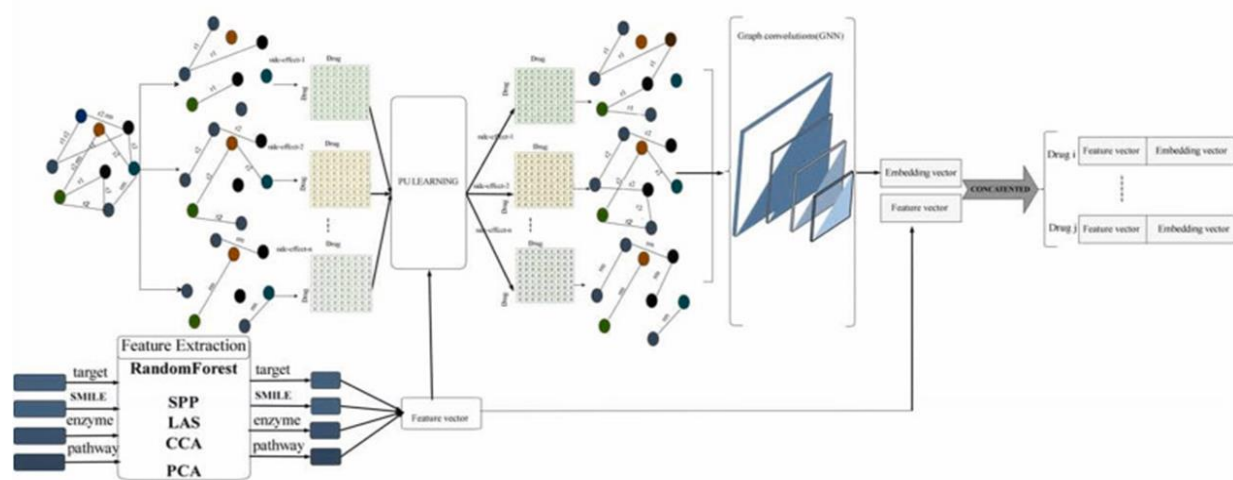


Figure 1. The general view of the proposed method.

Feature Vector Extraction

This study applies sophisticated machine learning methods to deepen understanding of the associations between medications and their adverse effects. The feature extraction process includes two primary elements.

(1) Random forest classifier for feature extraction: The process starts with a Random Forest classifier to identify the most relevant features for each drug across different side effects. This model is well-suited to capturing intricate interactions and effectively identifies which attributes most influence predictions of particular adverse reactions. The result is a feature vector for each drug that summarizes its key properties linked to known side effects.

In addition to the Random Forest, dimensionality reduction is achieved through biclustering. In this biclustering process, each identified cluster is represented as a column in the feature matrix.

(2) Modeling structural relationships: The method then addresses the relational structure of drug-side effect networks. A Graph Neural Network (GNN) is utilized to produce embedding vectors that encode the topological characteristics of these networks [43]. This captures connectivity patterns among drugs and side effects, providing valuable additional context from the overall network architecture.

The drug feature vectors obtained from the Random Forest model are subsequently merged with the topological embeddings generated by the GNN [43]. The resulting combined feature vector incorporates essential

chemical and biological attributes of each drug while also reflecting its contextual relationships within the larger network of drugs and side effects.

Reducing uncertainty in input data using positive unlabeled (PU) learning

Uncertainty in input data poses a major obstacle across many fields, especially in healthcare. In the context of polypharmacy, combinations of two drugs that lack experimental validation are frequently treated as having no interaction (i.e., labeled as zero). Such assumptions can significantly impair predictive model performance. To mitigate this problem, the second stage of the PU-MLP framework applies a Positive-Unlabeled (PU) learning algorithm. This step assigns labels to previously unknown samples, producing a cleaner, more reliable dataset.

The study adopts the Positive-Unlabeled Learning technique proposed by Elkan and Noto [39], which is highly suitable when only positive examples and unlabeled instances are available. The fundamental principle involves extracting knowledge from the confirmed positive samples while effectively estimating the distribution of the unlabeled data. The Elkan-Noto PU classifier reduces classification errors through iterative model refinement, enabling better separation of true positives from likely negatives within the unlabeled set.

We implemented the Elkan-Noto PU Classifier from the pulearn library, which executes this strategy by providing reliable probability estimates for both positive and

negative classes. This approach helps the model better manage the inherent uncertainty in unlabeled data. By emphasizing maximizing the likelihood of the positive class while properly accounting for unlabeled examples,

the method delivers superior classification results compared to standard supervised learning techniques.

Figure 2 depicts the PU-learning procedure employed in the second stage of the proposed method.

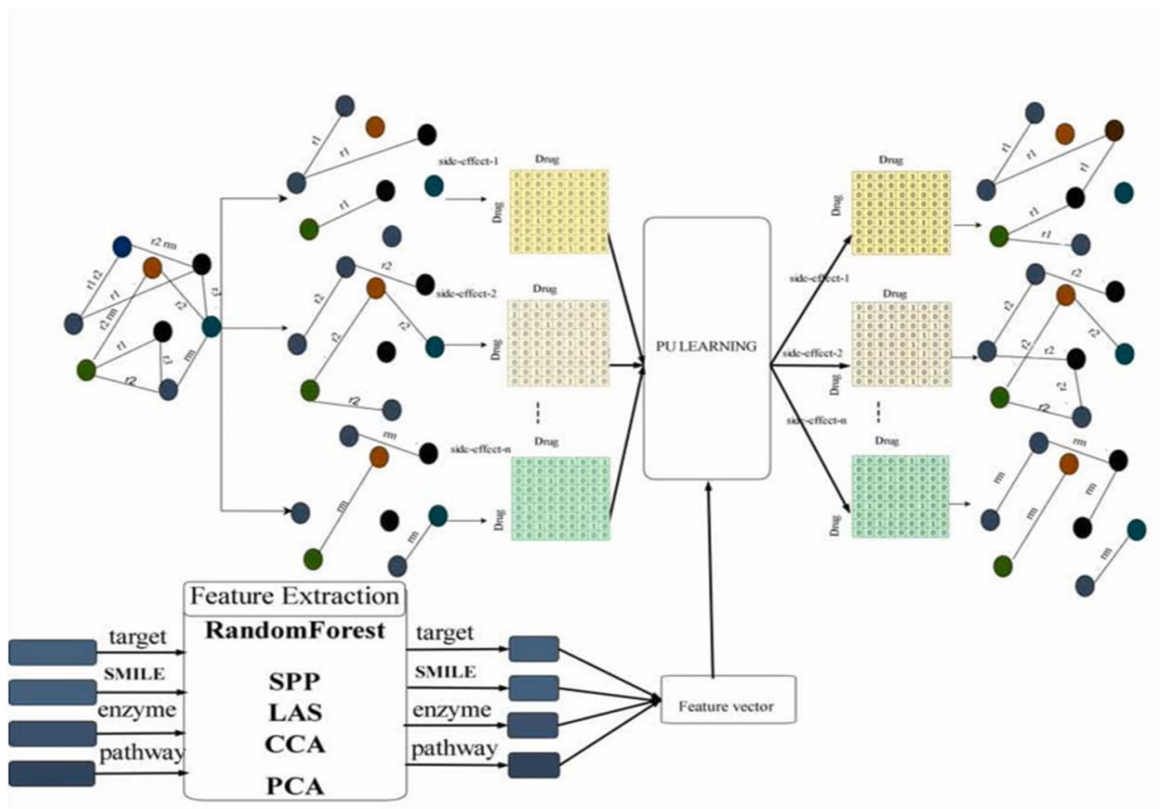


Figure 2. Positive Unlabeled (PU) learning in the second step of the proposed method.

Prediction of polypharmacy side effects

After extracting high-quality features and reducing input-data uncertainty in earlier stages, the framework forecasts side effects arising from drug combinations. This is accomplished using a Multi-Layer Perceptron (MLP) model tailored for binary classification. The network architecture includes an input layer, two hidden layers, and an output layer.

Training the neural network model

To determine the most effective neural network configuration using the training data, the following hyperparameters were explored:

- Number of hidden layers: {1, 2, 3, 4}
- Number of neurons in hidden layers: The output size of each layer is set to half the size of its input.
- Activation functions: {Rectified Linear Unit (ReLU), hyperbolic tangent (tanh), and sigmoid}

- Dropout rate: {0.05, 0.1, 0.2, 0.3}
- Learning rate: {0.01, 0.005, 0.001}
- Momentum: {0.7, 0.8, 0.9}

Multiple networks with one to four hidden layers were trained while varying the number of neurons. The single-hidden-layer configuration delivered superior performance without substantially extending training time compared to models with two to four hidden layers.

Architecture overview

1. Input layer: Receives the prepared feature vectors for each drug pair.
2. Hidden layers: The model uses one hidden layer. Each hidden layer's output dimension is half its input dimension, which helps normalize the data and stabilize training.
3. Activation functions:

- Hidden layers: ReLU (Rectified Linear Unit) is applied in the hidden layer(s). ReLU is selected because it helps alleviate the vanishing gradient issue and promotes quicker convergence during training.
 - Output layer: The sigmoid activation function is used, producing values between 0 and 1 suitable for binary classification.
 - ReLU introduces necessary non-linearity to the model.
4. Batch normalization: Applied to normalize layer inputs and improve overall training stability.
 5. Dropout: A 10% dropout rate is used to reduce the risk of overfitting.
 6. Learning rate: Set to 0.01 to support stable training.
 7. Momentum: A value of 0.9 is used to accelerate and enhance training efficiency.
 8. Optimizer: Stochastic Gradient Descent (SGD) with momentum is used to speed convergence. Learning rate decay is also applied to further stabilize the optimization process.
 9. Loss function: Binary cross-entropy is utilized, which is well-suited for binary classification problems.
 10. Early stopping: `early_stopping = EarlyStopping(monitor='val_loss', patience=10, verbose=0, mode='auto')`. This mechanism tracks validation loss and terminates training if no improvement is observed for 10 epochs, helping to avoid overfitting.

Training and functionality

The neural network is trained by minimizing classification errors using the binary cross-entropy loss function and an optimizer such as SGD or Adam. Through its multi-layer design and carefully chosen activation functions, the model learns to identify distinguishing patterns between the two classes, thereby improving its predictive capability.

Figure 3 presents the complete schematic diagram of the proposed method for predicting polypharmacy side effects.

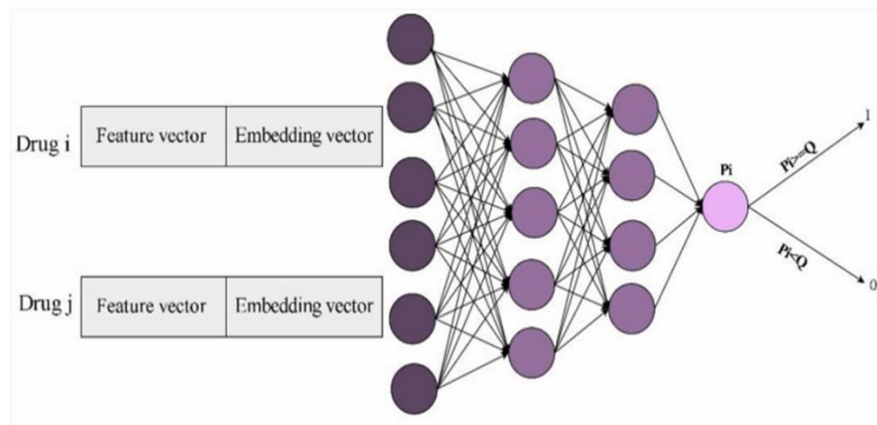


Figure 3. Illustration of polypharmacy side effects prediction during the third phase of the suggested approach.

Results and Discussion

Evaluation measures

The task of forecasting side effects caused by the simultaneous use of multiple drugs is formulated as a multi-label multi-class classification problem. Model effectiveness is measured with common performance indicators such as AUROC, AUPRC, and F1 score. To ensure reliable assessment, all tests were performed using 5-fold cross-validation.

Ablation study

An ablation analysis was performed to clarify the role and contribution of individual elements within the proposed models. The study is organized into three main parts: (1) assessing the inclusion versus exclusion of PU-Learning, (2) comparing different classifiers used for feature extraction, and (3) investigating how various dimensionality reduction techniques affect the drug feature representations.

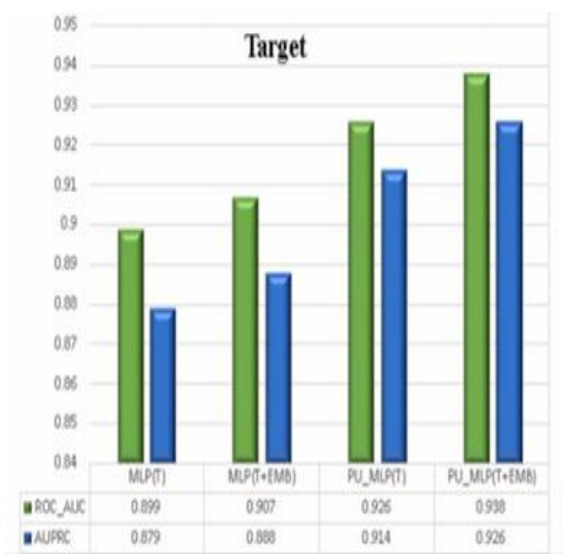
Impact of employing or omitting positive-unlabeled learning

This set of experiments explores the influence of Positive-Unlabeled (PU) Learning on overall predictive accuracy. The method applies PU Learning to better leverage unlabeled instances by relying on available positive examples when negative labels are missing. Evaluations were conducted separately on the Target (T), Pathway (P), Enzyme (E), and SMILES (S) datasets. For each one, we also tested the advantage gained by fusing the extracted feature vector with embeddings produced by the Graph Neural Network (GNN) subnetwork. Outcomes comparing the individual and integrated configurations are shown in **Figure 4**. These findings underscore the benefits of PU Learning within the model. The columns in **Figure 4** correspond to the following setups:

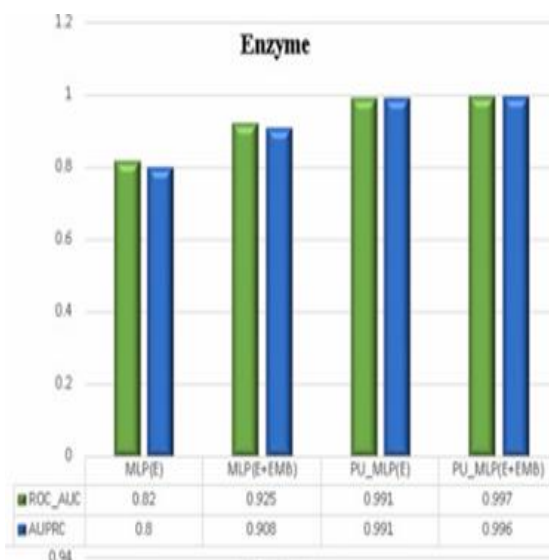
- Column 1: Features Extracted from Random Forest (without PU Learning). Features are obtained solely from a Random Forest model without applying PU Learning. This configuration identifies primary patterns in the data and provides a reliable reference point, capitalizing on Random Forest's proficiency in handling feature interactions and generating reliable importance scores.
- Column 2: Features from Random Forest + GNN Embedding (without PU Learning). Random Forest

features are combined with GNN-learned embeddings. The fusion harnesses Random Forest's feature selection capabilities and the GNN's ability to model relational information, which is expected to uncover complex associations that might otherwise go undetected.

- Column 3: Features Extracted from Random Forest (with PU Learning). PU Learning is incorporated alongside features generated by Random Forest. This combination is crafted to strengthen learning under positive-unlabeled conditions, thereby improving pattern recognition and generalization to new data.
- Column 4: Features from Random Forest + GNN Embedding (with PU Learning). The most elaborate variant merges PU Learning with both Random Forest features and GNN embeddings. By integrating strong feature derivation, graph-based relational insight, and robust handling of incomplete labels, this setup is projected to achieve the highest classification performance. A side-by-side comparison of the columns demonstrates how PU Learning interacts with alternative feature extraction strategies and illuminates their collective impact on model results.



(a)



(b)

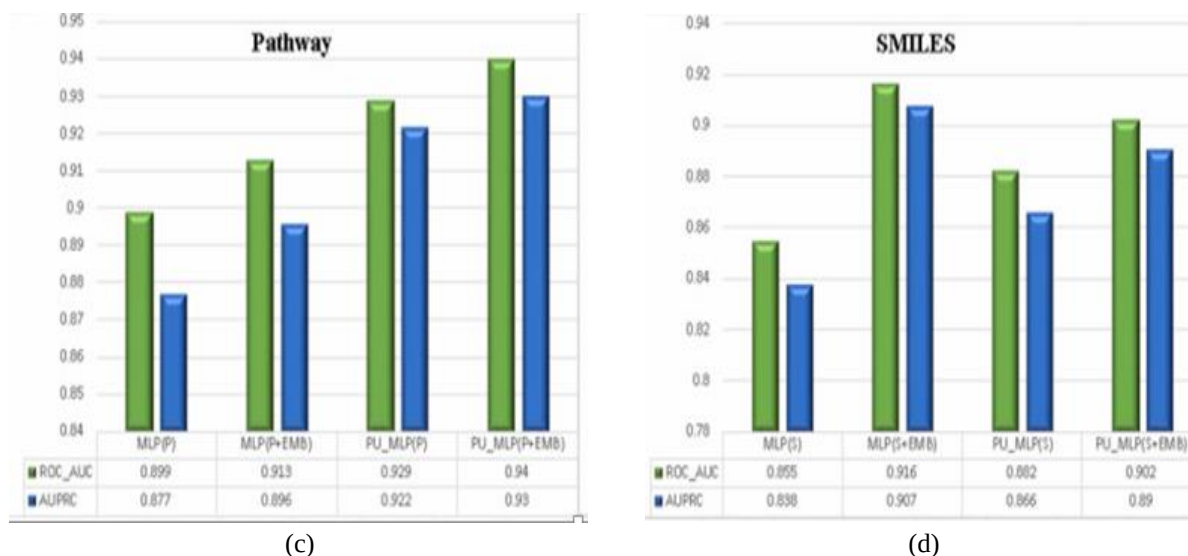


Figure 4. Impact of employing or excluding PU-Learning across various drug feature vector configurations: T: Target, E: Enzyme, P: Pathway, S: SMILES, GNN: GNN-based features.

Effects of different classifiers on feature vector extraction

In this part of the investigation, a feature-extraction technique is first applied to uncover important attributes in the dataset. Several classifiers are then evaluated to boost the overall predictive performance. The tested classifiers included Decision Trees, Neural Networks, and Gaussian Naive Bayes. Among them, the Random Forest classifier achieved the highest accuracy. **Table 2**

provides a detailed comparison of the accuracy scores obtained by these classifiers. The Random Forest approach consistently surpassed the alternatives and was therefore chosen as the optimal setup for the final model. This systematic process helps pinpoint the most influential factors affecting the target outcome, offering valuable guidance for enhancing model effectiveness in subsequent work.

Table 2. Effects of using different classifiers for feature vector extraction. Abbreviations: T = Target, NB = Gaussian NB, LR = Logistic Regression, RF = Random Forest, SVM = Support Vector Machine, DT = Decision Tree Classifier, MLP = Multi-layer Perceptron.

Model	AUROC	AUPRC
PU-MLP_(RF(T))	0.957	0.951
PU-MLP_(MLP(T))	0.953	0.943
PU-MLP_(NB (T))	0.952	0.946
PU-MLP_(SVM (T))	0.954	0.945
PU-MLP_(DT (T))	0.952	0.942
PU-MLP_(LR (T))	0.951	0.946

Impact of various dimensionality reduction techniques

In this work, multiple biclustering approaches were applied to efficiently organize and group drug-related features. The main objective was to detect meaningful biclusters within the drug characteristics, which could subsequently be used to construct representative feature vectors for individual drugs.

The clustering process generated binary feature vectors, which were then merged with embedding-based features that encode supplementary drug information, including

side effects and additional pharmacological properties. Such fusion improves the richness and quality of drug representations, supporting more effective downstream analysis. The evaluation involved several types of drug features—namely Target, Enzyme, Pathway, and SMILES. For every feature vector, experiments were conducted under two scenarios: with and without the inclusion of SPP (Similar Patterns of target-Proteins) [43], LAS (Large Average Submatrices) [47], CCA (Canonical Correlation Analysis) [48], other bi-

clustering algorithms, and PCA (Principal Component Analysis) [49]. We further examined different

combinations of these features. The outcomes of this evaluation are presented in **Figure 5**.

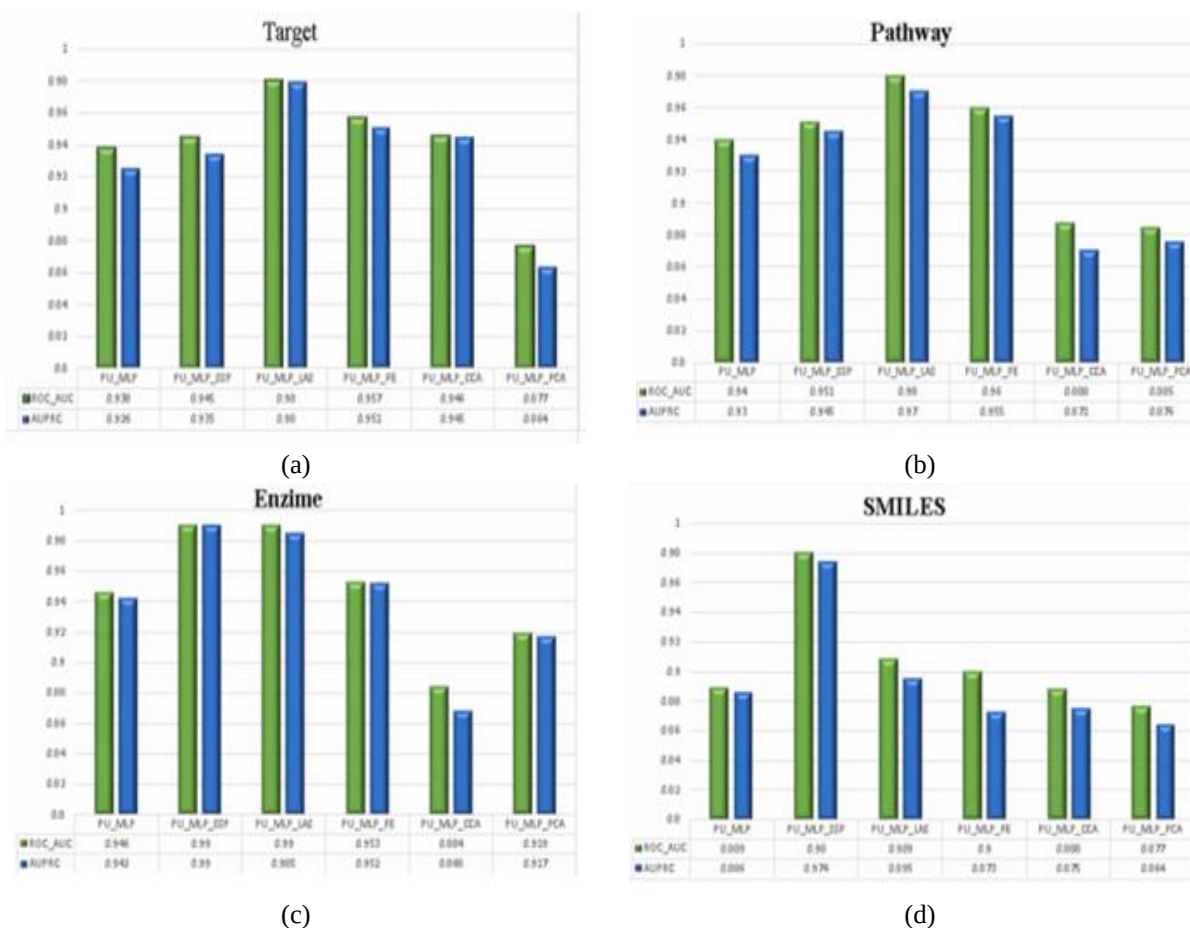


Figure 5. Outcomes achieved by applying different drug features along with their various combinations in the suggested model. Abbreviations: FE = Feature Extraction, SPP = Similar Patterns of target-Proteins, LAS = Large Average Submatrices, CCA = Canonical Correlation Analysis, PCA = Principal Component Analysis, PU = Positive-Unlabeled Learning.

Comparison with other methods

To verify the strength of our developed framework, performance comparisons were carried out against multiple established state-of-the-art techniques on both the DS1 and DS2 datasets, as summarized in **Table 1**. Reliable assessment was ensured through 5-fold cross-validation. Experiments on the DS dataset used both balanced and imbalanced distributions of 0/1-labeled samples.

Comparison results on the DS1 dataset (balanced version)

A balanced version of the data was prepared by sampling positive and negative examples equally. After splitting

the data into training, validation, and test sets, we ran several benchmarking experiments. Evaluation relied on F1, AUPR, and AUROC metrics. The proposed framework was benchmarked against VGAE [31], NNPS [34], MDF-PSE [28], PSECNN [24], Trivec [22], TIP [23], KB-LRN [21], and PU_GNN [43]. The corresponding results for this balanced setting are shown in **Figure 6**. Overall, the proposed approach achieved superior performance across all evaluated metrics compared to all other methods.

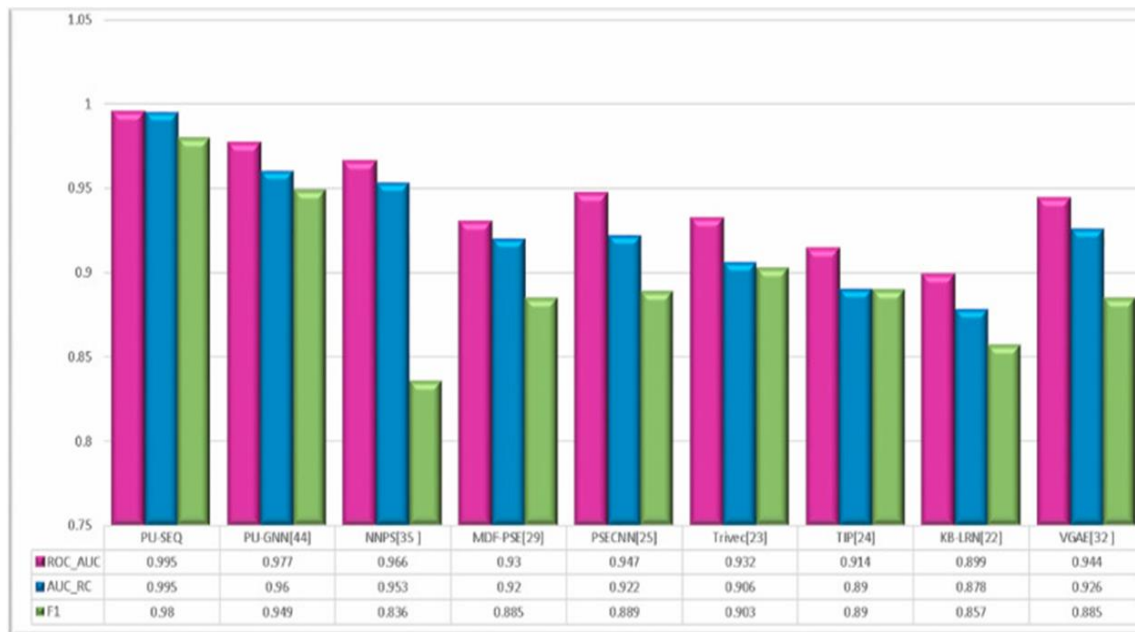


Figure 6. Comparison between the proposed method and other existing techniques in the balanced setting (DS1 dataset).

Comparison results on the DS1 dataset (imbalanced version)

For the imbalanced dataset scenario, the complete set of samples from DS1 was utilized, including all available positive and negative instances. The data was divided into training, validation, and test sets, followed by multiple benchmarking experiments. Performance was

assessed using F1, AUPR, and AUROC. The proposed model was evaluated against several contemporary methods: PSECNN [24], DeSIDE-DDI [30], DEDICOM [20], Decagon [11], and PU_GNN [43]. **Figure 7** shows results for the imbalanced case. These results confirm that the proposed approach consistently outperforms the competing methods across all performance measures

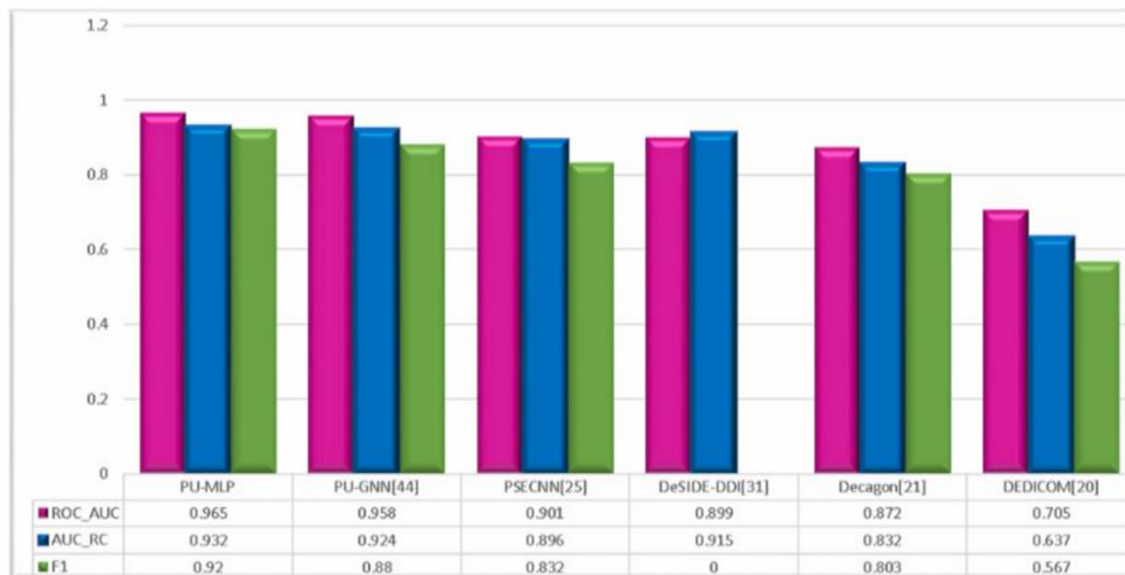


Figure 7. Comparison between the proposed model and competing techniques under the imbalanced condition on the DS1 dataset.

Comparison results on the DS2 dataset

After splitting the dataset into training, validation, and test sets, we ran the benchmarking process. PU_MLP achieved markedly higher performance than the other methods with respect to F1 score, precision, recall, and accuracy. To enable a more detailed evaluation, we directly compared the outcomes of our PU_MLP model

with those of the top-performing method across all examined methods. In **Table 3**, our approach is compared against eight prominent existing techniques—DPSP [35], GNN-DDI [27], MSTE [50], MDF-SA-DDI [7], NNPS [34], PU-GNN [43], CNN-Siam [51], CNN-DDI [27], and DDIMDL [52]—across ten repeated runs.

Table 3. Results of comparison of the PU-MLP with some of the machine-learning methods on DS2.

model	AUROC	AUPRC	F1	Precision	Recall	ACC
PU-MLP(E)	0.992	0.999	0.980	0.983	0.978	0.998
PU-GNN [43]	0.952	0.996	0.980	0.980	0.980	0.972
CNN-Siam[51]	0.963	0.999	0.924	0.924	0.924	0.924
CNN-DDI[27]	0.925	0.998	0.750	0.856	0.722	0.887
DPSP[35]	0.977	0.999	0.931	0.931	0.931	0.934
GNN-DDI[27]	0.971	0.999	0.899	0.899	0.899	0.918
MSTE[50]	0.934	0.998	0.847	0.847	0.847	0.858
MDF-SA-DDI[7]	0.977	0.999	0.912	0.938	0.891	0.929
NNPS[34]	0.970	0.999	0.882	0.882	0.882	0.911
DDIMDL[52]	0.921	0.998	0.759	0.847	0.718	0.885

Conclusion

Administering several medications at once, commonly referred to as polypharmacy, has become a widespread practice in today's medical field. This is especially true for elderly individuals and those dealing with intricate medical conditions. Although using multiple drugs is often required to control persistent health problems, such combinations can interfere with each drug's intended action and produce unanticipated adverse reactions. Consequently, a growing body of research has examined the link between polypharmacy and the potential for harmful side effects.

This study introduces a three-phase framework to strengthen computational prediction of side effects from drug combinations. The main obstacles to accurately forecasting these interactions include obtaining optimal feature representations for each drug and properly accounting for uncertainty in the available data. To address these challenges, we developed a multi-class, multi-label model, PU-MLP, based on a Multi-Layer Perceptron architecture.

The PU-MLP system proceeds in three essential phases. Initially, it identifies and extracts the most relevant features for every drug. Next, Positive Unlabeled (PU) learning is applied as a preprocessing technique to reduce data uncertainty and improve predictive reliability by

distinguishing likely positive and negative samples from unlabeled ones. In the last phase, a Multi-Layer Perceptron is employed to forecast side effects when two drugs are taken together.

The comprehensive evaluation demonstrates that the proposed framework significantly improves the computational identification of polypharmacy side effects and delivers superior results compared to previously published methods.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Verrotti A, Tambucci R, Di Francesco L, Pavone P, Iapadre G, Altobelli E, et al. The role of polytherapy in the management of epilepsy: suggestions for rational antiepileptic drug selection. *Expert Rev Neurother.* 2020;20(2):167-73.
2. Sun X, Zhang Y, Zhou Y, Lian X, Yan L, Pan T, et al. NPCDR: natural product-based drug combination

- and its disease-specific molecular regulation. *Nucleic Acids Res.* 2022;50(D1):D1324-33.
3. Liu H, Zhang W, Zou B, Wang J, Deng Y, Deng L. Drug Comb DB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res.* 2020;48(D1):D871-81.
 4. Qato DM, Wilder J, Schumm LP, Gillet V, Alexander GC. Changes in prescription and over-the-counter medication and dietary supplement use among older adults in the United States, 2005 vs 2011. *JAMA Intern Med.* 2016;176(4):473-82.
 5. Kantor ED, Rehm CD, Haas JS, Chan AT, Giovannucci EL. Trends in prescription drug use among adults in the United States from 1999-2012. *JAMA.* 2015;314(17):1818-30.
 6. Sun Y, Sheng Z, Ma C, Tang K, Zhu R, Wu Z, et al. Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. *Nat Commun.* 2015;6(1):8481.
 7. Lin S, Wang Y, Zhang L, Chu Y, Liu Y, Fang Y, et al. 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.* 2022;23(1):bbab421.
 8. Feng YH, Zhang SW, Zhang QQ, Zhang CH, Shi JY. deepMDDI: a deep graph convolutional network framework for multi-label prediction of drug-drug interactions. *Anal Biochem.* 2022;646:114631.
 9. Ding Y, Tang J, Guo F, Zou Q. Identification of drug-target interactions via multiple kernel-based triple collaborative matrix factorization. *Brief Bioinform.* 2022;23(2):bbab582.
 10. Deng Y, Qiu Y, Xu X, Liu S, Zhang Z, Zhu S, et al. META-DDIE: predicting drug-drug interaction events with few-shot learning. *Brief Bioinform.* 2022;23(1):bbab514.
 11. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics.* 2018;34(13):i457-66.
 12. Shah BM, Hajjar ER. Polypharmacy, adverse drug reactions, and geriatric syndromes. *Clin Geriatr Med.* 2012;28(2):173-86.
 13. Valenza PL, McGinley TC, Feldman J, Patel P, Cornejo K, Liang N, et al. Dangers of polypharmacy. In: *Vignettes in patient safety-volume 1.* IntechOpen; 2017.
 14. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, et al. Uncovering disease-disease relationships through the incomplete interactome. *Sci Transl Med.* 2015;347(6224):1257601.
 15. Zhang P, Wang F, Hu J, Sorrentino R. Label propagation prediction of drug-drug interactions based on clinical side effects. *Sci Rep.* 2015;5(1):12339.
 16. Nickel M, Tresp V, Kriegel HP. A three-way model for collective learning on multi-relational data. In: *ICML.* 2011;11:3104482-3104584.
 17. Zong N, Kim H, Ngo V, Harismendy O. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. *Bioinformatics.* 2017;33(15):2337-44.
 18. Žitnik M, Zupan B. Data fusion by matrix factorization. *IEEE Trans Pattern Anal Mach Intell.* 2014;37(1):41-53.
 19. Huang L, Li F, Sheng J, Xia X, Ma J, Zhan M, et al. DrugComboRanker: drug combination discovery based on target network analysis. *Bioinformatics.* 2014;30(12):i228-36.
 20. Papalexakis EE, Faloutsos C, Sidiropoulos ND. Tensors for data mining and data fusion: models, applications, and scalable algorithms. *ACM Trans Intell Syst Technol.* 2016;8(2):1-44.
 21. Malone B, García-Durán A, Niepert M. Knowledge graph completion to predict polypharmacy side effects. In: *International conference on data integration in the life sciences.* Cham: Springer International Publishing; 2018. p. 144-9.
 22. Nováček V, Mohamed SK. Predicting polypharmacy side-effects using knowledge graph embeddings. *AMIA Jt Summits Transl Sci Proc.* 2020;2020:449.
 23. Xu H, Sang S, Lu H. Tri-graph information propagation for polypharmacy side effect prediction. *arXiv preprint arXiv:2001.10516.* 2020.
 24. Lakizadeh A, Babaei M. Detection of polypharmacy side effects by integrating multiple data sources and convolutional neural networks. *Mol Divers.* 2022;26(6):3193-203.
 25. Liu S, Zhang Y, Cui Y, Qiu Y, Deng Y, Zhang Z, et al. Enhancing drug-drug interaction prediction using deep attention neural networks. *IEEE/ACM Trans Comput Biol Bioinform.* 2022;20(2):976-85.
 26. Zhang C, Lu Y, Zang T. CNN-DDI: a learning-based method for predicting drug-drug interactions using convolution neural networks. *BMC Bioinformatics.* 2022;23(Suppl 1):88.

27. Al_Rabeah MH, Lakizadeh A. GNN-DDI: a new data integration framework for predicting drug-drug interaction events based on graph neural networks. 2022. doi:10.21203/rs.3.rs-1803318/v1
28. Lin S, Zhang G, Wei DQ, Xiong Y. DeepPSE: prediction of polypharmacy side effects by fusing deep representation of drug pairs and attention mechanism. *Comput Biol Med.* 2022;149:105984.
29. Lukashina N, Kartysheva E, Spjuth O, Virko E, Shpilman A. SimVec: predicting polypharmacy side effects for new drugs. *J Cheminform.* 2022;14(1):49.
30. Kim E, Nam H. DeSIDE-DDI: interpretable prediction of drug-drug interactions using drug-induced gene expressions. *J Cheminform.* 2022;14(1):9.
31. Ngo NK, Hy TS, Kondor R. Modeling polypharmacy and predicting drug-drug interactions using deep generative models on multimodal graphs. *arXiv preprint arXiv:2302.08680.* 2023.
32. Lakizadeh A, Mir-Ashrafi SMH. Drug repurposing improvement using a novel data integration framework based on the drug side effect. *Inform Med Unlocked.* 2021;23:100523.
33. Ghasemi S, Lakizadeh A. Drug-disease association prediction based on end-to-end multi-layer heterogeneous graph convolutional encoders. *Inform Med Unlocked.* 2023;37:101177.
34. Masumshah R, Aghdam R, Eslahchi C. A neural network-based method for polypharmacy side effects prediction. *BMC Bioinformatics.* 2021;22:1-17.
35. Masumshah R, Eslahchi C. DPSP: a multimodal deep learning framework for polypharmacy side effects prediction. *Bioinform Adv.* 2023;3(1):vbad110.
36. Nejati M, Lakizadeh A. GADNN: a graph attention-based method for drug-drug association prediction considering the contribution rate of different types of drug-related features. *Inform Med Unlocked.* 2024;44:101429.
37. Zhang Y, Qiu Y, Cui Y, Liu S, Zhang W. Predicting drug-drug interactions using multi-modal deep auto-encoders based network embedding and positive-unlabeled learning. *Methods.* 2020;179:37-46.
38. Liu B, Dai Y, Li X, Lee WS, Yu PS. Building text classifiers using positive and unlabeled examples. In: *Third IEEE International Conference on Data Mining.* IEEE; 2003. p. 179-86.
39. Elkan C, Noto K. Learning classifiers from only positive and unlabeled data. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Las Vegas, NV; 2008. p. 213-20.
40. Mordelet F, Vert JP. A bagging SVM to learn from positive and unlabeled examples. *Pattern Recognit Lett.* 2014;37:201-9.
41. Zheng Y, Peng H, Zhang X, Zhao Z, Gao X, Li J. DDI-PULearn: a positive-unlabeled learning method for large-scale prediction of drug-drug interactions. *BMC Bioinformatics.* 2019;20(19):1-12.
42. Hameed PN, Verspoor K, Kusljic S, Halgamuge S. Positive-unlabeled learning for inferring drug interactions based on heterogeneous attributes. *BMC Bioinformatics.* 2017;18:1-15.
43. Keshavarz A, Lakizadeh A. PU-GNN: a positive-unlabeled learning method for polypharmacy side-effects detection based on graph neural networks. *Int J Intell Syst.* 2024;2024(1):4749668.
44. Jin Q, Xie J, Huang D, Zhao C, He H. Msff-ma-ddi: multi-source feature fusion with multiple attention blocks for predicting drug-drug interaction events. *Comput Biol Chem.* 2024;108:108001.
45. Li YC, You ZH, Yu CQ, Wang L, Wong L, Hu L, et al. PPAEDTI: personalized propagation auto-encoder model for predicting drug-target interactions. *IEEE J Biomed Health Inform.* 2022;27(1):573-82.
46. He S, Yun L, Yi H. Fusing graph transformer with multi-aggregate GCN for enhanced drug-disease associations prediction. *BMC Bioinformatics.* 2024;25(1):79.
47. Abdul-Jabbar SS, Saeed HS, Shihab SS. Comparative study for Bi-clustering algorithms: historical and methodological notes. *Solid State Technol.* 2020;63(5):5142-52.
48. Adossa N, Khan S, Rytönen KT, Elo LL. Computational strategies for single-cell multi-omics integration. *Comput Struct Biotechnol J.* 2021;19:2588-96.
49. Hammoumi D, Al-Aizari HS, Alaraidh IA, Okla MK, Assal ME, Al-Aizari AR, et al. Seasonal variations and assessment of surface water quality using water quality index (WQI) and principal component analysis (PCA): a case study. *Sustainability.* 2024;16(13):5644.
50. Yao J, Sun W, Jian Z, Wu Q, Wang X. Effective knowledge graph embeddings based on

- multidirectional semantics relations for polypharmacy side effects prediction. *Bioinformatics*. 2022;38(8):2315-22.
51. Yang Z, Tong K, Jin S, Wang S, Yang C, Jiang F. CNN-Siam: multimodal siamese CNN-based deep learning approach for drug–drug interaction prediction. *BMC Bioinformatics*. 2023;24(1):110.
52. Deng Y, Xu X, Qiu Y, Xia J, Zhang W, Liu S. A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics*. 2020;36(15):4316-22.