Dual representation learning for predicting drug-side effect frequency using protein target information

Sungjoon Park, Sangseon Lee, Minwoo Pak, and Sun Kim

***Abstract*—Knowledge of unintended effects of drugs is critical in assessing the risk of treatment and in drug repurposing. Although numerous existing studies predict drug-side effect presence, only four of them predict the frequency of the side effects. Unfortunately, current prediction methods (1) do not utilize drug targets, (2) do not predict well for unseen drugs, and (3) do not use multiple heterogeneous drug features. We propose a novel deep learning-based drug-side effect frequency prediction model. Our model utilized heterogeneous features such as target protein information as well as molecular graph, fingerprints, and chemical similarity to create drug embeddings simultaneously. Furthermore, the model represents drugs and side effects into a common vector space, learning the dual representation vectors of drugs and side effects, respectively. We also extended the predictive power of our model to compensate for the drugs without clear target proteins using the Adaboost method. We achieved state-of-the-art performance over the existing methods in predicting side effect frequencies, especially for unseen drugs. Ablation studies show that our model effectively combines and utilizes heterogeneous features of drugs. Moreover, we observed that, when the target information given, drugs with explicit targets resulted in better prediction than the drugs without explicit targets. The implementation is available at https://github.com/eskendrian/sider.**

***Index Terms*—drug-side effect frequency, dual representation learning, adverse drug reaction, drug target protein**

I. Introduction

DRUGS contributed to human lives by alleviating pain and curing diseases. However, they are not panaceas and often involve unintended side effects. Predicting side effects of drugs is one of the most important challenges of modern medicine. Adversarial side effects may result in poisonous effects. Thalidomide, as a notorious example, had unexpected side effect of causing deformities in pregnant fetuses, and about 10,000 victims suffered the consequences [1]. Even now, side effect is the fourth major cause of death in the United States [2], [3], [4]. Predicting side effect frequencies can help weigh the potential risks and benefits of using certain drugs [5].

Side effects also account for the second reason for failure to develop new drugs, following lack of efficacy [6]. Drug discovery pipeline consists of virtual screening followed by pre-clinical and clinical testing of efficacy and toxicity. Therefore, early detection of side effects in the stage of virtual screening can reduce enormous costs in the whole drug discovery process. Post-marketing drug withdrawal causes the health risk of public [7] and financial costs [8], [9].

Unintended side effects do not always lead to bad results. Side effects sometimes lead to drug repurposing [10]. For instance, Sildenafil was originally developed to treat angina, but the desired cardiovascular effect was not observed. Instead, it is nowadays more known as a medication treated for erectile dysfunction [11]. The knowledge of the frequencies of side effect for existing drugs can facilitate the introduction of new drugs [12], [13].

Although it is important to know the side effects of drugs, it is a challenging work. Side effects are rare by nature, and only observed in a small population of patient groups. Information about side effects of drugs is incomplete and unobserved side effects can be newly reported at any time. This makes empirical evaluation of side effects costly and time-consuming. For these reasons, many researchers have tried to develop resources and computational methods for side effects of drugs. SIDER was introduced in 2010 as a database that collects known side effects of drugs from public documents and package inserts [14], [15]. Since the availability of such curated data, many studies focused on predicting drug side effects [16], [17], [18], [19], [20], [21], [22].

Lately, predicting frequencies of side effects has arisen as a new problem. Most drugs, even when treated with appropriate dosage, can result in side effects for some proportion of those who take the medication. Measuring the frequency of these side effects can be useful in determining the therapeutic efficacy of a drug in clinical settings. However, frequency of side effects has not been widely studied until reccently. To this day, only four studies have focused on predicting the frequencies of drug-side effects.

Manuscript received May 10, 2023. This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science & ICT (NRF-2022M3E5F3085677, NRF-2022M3E5F3085681, RS-2023-00257479), supported by the same foundation funded by the Ministry of Agriculture, Food and Rural Affairs (RS-2023-00231699), supported by a grant (No. DY0002259501) from Ministry of Food and Drug Safety, supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.2021-0-01343, Artificial Intelligence Graduate School Program (Seoul National University)], and funded by AIGENDRUG. The ICT at Seoul National University provides research facilities for this study.

S. Park and M. Pak are with the Department of Computer Science and Engineering, Seoul National University, Gwanak-ro 1, Gwanak-gu, Seoul, 08826, Republic of Korea (e-mail: stj@snu.ac.kr; mwpak@snu.ac.kr).

S. Lee is with Institute of Computer Technology, Seoul National University, Gwanak-ro 1, Gwanak-gu, Seoul, 08826, Republic of Korea (e-mail: sangseon486@snu.ac.kr).

S. Kim is with the Department of Computer Science and Engineering, Interdisciplinary Program in Artificial Intelligence, and Interdisciplinary Program in Bioinformatics, Seoul National University, and AIGENDRUG Co., Ltd., Gwanak-ro 1, Gwanak-gu, Seoul, 08826, Republic of Korea (e-mail: sunkim.bioinfo@snu.ac.kr).

(*Sungjoon Park and Sangseon Lee are co-first authors.) (Corresponding author: Sun Kim.*)

A. Related Works

SIDER database [15] currently provides drug-side effect association among 1,430 drugs and 5,868 side effects, and 40% of them have side effect frequency information. Galeano et al. [5] screened the database to integrate heterogeneous frequency labels of 759 drugs and 994 side effects into five classes ranging from very rare (label=1) to very frequent (label=5). Then, the authors formulated drug-side effect frequency prediction problem as a recommendation system and used matrix factorization. Since its publication, follow-up papers predicting drug-side effect frequency commonly used the curated dataset of 759 drugs and 994 side effects.

Numerous research endeavors have been undertaken to predict the presence of drug-side effects, employing a diverse array of methodologies. For example, Zhang et al. [23] utilized recommender techniques to predict potential associations between drugs and side effects. In this study, the researchers constructed an ensemble model that combined an integrated neighborhood-based approach with a method based on the restricted Boltzmann machine, culminating in an enhanced framework for predicting drug-side effect relationships. Another study by Zhang et al. [24] utilized matrix factorization coupled with graph regularization to generate vectors representing both drugs and side effects. In this work, it was assumed that drug targets significantly influence metabolic processes, thereby influencing the emergence of side effects. DeepSide [25] developed multi-modal and multi-task neural networks to forecast drug side effects. This work used drug-induced gene expression profiles from the LINCS L1000 dataset [26] to optimize the prediction accuracy for drug side effects. More recently, Galeano et al. [27] developed the development of a geometric self-expressive model (GSEM) to predict drug-side effect associations. This method involved the computation of two matrices: drug-drug similarity and side effect-side effect similarity matrices, which collectively facilitated the prediction of presence scores. Importantly, this framework showcased its adaptability to proprietary datasets. In another study, Liang et al. [28] adopted a distinctive approach to the drug-side effect prediction problem by employing positive-unlabeled learning. The authors used transductive matrix co-completion as a means for the prediction of unobserved side effects, with a notable emphasis on the pivotal role of drug-target interactions for the prediction of side effects.

However, these studies are developed for the prediction of the existence of side effects. In practice, medical practitioners prescribe drugs while considering a delicate equilibrium between the benefits and potential side effects. Therefore, the prediction of the quantitative frequency of side effects can be more informative. Recently, there are studies for the prediction of frequencies of side effects as summarized below.

Zhao et al. proposed MGPred [29], a graph attention network (GAT) [30] model to predict the drug-side effect frequency. The paper constructed heterogeneous graph of drugs and side effects, and learned important features based on the heterogeneous neighbors. That is, given drug and side effect, GAT extracts drug features based on drugs that carry the given side effect, and side effect features based on other side effects that are carried by given drug. MGPred first introduced GAT to enhance the prediction of drug-side effect frequency. However, it utilized the frequency of some side effects as input features, limiting the model only to work on known drugs.

Later, Zhao et al. developed SDPred [31] and investigated the role of similarities of drugs and side effects in their frequency. In the paper, the authors used 10 drug similarities and 4 side effect similarities fed into deep learning models such as multiple layer perceptrons (MLPs) and convolutional neural networks (CNNs). The paper showed that similarities from different sources can be used to predict the adverse reactions, even without chemical property of drugs or known relation to side effects. However, the authors note that SDPred is dependent to the complete similarity information of new drugs.

Most recently, Xu et al. suggested DSGAT [32], a novel GAT model to predict drug-side effect frequency. Unlike MGPred that applies GAT on drug-drug network graph, DSGAT used molecular structure of drugs as graph data to feed to GAT. This way, DSGAT works without requiring prior knowledge of novel drugs other than its molecular formula. The authors emphasize that DSGAT performs best for predicting side effect frequencies of unseen drugs, or "cold start" drugs. However, the model neglects other available information of drugs such as protein target information, which leaves room for improvement.

II. Motivation

The problem of drug-side effects is a critical issue, and previous research has highlighted the significance of protein targets as ingredients in determining the side effects of a drug [33]. Incorporating target information has also enhanced the detection of side effects [34], [35], [36], [37]. However, a more complex problem of estimating the frequency of drug-side effects has emerged, which has only been covered in four studies so far. Still, the useful target information is not currently getting attention in drug-side effect *frequency* studies.

In the subsequent sections, we outline the limitations of the existing studies that have examined drug-side effect frequency. Subsequently, we propose our approach to address the shortcomings and challenges faced by previous methods. We devised a dual representation learning method, that puts drugs and side effects onto a common embedding space, along with emerging technologies such as GAT (graph attention networks) and network propagation.

A. Limitations and Challenge

There are three major limitations for the existing side effect frequency prediction models. In this subsection, we list and describe the common limitations of existing studies. The limitations can be summarized as follows:

- Protein target information is a key to estimating side effects, which, however, has not been utilized.
- Existing models are not designed to effectively predict side effect frequencies for unseen novel drugs.

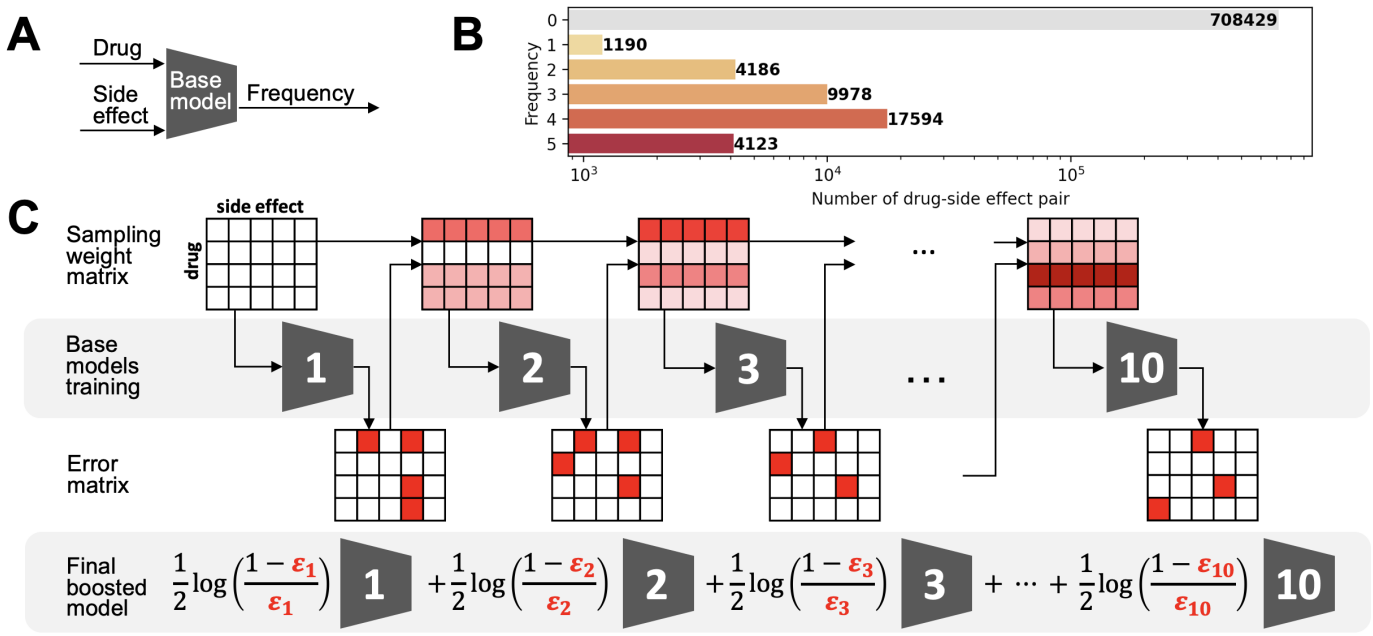

Fig. 1: Overview of training our drug-side effect frequency prediction framework. (A) Inputs and outputs of the base model for predicting drug-side effect frequency. Our model takes drug and side effect as input, and outputs frequency of the given drug-side effect pair. (B) Value distribution of frequency range of our data of 750 drugs and 994 side effects. (C) Adaboost boosting scheme. At each step, the base model is trained for 50 epochs, using the data with its sampling weight matrix. Here, the sampling weights assigned to drugs are augmented proportionally in relation to the rate of incorrect predictions so that they are better trained in the subsequent base model. Final model is a linear combination of base models using the log of overall error $\varepsilon_b (b = 1, 2, ..., 10)$ as coefficients.

- Heterogeneous features, such as molecular graph, fingerprints, drug-drug chemical similarity, and protein targets are not utilized simultaneously.

First, protein targets are not utilized for side effect frequency prediction. It is established that on- and off-targets are responsible for side effects of drugs. Studies also show that target is a key information related to drug side effects [38]. Some studies have found drug targets using side effect information [34], [39], while others investigated drug protein target to explain the side effects of the drugs [40], [41], [42]. Moreover, target validation is one of the first steps in discovering a new drug [43], thus targets are likely to be known by the time of screening candidate compounds and predicting their side effects [44]. Therefore, it is natural to leverage target information available at the stage of predicting side effect frequencies.

Second, Galeano et al. [5], MGPred [29], and SDPred [31] are not designed to predict unseen novel drugs. MGPred and SDPred directly use the known side effect frequency of drugs to predict the frequency of unknown side effects. Additionally, matrix factorization proposed by Galeano et al. decomposes drug vectors and side effect vectors, thus it is innately inapplicable for unseen drugs. In real-life drug discovery, known frequencies of side effects for novel drug candidates are often not available. We present in the Result section, that these models do not perform well.

Finally, the existing studies lack utilization of heterogeneous information from multiple sources. Galeano et al. [5], SDPred [31], and DSGAT [32] used drug-side effect frequency, drug-drug similarity, and molecular graph structure as the only input features, respectively. Side effect can be attributed from different factors of drugs, and current studies are limited to using only a single type of data.

B. Our approach

In this study, we developed a novel deep learning-based drug-side effect frequency prediction model. Our model is the first to introduce drug protein target as an important feature in predicting the frequencies of drug-side effects. We used various sources of drug features such as molecular graph, fingerprints, chemical similarity, as well as drug target proteins. We constructed a shared vector space for the dual representation for drugs and side effects in a single space. However, simply adding more heterogeneous input features to models does not guarantee accurate prediction of drug-side effect frequency. To tackle this problem, we ensembled multiple base models complementing different features using Adaboost technique. Our model showed significant performance gain compared to the current state-of-the-art method.

Utilization of drug protein target information. Information about drug protein on- and off-targets is available from DrugBank [45]. To simulate the downstream effect of the cellular perturbation by drug targets, we performed network propagation [46] on a protein-protein interaction (PPI) network with the seed genes as protein targets. This way, we not only encoded protein targets of drugs, but also considered the

downstream effect of perturbation of the targets in biological network.

Representation learning of drugs and side effects on common vector space. Our model is designed to be applicable to novel drugs for their side effect frequency. First, we created drug vectors from four embeddings of drug features, and side effect vectors from one embedding of side effect features, respectively. The dimensions of drug vectors and side effect vectors are kept the same. Then, we trained the representation of drugs and side effects while interpreting the (cosine) distance between drug and side effect vectors as their frequency. The continuous mapping of drugs and side effects achieved good generalization for unseen drugs.

Using heterogeneous multiple information sources. We inherited drug features, such as molecular graph structure, fingerprints, and drug-drug similarity, that are traditionally used to determine side effects. Additionally, we adopted drug protein features, and created embedding for each of the input feature. We used an ensemble technique to complement the models that use heterogeneous information.

III. Methods

To predict drug-side effect frequency, we developed a novel deep learning model that exploits multiple features complementarily in an Adaboost framework. Figure 1 illustrates the bird-eye view of our Adaboost boosting framework. In the following subsections, we describe the dataset, formulation of the research problem, and pre-processing of features, and depict our model architecture with its training process.

A. Dataset

Since the publication of Galeano et al. [5], studies predicting drug-side effect frequency commonly used the same dataset [29], [31], [32]. We also used the same benchmark dataset from the models to train and validate our model for fair comparison. The dataset includes 750 drugs and 994 side effects with $37{,}071$ drug-side effect pairs with observed frequencies. For a set of drugs $\mathcal{D} = \{d_1, d_2, ..., d_N\}$, a set of side effects $\mathcal{S} = \{s_1, s_2, ..., s_M\}$, and frequency range $\mathcal{F} = \{0, 1, 2, 3, 4, 5\}$, our goal is to build a deep learning model that best approximates the frequency function f such that

$$f : \mathcal{D} \times \mathcal{S} \rightarrow \mathcal{F} \tag{1}$$

In our case, $N = 750$ is the number of drugs, $M = 994$ the number of side effects, and the elements of $\mathcal{F}$ indicates 0: unobserved, 1: very rare, 2: rare, 3: infrequent, 4: frequent, 5: very frequent.

Among $745{,}500$ combinatorial pairs of 750 drugs and 994 side effects, only $37{,}071$ pairs have non-zero side effect frequency. In evaluating our model in terms of predicting presence of drug-side effects, we regarded zero frequency as non-existence of side effects.

However, though presently unobserved, the drug-side effect pairs with zero frequency could potentially be identified later. Even in such scenario, the identified frequency is expected to be low, because all drugs included in our study have undergone rigorous clinical trials and obtained drug approval. This further strengthens our perspective that prediction of the quantitative frequencies of side effects is more informative.

B. Heterogeneous feature generation

Campillos et al. [34] inspired us that side effects of drugs may be attributed to various factors, including compound structures, drug targets, and therapeutic categories, either individually or in combination. Therefore, we used heterogeneous set of features of drugs and side effects as features to construct different embeddings. Specifically, we used molecular graph structure, drug-drug similarity score, and drug target information as drug features. Likewise, we used MedDRA categories and Glove word embedding as side effect features.

To take the structural information of molecular graph into account, we adopted the molecular graph encoding similar to embeddings used in existing studies [32]. The atoms and bonds in drug compound are interpreted as nodes and edges of graph, respectively. For node feature, we use multi-hot vector encoding 109 atomic properties of each node. They include properties such as atom type, degree, total number of adjacent hydrogens, valence, formal charge, hybridization, and aromaticity of the atom. We denote the drug molecular graph $\texttt{DG}$ as

$$\texttt{DG} = (G_1, G_2, ..., G_N), \quad G_i = (V_i, E_i) \tag{2}$$

$$V_i = \{x_{i1}, x_{i2}, ..., x_{ik}\} \tag{3}$$

Here, $G_i = (V_i, E_i)$ is a graph for drug d_i. V_i denotes a set of atoms and E_i a set of bonds. All k atoms in V_i have 109-dimensional multi-hot features $x_{i1}, x_{i2}, ..., x_{ik}$.

Existing studies [29], [31] showed that both chemical similarity of drugs and drug-drug interaction are effective predictors of side effects. MGPred [29] used the `Combined_score` value of chemical-chemical link from STITCH database [47], which is then interpolated using chemical structure similarity calculated using a chemical development kit (cdk) [48]. Later, we use independent validation data to show that similarity to only 750 drugs from our dataset is enough to predict side effect frequencies of unseen drugs. We denote the drug similarity feature $\texttt{DS}$ as

$$\texttt{DS} = (w_1, w_2, ..., w_N) \tag{4}$$

where w_i is the similarity vector of drug d_i to all 750 drugs.

Drugs that share similar protein targets can also share similar side effects. We collected 843 protein targets for the 750 drugs from the DrugBank [45] database. There were 611 drugs that had at least one protein target, and the maximal number of protein targets for a single drug was 70. We define 843-dimensional multi-hot protein target vector z_i for each drug d_i. We denote the drug target information $\texttt{DT}$ as

$$\texttt{DT} = (z_1, z_2, ..., z_N). \tag{5}$$

Molecular fingerprints are multi-hot vector of pre-defined substructures. Its utility, especially when combined with graph neural network, is known to be useful [49]. Here, we use

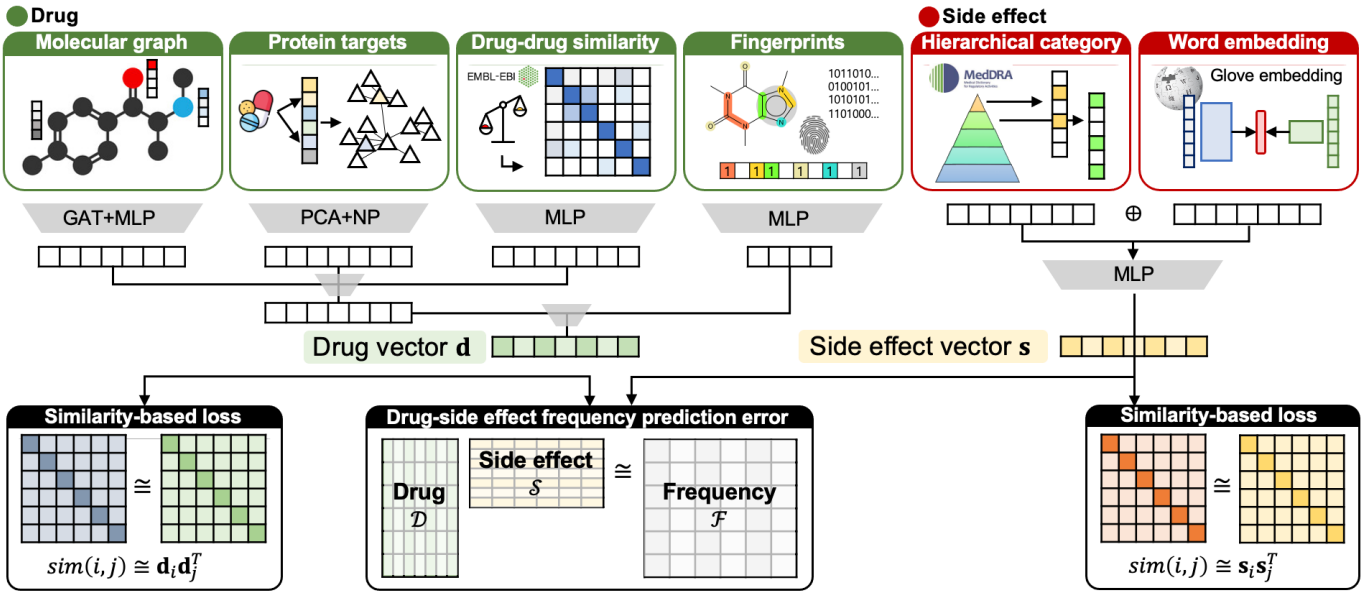

Fig. 2: Architecture of our base model. Drug and side effect are encoded into drug vector **d** and side effect vector **s**. Drug features include molecular graph, protein targets, drug-drug similarity, and fingerprints. Graph attention network, network propagation, and multi-layer perceptron are used to construct the embeddings for drug, and the embeddings are merged into a drug vector. On the other side, side effect vector is constructed using MedDRA hierarchical category and Glove word embedding as initial features. The final prediction of the drug-side effect frequency is an element-wise multiplication of the drug and side effect vectors. To guide the embedding training during Adaboost, similarity-based loss for drug vectors and side effect vectors are defined.

1024-bit Morgan fingerprint computed by rdkit to further complement the prediction of model. We denote molecular fingerprint DF as

$$\mathtt{DF} = (v_1, v_2, ..., v_N). \tag{6}$$

Side effects in SIDER database are provided in MedDRA [50] labels. MedDRA terms consist of hierarchical structures, starting with the System Organ Class at the top, followed by High Level Group Term, High Level Term, Preferred Term, and Lowest Level Term. In this paper, We encode side effects into 243-dimensional multi-hot MedDRA categorical vector u_i using 243 categories of System Organ Class and High Level Group Term. Therefore, we denote the side effect class feature SC as

$$\mathtt{SC} = (u_1, u_2, ..., u_M) \tag{7}$$

We also used word embedding used in language models to capture the contextual similarity among side effects. Glove [51] is a method that learns vector space representation of words using word co-occurrence. MGPred [29] used a Glove model pre-trained on Wikipedia to create semantic features of side effects.

$$\mathtt{SE} = (v_1, v_2, ..., v_M) \tag{8}$$

where v_i is a 300-dimensional vector for side effect s_i.

C. Dual representation learning of drugs and side effects

Our model constructs drug vectors and side effect vectors from the features mentioned above (Figure 2). For the drug vectors, we defined four different kinds of embeddings using molecular graph (DM), drug-drug similarity (DS), protein target (DT), and molecular fingerprint (DF), respectively. Then, the four embeddings are merged into a single drug vector (See Equation (15)).

First, we describe the embedding driven by molecular graph features DM that is given as node features with edge information of the graph topology. Here, we use GAT [30] layer to transform node features of dimension F into another dimension F'. Thus, the formula for a single GAT layer is described as follows. For input $\vec{h}_i$ of atom node i and output $\vec{h}'_i$,

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\vec{\mathbf{a}}^T[W\vec{h}_i || W\vec{h}_j]))}{\exp(\sum_{k\in\mathcal{N}_i}\text{LeakyReLU}(\vec{\mathbf{a}}^T[W\vec{h}_i || W\vec{h}_k]))} \tag{9}$$

$$\vec{h}'_i = \Big\Vert_{l=1}^{L} \sigma\Big(\sum_{j\in\mathcal{N}_j} \alpha_{ij}^{(l)} W^{(l)}\vec{h}_j\Big) \tag{10}$$

where α_{ij} indicates normalized importance of node j's features to node i, vector $\vec{\mathbf{a}}$ a weight vector, matrix W a linear transformation matrix, and $\mathcal{N}_i$ is a set of neighbor nodes that are connected to node i with covalent bond. Furthermore, $\|$ denotes concatenation in both Equation (9) and (10), and L is the number of attention heads.

Specifically, We used four GAT layers to transform the molecular graph features for atoms. We utilized multi-head attention ($L = 10$) until the penultimate layer, and $L = 1$ for the final GAT layer. The output of the consecutive four GAT layers is then aggregated using max pooling and fed to two

fully connected layers to project into embedding space. Thus, the molecular graph-based drug embedding

$$\vec{\mathbf{x}}_i = \sigma(W(\sigma(W(G_i') + b) + b) \tag{11}$$

where G_i' is the output of G_i fed to the four GAT layers followed by activation function, W and b denote weight matrix and bias vector of a linear layer.

Next, drug-drug similarity matrix `DS` is used to construct another embedding of drug. Here, we simply used two fully connected layers. That is, the similarity drug embedding

$$\vec{\mathbf{w}}_i = \sigma(W(\sigma(W(w_i) + b) + b) \tag{12}$$

where W and b denote weight matrix and bias vector of a linear layer.

Moreover, we constructed drug embedding of protein targets using network propagation proposed by Pak et al. [52]. We performed network propagation on STRING protein-protein interaction (PPI) network [53] to reflect the underlying biological mechanism of drug-target interaction (DTI). First, we constructed drug-specific PPI network for each drug in a way that only leaves the edges that are either with edge weight greater than 800 or connected to one of the drug targets. Then, we performed network propagation on the constructed graph with the drug protein targets `DT` being the seeds of the propagation. Genes with top 200 node values after convergence were selected. We performed gene set enrichment analysis (GSEA) with the selected genes to determine the enriched pathway from Reactome Pathway Database [54]. Next, the genes included in the biological pathway that are enriched with p-value less than 0.05 were selected as the next seed genes for the network propagation process. The process was repeated until convergence, and then the final values for all 19,127 nodes (genes) were fed to two fully connected layers for dimension reduction. Therefore, the target protein drug embedding

$$\vec{\mathbf{z}}_i = \sigma(W(\sigma(W(z_i') + b) + b) \tag{13}$$

where z_i' is the output of z_i after the network propagation process, W and b denote weight matrix and bias vector of a linear layer.

Lastly, we constructed drug embedding of molecular fingerprint using two fully connected layers. The fingerprint drug embedding is thus

$$\vec{\mathbf{v}}_i = \sigma(W(\sigma(W(v_i) + b) + b) \tag{14}$$

With four embeddings of drugs, we define the final drug vector $\vec{\mathbf{d}}$, using the embeddings $\vec{\mathbf{x}}$, $\vec{\mathbf{w}}$, $\vec{\mathbf{z}}$, and $\vec{\mathbf{v}}$. The first three embeddings are aggregated by a single fully connected layer. After that, the fingerprint embedding is concatenated to the aggregated embedding, which is to highlight the existence of particular subgraphs of a modelcule. All embeddings of a molecule are input to a fully connected layer and an activation layer. This way, the final embedding captures the signal from among molecular graph, drug similarity, protein target information, and molecular fingerprints.

$$\vec{\mathbf{d}} = \tanh(W(W(\vec{\mathbf{x}} \Big|\Big| \vec{\mathbf{w}} \Big|\Big| \vec{\mathbf{z}}) + b) \Big|\Big| \vec{\mathbf{v}}) + b \tag{15}$$

where $||$ is concatenation, W and b denote weight matrix and bias vector of a linear layer.

For side effect vector, we concatenate MedDRA categorical vectors `SC` and Glove word embedding `SE` and feed the concatenated vector into two fully connected layer followed by activation function.

$$\vec{\mathbf{s}} = \tanh(W_2(\text{ReLU}(W_1[u||v] + b_1)) + b_2) \tag{16}$$

where W_i and b_i denote weight matrices and bias vectors of linear layers.

Finally, we inner-product drug vector and side vector for the prediction of drug-side effect frequency. Here, $\vec{\mathbf{d}}$ and $\vec{\mathbf{s}}$ are both unit vectors, as their range is an output of a hyperbolic tangent function. To scale this range to cover our frequency values from 0 to 5, we multiply 5 to the inner product of two vectors.

$$f(d, s) = 5 \cdot \vec{\mathbf{d}} \cdot \vec{\mathbf{s}} \tag{17}$$

Dimensions of the GAT layers and hidden layers from Equation (11)-(16) are noted in the supplementary material (Supplementary Table S1).

D. Model training loss

Given drug d, side effect s, and their frequency y, we train the model using root mean square loss as follows:

$$\mathcal{L}_{rmse} = \sqrt{\sum_{d,s} \left(\alpha(f(d, s) - y)\right)^2 / NM} \tag{18}$$

where $f(\cdot)$ denotes our prediction function, α scaling weights for the drug-side effect pairs (observed: 1; unobserved: 0.03) to complement the class imbalance, N the total number of drugs, and M the total number of side effects.

E. Adaboost framework

Drug-side effect frequency is determined by multiple different factors. We conjecture that different drugs have different input features of $\vec{\mathbf{x}}$, $\vec{\mathbf{w}}$, $\vec{\mathbf{z}}$, or $\vec{\mathbf{v}}$ that need to be focused in order to predict side effect frequency. However, we have limited numbers of available drugs, side effects, and proteins. To overcome the issue of overfitting and improve predictions for all drug groups, we utilize a boosting strategy inspired by Adaboost. Initially, we define the sampling weights of our training data as $W_1 = \{1/NM\}_{i=1}^{NM}$ where N is the number of drugs and M the number of side effects. Then, we train the base model $f_1 : x \rightarrow y$ for 50 epochs. To utilize Adaboost classifier, we interpret our base model as a binary classifier for now and determine the presence/absence of drug-side effects using the best threshold computed. In other words: $f_i(x),\ y_i \in \{0, 1\}$. For the total prediction error e after training, the influence of the classifier is defined

$\alpha_1 = \log((1-e)/e)/2$. Then, we update the sampling weights of the training data as following:

$$W_{i+1} = \begin{cases} \frac{1}{Z} e^{-\alpha_i} W_i, & \text{if} f_i(x) = y_i \\ \frac{1}{Z} e_i^{\alpha} W_i, & \text{otherwise} \end{cases} \quad (19)$$

Then we valuated sampling weight for each of drug-side effect and averaged the sampling weights per drugs. This way, we maximize the boosting effect and prevent overfitting to outlier data. We iteratively train a base model $f_i(x)$ using sample weights W_i for 50 epochs. The final model $F(x)$ after boosting 10 base models, $f_1(x), f_2(x), ..., f_{10}(x)$ is:

$$F(x) = \sum_{i=1}^{10} \alpha_i f_i(x) \quad (20)$$

We reused the similarity of drugs and side effect vectors at each boosting step. Generally, Adaboost assigns lower sampling weight for the data that is well predicted and focus more on the data failed to predict. To compensate the downlplayed data and guide to train other vectors, we utilize the vector similarity matrix to introduce the embedding loss. At the end of the training of a base model, drug-drug similarity matrix and side effect-side effect similarity matrix are defined. Then, we adopt embedding loss that minimizes the difference of next drug/side effect vectors similarity and the defined matrices. Contrary to the prediction loss, embedding loss is inverse proportional to the sampling weights. This way, the next base model can learn from previously successful embeddings. For drug vector d and side effect vector s, we define the embedding loss as following.

$$\begin{aligned} \mathcal{L}_{emb} = & \sum_{i,j} \left(\frac{d_i^T d_j}{\|d_i\| \|d_j\|} - Sim_D(i,j) \right) \omega_d^{(i)} \omega_d^{(j)} \\ & + \sum_{i,j} \left(\frac{s_i^T s_j}{\|s_i\| \|s_j\|} - Sim_S(i,j) \right) \omega_s^{(i)} \omega_s^{(j)} \end{aligned} \quad (21)$$

where Sim_D denotes drug-drug similarity matrix, Sim_S side effect-side effect similarity matrix, and $\omega_d^{(k)}$ or $\omega_s^{(k)}$ the maximal value of sampling weights minus the sampling weight of drug or side effect k. Finally, the final loss function is computed as

$$\mathcal{L}_{total} = \mathcal{L}_{rmse} + \lambda \mathcal{L}_{emb}. \quad (22)$$

IV. Result

A. Evaluation metrics

For comparison, we used exactly the same evaluation metrics used in the current state-of-the-art model, DSGAT [32]. For drug-side effect frequency prediction, we computed Spearman's correlation coefficient (SCC), root mean square error (RMSE), absolute mean error (MAE) as following:

$$\text{SCC} = \frac{Cov(R(f(x), R(y))}{\sigma_{R(f(x))} \sigma_{R(y)}} \quad (23)$$

$$\text{RMSE} = \sqrt{\sum (f(x) - y)^2 / NM} \quad (24)$$

$$\text{MAE} = \sum |f(x) - y| / NM \quad (25)$$

where $R(\cdot)$ denotes rank variables of the data.

For drug-side effect presence/absence prediction, we compute area under the receiver operating characteristic curve (AUROC) and mean average precision (mAP; also known as AUPRC) for association prediction

$$\text{AUROC} = \int_0^1 TPR(FPR^{-1}(x))dx \quad (26)$$

$$\text{mAP (AUPRC)} = \frac{1}{N} \sum_{d=1}^{N} TP_d / (FP_d + TP_d) \quad (27)$$

We also compute normalized discounted cumulative gain (nDCG), recall and precision of the top 1, 15 items with prediction score for recommendation. Here, we used the best threshold computed from the presence/absence prediction above, to determine the predicted labels for drug-side effects.

$$\text{nDCG@K} = Z_K \sum_{i=1}^{K} \frac{r(i)}{\log_2(i+1)} \quad (28)$$

$$\text{Precision} = \frac{TP}{N} \quad (29)$$

$$\text{Recall} = \frac{TP}{T} \quad (30)$$

where $r(i) \in \{0, 1\}$ is a relevance score, and Z_K the normalizer that nDCG@K estimation for the ideal possible recommendation would be 1.

B. Performance comparison with baseline models

In the process of drug development, we may not have any side effect reported for the tested drug prior to the examination. Therefore, it is desirable and necessary to test the drugs that are never seen to the model in the training process. To evaluate the prediction power for side effect frequencies of unseen drugs, we split drugs into training data and test data, so that drugs in the test data have not seen while training our model.

We assessed our model in drug split with tenfold cross-validation. First, we randomly divided 750 drugs into 10 bins of equal size. Then, we trained our model using 9 bins to test each remaining bin. We repeated and averaged the evaluation metrics over 10 times. To prevent data leakage, the columns of drug similarity vector that correspond to test drugs are masked to zero. To prevent overfitting, dropouts were included in all trainings. The supplementary material reports the hyperparameters used to train our model (Supplementary Table S2).

Table I shows the results of performance comparison. Our suggested model outperformed the existing models, and scored the best performance in 7 out of 10 of the evaluation metrics.

In Table I, the performances of MGPred [29] and SDPred [31] were not as good as the other models because they were initially designed for predicting the side effect frequencies of drugs that are already known. To compare with these models,

TABLE I: Performance comparison of our model with other SOTA baseline models. We split drugs into 10 bins, and performed tenfold cross validation using training and testing dataset as 9:1 ratio, to evaluate prediction for unseen drugs. MGPred, SDPred, DSGAT and our model were compared in ten different evaluation metrics. Galeano's model is matrix factorization-based method and is not applicable to predicting side effects of unseen drugs. Our model shows superior performance in overall. Bold-faced numbers indicate the best performance of each evaluation metric.

Metric	MGPred [29]	SDPred [31]	DSGAT [32]	Our model
SCC	-0.065	0.258	0.431	**0.438**
RMSE	3.435	3.649	1.470	**1.407**
MAE	3.314	3.539	1.174	**1.057**
AUROC	0.746	0.845	0.878	**0.901**
mAP (AUPRC)	0.178	0.347	0.403	**0.436**
nDCG@10	0.201	0.778	0.813	**0.858**
Precision@1	0.019	0.668	0.701	**0.750**
Precision@15	0.021	0.476	0.513	**0.556**
Recall@1	0.000	0.026	0.030	**0.031**
Recall@15	0.004	0.241	0.265	**0.267**

we designed additional tenfold cross-validation using different data split. Specifically, we divided 750 drugs and 994 side effects into 10 bins, ensuring an equal distribution of known associations between drugs and side effects. Then, we used tenfold cross-validation to test each bin. Supplementary Table S3 shows the results. Note that we use imbalanced dataset. MGPred and SDPred were not scalable, and do not perform as effective as they claimed in their original paper where only part of the negative data were selected to make balanced datasets. On the other hand, our model showed good performance, and was comparable to the state-of-the-art model, DSGAT [32].

We additionally carried out another experiment to assess the performance of our model when only data indicating presence are accessible. We trained the model with the same dataset, this time binarizing the label prior to the training phase so that all labels are either one of 0 or 1. We achieved equally high scores as we did with the original dataset. Specifically, we achieved AUROC and AUPRC of 0.905 and 0.422 with binarized labels, comparable to the original values of 0.903 and 0.431. The results are added to Supplementary Note S1.

In summary, our model achieved good performance in overall. Drug discovery process in real life often involves situation where side effect information for candidate drugs is not available. In this case, our model has superior performance in most evaluation metrics. Furthermore, cross-validaiton in traditional data split shows our model is robust to data imbalance compared to other methods such as Galeano et al. [5], MGPred [29], and SDPred [31].

C. Drug feature importance and ablation studies

To demonstrate the power of all drug features, we performed feature importance study and feature ablation study. First, for feature importance study, we used only one drug embedding to create drug vector and masked others. Other than the feature set, we used the same model architecture with the same parameters. Table II shows the result of feature importance study. Here, each of DG, DS, DT, and DF denotes drug molecular graph, similarity to other drugs, drug target information, and molecular fingerprints, respectively.

TABLE II: Feature importance studies of drug features. −DG, −DS, −DT, and −DF in the header row indicates ablation model with no molecular graph, drug similarity, drug protein target information, or molecular fingerprints, respectively. Bold-faced numbers indicate the best performance of each evaluation metric.

Metric	Full model	DG	DS	DT	DF
SCC	0.438	0.453	0.455	**0.459**	0.440
RMSE	1.407	1.549	1.406	1.415	**1.361**
MAE	1.057	1.223	1.078	1.106	**1.038**
AUROC	**0.901**	0.892	0.878	**0.901**	0.895
mAP (AUPRC)	**0.436**	0.417	0.401	0.433	0.421
nDCG@10	**0.858**	0.837	0.626	0.850	0.836
Precision@1	**0.750**	0.716	0.737	0.746	0.726
Precision@15	**0.556**	0.550	0.541	0.554	0.547
Recall@1	**0.031**	0.027	0.029	0.028	0.028
Recall@15	**0.267**	0.258	0.255	0.261	0.257

TABLE III: Ablation studies of drug features. −DG, −DS, −DT, and −DF in the header row indicates ablation model with no molecular graph, drug similarity, drug protein target information, or molecular fingerprints, respectively. Bold-faced numbers indicate the best performance of each evaluation metric.

Metric	Full model	-DG	-DS	-DT	-DF
SCC	0.438	0.418	0.411	**0.445**	0.429
RMSE	1.407	1.435	1.627	1.410	**1.373**
MAE	1.057	1.078	1.310	1.085	**1.049**
AUROC	**0.901**	0.897	0.893	0.897	0.892
mAP (AUPRC)	**0.436**	0.428	0.421	0.425	0.419
nDCG@10	**0.858**	0.851	0.846	0.852	0.843
Precision@1	**0.750**	0.736	0.708	0.748	0.740
Precision@15	**0.556**	0.546	0.555	0.550	0.548
Recall@1	**0.031**	0.029	0.027	0.030	0.029
Recall@15	**0.267**	0.259	0.276	0.259	0.257

We observed that drug target information has the highest score in SCC, and AUROC. It was also observed to perform competitively, in other metrics. This corresponds with our emphasis on drug target protein, while also highlighting the importance of utilizing all drug features to predict drug-side effect frequency.

Table III shows the result of ablation studies. Each of -DG, -DS, -DT, -DF denotes that either molecule graph, similarity to other drugs, drug target information, or molecular fingerprint embedding vector is masked. Experiments without any feature lead to a notable drop in performance predicting the frequencies and presence, in most evaluation metrics. Therefore, all the features in our model play role in determining side effect frequency of drugs.

D. Evaluating independent drugs

So far, we used 750 drugs from Galeano et al. [5] and MGPred [29] to compare the performance of our model

TABLE IV: Independent nine drugs performance. We trained our model on all 750 drugs and 994 side effects, and tested on nine novel drugs. The prediction of drug-side effect frequency for these drugs do not deteriorate with external data.

Drug name	SCC	RMSE	MAE	AUROC	mAP (AUPRC)
balsalazide	0.351	1.067	0.693	0.953	0.712
carboplatin	0.204	0.976	0.786	0.952	0.478
cisatracurium	-*	0.517	0.448	0.971	0.292
doxercalciferol	0.327	1.404	1.166	0.955	0.420
esomeprazole	-0.144	1.678	1.317	0.937	0.554
everolimus	0.500	1.676	1.370	0.871	0.682
fidaxomicin	0.577	0.769	0.536	0.976	0.486
gadoteridol	0.434	1.059	1.796	0.898	0.470
ixabepilone	0.391	1.110	0.858	0.958	0.740
Avg.	0.330	1.140	0.997	0.941	0.537

* Only one-class label to predict

with former drug-side effect prediction algorithms. Originally, Galeano et al. curated dataset of 759 drugs. MGPred screened out nine drugs whose STITCH compound IDs were not available. We identified the STITCH compound IDs of the nine drugs and further validated our model on the mentioned compounds (Supplementary Table S4). Specifically, we trained our model using the whole 750 drugs and then evaluated the scoring metrics of the nine newly introduced drugs. Table IV shows SCC, RMSE, MAE, AUROC, and mAP values for the nine independent drugs.

The AUROC and mAP values show that most of the nine drugs show excellent prediction power in determining drug-side effect association. The average value of AUROC and mAP are 0.941 and 0.537, respectively. These scores are significantly higher than the original performance of our model. RMSE and MAE also shows that our model excels in predicting the frequency in terms of error. SCC was not satisfactory compared to that of the original scheme. The reason is mainly attributed to esomeprazole and carboplatin with their SCC being -0.144 and 0.204, respectively. Nonetheless, the AUROC and mAP values of these drugs indicate that predictions of side effect presence are reliable enough for practical use. Overall, the result shows that our model, trained with the whole 750 drugs, performs well for an independent dataset.

E. Performance boosting of the Adaboost framework in terms of drug target specificity

We studied the impact of Adaboost on different drug groups. Since our motivation mainly focuses on using target information of drugs, we use drug groups with and without explicit protein targets. Explicit protein target means that the drug is meant to disrupt the targeted protein's biological pathway. Genomics of Drug Sensitivity in Cancer (GDSC) provides data on screened compounds with targeted pathways and processes [55] We classified cancer drugs into *cytotoxic drugs* (targeting DNA replication and cytoskeleton) and *targeted drugs* (targeting signaling pathways). In general, cytotoxic drugs are less investigated for off-targets than targeted drugs, thus we used the terminologies, cytotoxic vs. targeted in this context. We used area under receiver operating characteristic

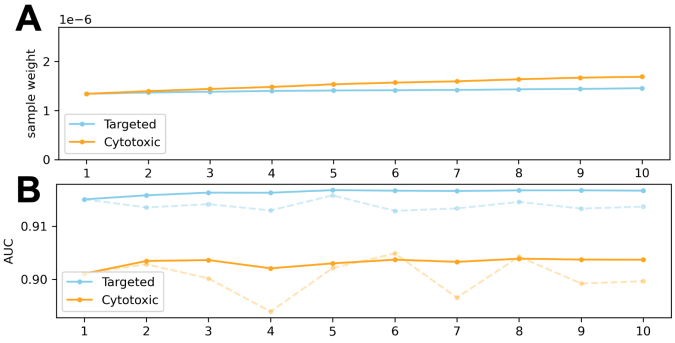

Fig. 3: Comparison of targeted (blue) and cytotoxic (orange) drugs. The x-axis indicates the ordinal number of base models in Adaboost. (A) Relative data sampling weights for targeted/cytotoxic drugs of the base model. (B) AUROC values of targeted/cytotoxic drugs of the base model. There are two lines for each group of drugs. Solid line denotes the AUROC values for 10 boosted models, and dashed line the AUROC values for 10 base models. Here, cytotoxic drugs greatly improves by Adaboost, compared to targeted drugs.

curve (AUROC) value to evaluate the scores for each group. The sampling weights and performance of two groups are shown in Figure 3. In Figure 3A, the sampling weights of both targeted and cytotoxic drugs increase with Adaboost, indicating that the selected cancer group is difficult predict their side effect values, compared to other type of drugs. Here, cytotoxic drug exhibits a larger increase. This shows that it is more difficult to train compared to the targeted ones. Figure 3B also shows the relative low AUROC score of cytotoxic drugs compared to the targeted drugs. However, as Adaboost complements with higher sampling weights, cytotoxic drugs show larger improvement than the targeted drugs. Specifically, while the average AUROC value of targeted drugs improves by 0.0017, that of cytotoxic drugs improves by 0.0026. To sum it up, while targeted drugs benefit from the target information given, cytotoxic drugs have low initial precision, which is complemented using Adaboost. Our Adaboost strategy successfully integrates to construct the final model that is competent to predict drug-side effect frequency for both targeted and cytotoxic drugs.

F. Common embedding space for drugs and side effects

Our model creates drug vectors and side effect vectors of the same dimension and magnitude (Equation 15, 16), inner products the two vectors and scale the resulting output to predict their frequency (Equation 17). Therefore, the drug vectors and side effect vectors are projected into the common spherical space with their cosine similarity being proportional to the frequency of side effects, forming a drug-side effect sphere like the constellation. In such space, the geography of drugs and side effects can be directly linked to reveal their frequency. We performed principal component analysis (PCA) to reduce the dimension and visualize the distribution of drugs and side effects.

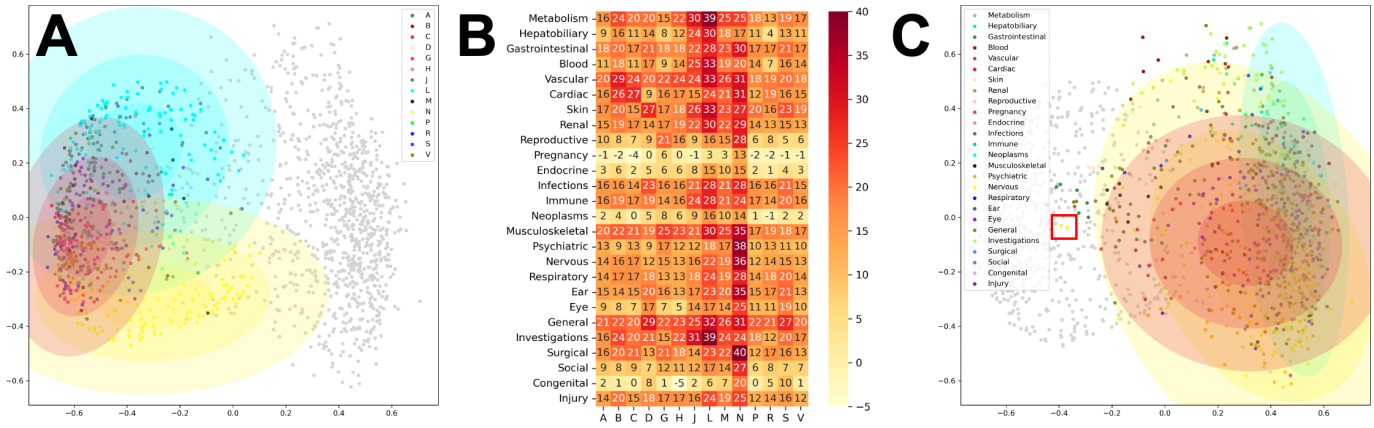

Fig. 4: Drugs and side effects projected in common vector space. The drug vectors and side effect vectors are reduced to 2-dimensional space using principal component analysis (PCA). (A) Drug vectors are represented in colored dots with their color indicating ATC code of drugs. Drugs for [C: cardiovascular system], [L: antineoplastic and immunomodulating agents], and [N: nervous system] are drawn with ellipses indicating the deviation of each class, respectively. Side effects are represented in gray marks. (B) Heatmap of average percentile cosine similarity among drugs and side effects in perspective of ATC code and `System Organ Class`. (C) Side effect vectors are represented in colored dots with their color indicating `System Organ Class` of side effects. Side effects of Cardiac, Neoplasm and Nervous system class are drawn with ellipses indicating the deviation of each class, respectively. Drugs are represented in gray marks.

Figure 4 displays drugs and side effects in a common space, with colors indicating corresponding ATC codes (Anatomical Therapeutic Chemical Classification System) and `System Organ Class`. We examine three pairs of drug and side effect groups, [C: cardiovascular-cardiac] (red), [L: antineoplastic-neoplasm] (blue), and [N: nervous-nervous] (yellow), and notate deviations with ellipses. Figure 4A and 4C show the same space in the same scale, with one focusing on drugs and the other on side effects.

While we clearly observe the clustering of drugs, the coverage of each group of drugs are explained by the cosine similarity among different classes, as in Figure 4B. Note that drug points are mostly located in the left-hand side of the space while side effect points are in the right-hand side. Cardiovascular drugs occupy the farthest side on the left, with its cosine similarity to side effects being one of the lowest among the drug groups. However, antineoplastic drugs and nervous system drugs are placed in center close to side effects. Their similarities to side effects are the largest, coherent to the reality. While an average drug gets 49.4 side effects, cardiovascular/antineoplastic/nervous system drugs get 28.1/56.4/110.8 side effects, respectively. Moreover, we observe that antineoplastic drugs have lower risk of developing side effects of neoplasm than developing side effects of other `System Organ Class`, while nervous system drugs on the contrary, often cause nervous system side effects. We observed the nervous system side effects that reside leftmost, as shown in the red box of Figure 4C. The names of the side effects and the true probability of a nervous system drug accompanying the side effects are: somnolence (0.80), tremor (0.65). The average probability for other nervous system side effects was 0.15. The side effect points that are closer to the nervous system drug cluster on the embedding space indeed showed more frequent occurrence.

V. Conclusion

Predicting drug-side effect frequency *in silico* is a major problem for drug discovery. Neglecting side effects of drugs can jeopardize the public health and incur social costs, even lives of people. On the other hand, knowledge of drug side effects can lead to understanding the benefit and risk of drugs. However, the task of predicting drug-side effect frequency is not a trivial task, and only four existing studies have covered the topic.

In this paper, we proposed a deep learning-based drug-side effect frequency prediction model. While we inherited important features such as molecular graph and drug-drug similarity from the previous studies, we also introduced drug protein target and molecular fingerprint to complement the model. We adopted GAT and network propagation to further extract latent features of given input, and finally aggregated four different embeddings to produce final vector representation for drugs. Our model achieved the best performance over the existing drug-side effect frequency prediction models. We are the first to utilize the direct drug protein target information to represent drug vectors for predicting frequencies of side effects. Moreover, Adaboost ensemble technique successfully integrated the utility of the heterogeneous features of drugs. We performed ablation studies to show that all components in our model contribute to good prediction.

The features of our model are limited to ingredients of the 4 drug embeddings, and exclude other information such as drug-induced gene expression of cell lines. Integrating biological data is another challenge to improve drug-side effect frequency prediction. We will pursue this topic as a future study.

References

[1] N. Vargesson, "Thalidomide-induced teratogenesis: History and mechanisms," *Birth Defects Research Part C: Embryo Today: Reviews*, vol. 105, no. 2, pp. 140–156, 2015.

[2] M. Hacker, "Adverse drug reactions," in *Pharmacology*. Elsevier, 2009, pp. 327–352.

[3] J. Sultana, P. Cutroneo, and G. Trifirò, "Clinical and economic burden of adverse drug reactions," *Journal of Pharmacology and Pharmacotherapeutics*, vol. 4, no. 1_suppl, pp. S73–S77, 2013.

[4] H. Le Louët and P. J. Pitts, "Twenty-first century global adr management: A need for clarification, redesign, and coordinated action," *Therapeutic Innovation & Regulatory Science*, pp. 1–4, 2022.

[5] D. Galeano, S. Li, M. Gerstein, and A. Paccanaro, "Predicting the frequencies of drug side effects," *Nature communications*, vol. 11, no. 1, pp. 1–14, 2020.

[6] R. K. Harrison, "Phase ii and phase iii failures: 2013–2015," *Nat Rev Drug Discov*, vol. 15, no. 12, pp. 817–818, 2016.

[7] I. J. Onakpoya, C. J. Heneghan, and J. K. Aronson, "Delays in the post-marketing withdrawal of drugs to which deaths have been attributed: a systematic investigation and analysis," *BMC medicine*, vol. 13, no. 1, pp. 1–11, 2015.

[8] B. Munos, "Lessons from 60 years of pharmaceutical innovation," *Nature reviews Drug discovery*, vol. 8, no. 12, pp. 959–968, 2009.

[9] P. Watkins, "Drug safety sciences and the bottleneck in drug development," *Clinical Pharmacology & Therapeutics*, vol. 89, no. 6, pp. 788–790, 2011.

[10] Y. Masoudi-Sobhanzadeh, Y. Omidi, M. Amanlou, and A. Masoudi-Nejad, "Drug databases and their contributions to drug repurposing," *Genomics*, vol. 112, no. 2, pp. 1087–1095, 2020.

[11] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nature reviews Drug discovery*, vol. 3, no. 8, pp. 673–683, 2004.

[12] V. Parvathaneni, N. S. Kulkarni, A. Muth, and V. Gupta, "Drug repurposing: a promising tool to accelerate the drug discovery process," *Drug discovery today*, vol. 24, no. 10, pp. 2076–2085, 2019.

[13] S. M. Corsello, J. A. Bittker, Z. Liu, J. Gould, P. McCarren, J. E. Hirschman, S. E. Johnston, A. Vrcic, B. Wong, M. Khan, *et al.*, "The drug repurposing hub: a next-generation drug library and information resource," *Nature medicine*, vol. 23, no. 4, pp. 405–408, 2017.

[14] M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, and P. Bork, "A side effect resource to capture phenotypic effects of drugs," *Molecular systems biology*, vol. 6, no. 1, p. 343, 2010.

[15] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, "The sider database of drugs and side effects," *Nucleic acids research*, vol. 44, no. D1, pp. D1075–D1079, 2016.

[16] X. Xia, "Bioinformatics and drug discovery," *Current topics in medicinal chemistry*, vol. 17, no. 15, pp. 1709–1726, 2017.

[17] W. Zhang, Y. Chen, S. Tu, F. Liu, and Q. Qu, "Drug side effect prediction through linear neighborhoods and multiple data source integration," in *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 2016, pp. 427–434.

[18] G. M. Dimitri and P. Lió, "Drugclust: a machine learning approach for drugs side effects prediction," *Computational biology and chemistry*, vol. 68, pp. 204–210, 2017.

[19] S. Shabani-Mashcool, S.-A. Marashi, and S. Gharaghani, "Nddsa: A network-and domain-based method for predicting drug-side effect associations," *Information Processing & Management*, vol. 57, no. 6, p. 102357, 2020.

[20] A. Cakir, M. Tuncer, H. Taymaz-Nikerel, and O. Ulucan, "Side effect prediction based on drug-induced gene expression profiles and random forest with iterative feature selection," *The Pharmacogenomics Journal*, vol. 21, no. 6, pp. 673–681, 2021.

[21] C. Y. Lee and Y.-P. P. Chen, "Descriptive prediction of drug side-effects using a hybrid deep learning model," *International Journal of Intelligent Systems*, vol. 36, no. 6, pp. 2491–2510, 2021.

[22] Y. Qian, Y. Ding, Q. Zou, and F. Guo, "Identification of drug-side effect association via restricted boltzmann machines with penalized term," *Briefings in Bioinformatics*, vol. 23, no. 6, p. bbac458, 2022.

[23] W. Zhang, H. Zou, L. Luo, Q. Liu, W. Wu, and W. Xiao, "Predicting potential side effects of drugs by recommender methods and ensemble learning," *Neurocomputing*, vol. 173, pp. 979–987, 2016.

[24] W. Zhang, X. Liu, Y. Chen, W. Wu, W. Wang, and X. Li, "Feature-derived graph regularized matrix factorization for predicting drug side effects," *Neurocomputing*, vol. 287, pp. 154–162, 2018.

[25] O. C. Uner, R. G. Cinbis, O. Tastan, and A. E. Cicek, "Deepside: a deep learning framework for drug side effect prediction," *Biorxiv*, p. 843029, 2019.

[26] A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, J. Gould, J. F. Davis, A. A. Tubelli, J. K. Asiedu, *et al.*, "A next generation connectivity map: L1000 platform and the first 1,000,000 profiles," *Cell*, vol. 171, no. 6, pp. 1437–1452, 2017.

[27] D. Galeano and A. Paccanaro, "Machine learning prediction of side effects for drugs in clinical trials," *Cell Reports Methods*, vol. 2, no. 12, 2022.

[28] X. Liang, Y. Fu, L. Qu, P. Zhang, and Y. Chen, "Prediction of drug side effects with transductive matrix co-completion," *Bioinformatics*, vol. 39, no. 1, p. btad006, 2023.

[29] H. Zhao, K. Zheng, Y. Li, and J. Wang, "A novel graph attention model for predicting frequencies of drug–side effects from multi-view data," *Briefings in Bioinformatics*, vol. 22, no. 6, p. bbab239, 2021.

[30] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *stat*, vol. 1050, p. 20, 2017.

[31] H. Zhao, S. Wang, K. Zheng, Q. Zhao, F. Zhu, and J. Wang, "A similarity-based deep learning approach for determining the frequencies of drug side effects," *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab449, 2022.

[32] X. Xu, L. Yue, B. Li, Y. Liu, Y. Wang, W. Zhang, and L. Wang, "Dsgat: predicting frequencies of drug side effects by graph attention networks," *Briefings in Bioinformatics*, vol. 23, no. 2, p. bbab586, 2022.

[33] S. I. Berger and R. Iyengar, "Role of systems pharmacology in understanding drug adverse events," *Wiley interdisciplinary reviews: systems biology and medicine*, vol. 3, no. 2, pp. 129–135, 2011.

[34] M. Campillos, M. Kuhn, A.-C. Gavin, L. J. Jensen, and P. Bork, "Drug target identification using side-effect similarity," *Science*, vol. 321, no. 5886, pp. 263–266, 2008.

[35] Y. Yamanishi, E. Pauwels, and M. Kotera, "Drug side-effect prediction based on the integration of chemical and biological spaces," *Journal of chemical information and modeling*, vol. 52, no. 12, pp. 3284–3292, 2012.

[36] Z. Isik, C. Baldow, C. V. Cannistraci, and M. Schroeder, "Drug target prioritization by perturbed gene expression and network information," *Scientific reports*, vol. 5, no. 1, p. 17417, 2015.

[37] Y. Ding, J. Tang, and F. Guo, "Identification of drug-side effect association via multiple information integration with centered kernel alignment," *Neurocomputing*, vol. 325, pp. 211–224, 2019.

[38] H. Zhou, M. Gao, and J. Skolnick, "Comprehensive prediction of drug-protein interactions and side effects for the human proteome," *Scientific reports*, vol. 5, no. 1, pp. 1–13, 2015.

[39] M. Kuhn, M. Al Banchaabouchi, M. Campillos, L. J. Jensen, C. Gross, A.-C. Gavin, and P. Bork, "Systematic identification of proteins that elicit drug side effects," *Molecular systems biology*, vol. 9, no. 1, p. 663, 2013.

[40] E. Lounkine, M. J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J. L. Jenkins, P. Lavan, E. Weber, A. K. Doak, S. Côté, *et al.*, "Large-scale prediction and testing of drug activity on side-effect targets," *Nature*, vol. 486, no. 7403, pp. 361–367, 2012.

[41] S. Mizutani, E. Pauwels, V. Stoven, S. Goto, and Y. Yamanishi, "Relating drug–protein interaction network with drug side effects," *Bioinformatics*, vol. 28, no. 18, pp. i522–i528, 2012.

[42] S. Bang, J. H. Jhee, and H. Shin, "Polypharmacy side-effect prediction with enhanced interpretability based on graph feature attention network," *Bioinformatics*, vol. 37, no. 18, pp. 2955–2962, 2021.

[43] X.-P. Chen and G.-H. Du, "Target validation: A door to drug discovery." *Drug discoveries & therapeutics*, vol. 1, no. 1, pp. 23–29, 2007.

[44] R. A. Blake, "Target validation in drug discovery," in *High Content Screening*. Springer, 2007, pp. 367–377.

[45] 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, *et al.*, "Drugbank 5.0: a major update to the drugbank database for 2018," *Nucleic acids research*, vol. 46, no. D1, pp. D1074–D1082, 2018.

[46] M. Pak, D. Jeong, J. H. Moon, H. Ann, B. Hur, S. Lee, and S. Kim, "Network propagation for the analysis of multi-omics data," in *Recent Advances in Biological Network Analysis*. Springer, 2021, pp. 185–217.

[47] M. Kuhn, C. von Mering, M. Campillos, L. J. Jensen, and P. Bork, "Stitch: interaction networks of chemicals and proteins," *Nucleic acids research*, vol. 36, no. suppl_1, pp. D684–D688, 2007.

[48] C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, and E. Willighagen, "The chemistry development kit (cdk): An open-source java library for chemo-and bioinformatics," *Journal of chemical information and computer sciences*, vol. 43, no. 2, pp. 493–500, 2003.

[49] H. Cai, H. Zhang, D. Zhao, J. Wu, and L. Wang, "Fp-gnn: a versatile deep learning architecture for enhanced molecular property prediction," *Briefings in bioinformatics*, vol. 23, no. 6, p. bbac408, 2022.

[50] I. Große-Michaelis, S. Proestel, R. M. Rao, B. S. Dillman, S. Bader-Weder, L. Macdonald, and W. Gregory, "Meddra labeling groupings to improve safety communication in product labels," *Therapeutic Innovation & Regulatory Science*, pp. 1–6, 2022.

[51] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[52] M. Pak, S. Lee, I. Sung, B. Koo, and S. Kim, "Improved drug response prediction by drug target data integration via network-based profiling," *Briefings in Bioinformatics*, p. bbad034, 2023.

[53] D. Szklarczyk, A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, *et al.*, "The string database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets," *Nucleic acids research*, vol. 49, no. D1, pp. D605–D612, 2021.

[54] M. Gillespie, B. Jassal, R. Stephan, M. Milacic, K. Rothfels, A. Senff-Ribeiro, J. Griss, C. Sevilla, L. Matthews, C. Gong, *et al.*, "The reactome pathway knowledgebase 2022," *Nucleic acids research*, vol. 50, no. D1, pp. D687–D692, 2022.

[55] W. Yang, J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J. A. Smith, I. R. Thompson, *et al.*, "Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells," *Nucleic acids research*, vol. 41, no. D1, pp. D955–D961, 2012.