Deep Learning for High-Order Drug-Drug Interaction Prediction

Bo Peng
peng.707@buckeyemail.osu.edu
The Ohio State University
Columbus, Ohio

Xia Ning*
ning.104@osu.edu
The Ohio State University
Columbus, Ohio

ABSTRACT

Drug-drug interactions (DDIs) and their associated adverse drug reactions (ADRs) represent a significant detriment to the public health. Existing research on DDIs is primarily focused on pairwise DDI detection and prediction. It is highly needed to develop effective computational tools for high-order DDI prediction. Here we show that deep learning can be effectively utilized to predict ADRs induced from high-order DDIs. In this manuscript, we present a deep learning model D^3I for cardinality-invariant and order-invariant high-order DDI prediction. The D^3I models achieve 0.740 F1 value and 0.847 AUC value on balanced high-order DDI prediction, and outperform other models on order-2 DDI prediction. These results demonstrate the strong potential of D^3I and deep learning models in tackling the prediction problems of high-order DDIs and their induced ADRs. In addition, D^3I is able to derive single drug representations, which conform to our current knowledge on single drugs, from their behaviors in drug combinations. D^3I can also correctly predict ADRs for drug combinations in which no single drugs on their own induce ADRs, and improve ADR prediction on drug pairs by learning from all drug combinations. To the best of our knowledge, D^3I is the first deep model for high-order DDI prediction.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Applied computing** → **Health informatics**; • **Information systems** → **Data mining**.

KEYWORDS

High-order drug-drug interactions; adverse drug reactions; deep learning

ACM Reference Format:
Bo Peng and Xia Ning. 2019. Deep Learning for High-Order Drug-Drug Interaction Prediction. In *10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB '19), September 7–10, 2019, Niagara Falls, NY, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3307339.3342136

ACM-BCB '19, September 7–10, 2019, Niagara Falls, NY, USA

ACM ISBN 978-1-4503-6666-3/19/09…$15.00
https://doi.org/10.1145/3307339.3342136

1 INTRODUCTION

Drug-drug interactions (DDIs) and their associated adverse drug reactions (ADRs) represent a significant detriment to the public health. Approximately 195,000 hospitalizations and 74,000 emergency room visits are resulted out of DDIs in the United States [1]. The increasing rates of polypharmacy, particularly among aging population [1], will further deteriorate this situation [24]. Consequent upon these facts, significant research efforts have been dedicated to detecting DDIs, including DDI extraction from medical literature [11, 18] or social media [16, 32, 39], and biochemical and molecular information integration for DDI scoring [5, 5, 22, 29, 33], etc.

However, most of the existing DDI studies are limited to interactions between pairs of drugs (i.e., order-2 DDIs), while DDIs among multiple drugs (i.e., high-order DDIs) occupy a significant portion in real-life cases. It is reported that more than 76% of the elderly Americans take two or more drugs daily [1]. Another study [18] estimates that about 29.4% of elderly American patients taking six or more drugs every day. Therefore, understanding high-order DDIs and their associated ADRs becomes urgent and critical [12, 13, 24].

Unfortunately, very limited efforts, to the best of our knowledge, have been dedicated to representing, quantifying, discovering and visualizing relations among high-order DDIs. Emerging methods on high-order DDI studies are only focused on the discovery of high-order DDIs through mining frequent drug combinations efficiently. Meanwhile, as the cardinality of drug combinations (i.e., the number of drugs in drug combinations; also refereed to as the order of drug combinations) increases, modeling of DDI relations, particularly of arbitrary cardinalities/orders in a unified framework, becomes increasingly difficult.

In this manuscript, we present a new deep model to conduct cardinality- and order-invariant high-order DDI prediction, refereed to as **D**eep **DDI** model and denoted as D^3I. D^3I is invariant of drug combination cardinalities and the order in which the drugs are considered in the model, that is, D^3I is able to predict ADR labels for combinations of arbitrary numbers of drugs in arbitrary input orders. Meanwhile, D^3I is able to generate embeddings for single drugs and aggregate single drug embeddings into drug-combination embeddings. Thus, these drug-combination embeddings are able to capture the synergistic latent signals that are related to ADRs among the constituent single drugs. We conducted extensive experiments on two public datasets of high-order DDIs, and tested multiple D^3I variations on the datasets. Our experimental results demonstrate that D^3I is able to achieve 0.740 F1 value and 0.847 AUC value on balanced high-order DDI prediction, and outperform other models on order-2 DDI prediction. The experiments also show that by integrating DDIs of high orders, D^3I models are even able

to further improve prediction performance on order-2 DDIs. In addition, the single drug embeddings produced from D^3I models also represent clustering structures that conform to domain knowledge.

To the best of our knowledge, D^3I is the first deep model for high-order DDI prediction. High-order DDI prediction is particularly non-trivial compared to high-order DDI detection due to several very unique facts. First of all, it could be one or more drugs, but not necessarily all the drugs, in a drug combination, that interact and induce ADRs. However, it is often unknown which drugs in the drug combination induce the ADRs, or their machenisms. Moreover, those ADR-inducing drugs in the combination could be safe on their own if taken alone. These facts together unstantially increase the difficulty for predicting high-order DDIs, and demand strong prediction power for such prediction problems. Our D^3I and its experimental results demonstrate promising progress toward accurate high-order DDI prediction.

The rest of this manuscript is organized as follows. Section 2 presents the literature review. Section 3 presents the definitions and notations used in this manuscript. Section 4 presents the D^3I method. Section 5 presents the datasets used for the experiments. Section 6 presents the experimental protocol. Section 7 presents the experimental results. Section 8 presents conclusions and future research.

2 LITERATURE REVIEW

2.1 DDI Detection and Prediction

Current research on detecting DDIs can be broadly classified into four categories [8, 24, 33, 37]. The first category of methods focus on text mining from medical literature and electronic medical records, and they extract mentioned drug pairs [11, 18, 20, 38, 42]. A second category of methods integrate various biochemical and molecular drug/target data to measure drug-drug similarities and score/predict DDIs using the similarities. These data include chemical structures [10, 22], target information [33, 43], compound-target docking results [15], phenotypic and genomic information [5], and drug side effects [29], etc. The collected data are used in various data-driven computational methods such as classification [5], regression [29], statistical testing [21] to detect DDIs. For example, Zhang *et al.* [41] applies multiple methods such as neighbor-based recommendation, random walk and matrix perturbation for pairwise DDI ranking and prediction. The third category of methods leverage healthcare information on social media and online communities to detect DDIs that have been mentioned/inferred in online discussions and posts [16, 32, 39]. The last category of methods predict the probability of ADR event counts due to high-order DDIs [6, 36] and use either electronic medical records or pharmacokinetic modeling to validate potential DDIs. A notable shortcoming of these methods is that they work for low-order or fixed-order DDIs but do not scale well to arbitrary orders. A very related research area is on combinatorial drug therapy against cancer and infection [3], in which numerical models are built to predict dose responses to multiple drugs [30, 44] in order to quantify efficacy of multiple drugs admitted together for disease treatment. Therefore, drug efficacy rather than ADR is more emphasized, and *in vitro* or *in vivo* experiments are typically conducted to validate the numerical predictions. These methods also usually suffer from scalability to arbitrary number of comediated drugs.

2.2 Deep Learning based DDI Detection and Prediction

The interactions between drugs are very complex and may go far beyond simple or linear relations. Thus, it inspires the use of Deep Learning (DL) in this field due to the strong capability of DL in approximating complex relations. High-order DDIs prediction has some analogies to multi-instance learning [35] over bags of instances. Wang *et al.* [35] proposed a deep framework for multi-instance learning, which first learns an embedding for each of the instances in the bag, and then applies an aggregator to combine these embeddings into a bag-level representation for classification. llse *et al.* [17] proposed an attention-based deep model to integrate instance embeddings into bag embeddings. One drawback of this method is that it combines instances linearly, which might not always be optimal. Zaheer *et al.* [40] introduced constraints on the weight matrix of the deep model to learn over sets, and enforced symmetry of the learned weight matrix to enable order-invariant property into the model. Wang *et al.* [34] incorporated different types of drug features to learn a drug embedding for single drugs, and used a deep neural network architecture to predict potential side-effects of single drugs.

Deep learning technologies are also used in detecting and predicting DDIs. Segura-Bedmar *et al.* [27] proposed to use convolutional neural networks (CNNs) to extract DDIs from biomedical text. Text information is represented as a matrix, in which each column or row is a word vector [23]. Then CNN layers are applied to the matrix to extract features and do the prediction. This work achieves the second place in the 2013 ranking of the DDIs extraction challenge. In Sahu *et al.* [26], instead of using CNNs, Long Short-Term Memory (LSTM) model is used to extract features from text and then do the prediction. Graph Convolutional neural Network (GCN) is also introduced to predict pairwise DDIs. Zitnik *et al.* [45] views pairwise DDI prediction as a link prediction task over drug-drug graphs. They applied GCN on the constructed DDI graph to learn embeddings for each drug, and calculated link probabilities (i.e., DDI probabilities) based on learned embeddings.

3 DEFINITIONS AND NOTATIONS

Table 1: Notations

notation	meaning
d	a drug
D	a drug combination
$\mathbf{f}$	a vector of drug features
$\mathbf{e}$	a vector of drug embedding
$\mathbf{E}$	a vector of drug combination embedding

The key notations used in this manuscript are listed in Table 1. In this manuscript, all the vectors are by default row vectors and represented using lower-case bold letters (e.g., $\mathbf{e}$); all the matrices are represented using upper-cases letters (e.g., X). The key definitions used in this manuscript are listed as follows:

- *Drug combination*: a set of drugs that are prescribed/taken together, denoted as D.
- *Cardinality of drug combinations*: the number of drugs in a drug combination D is the cardinality of the drug combination, denoted as $\|D\|$. Drug combination cardinality is also refereed to as drug order of the combination.
- *Cardinality invariance*: the model is able to predict for drug combinations of arbitrary cardinalities; the prediction mechanism is invariant of the cardinalities of input drug combinations.
- *Order invariance*: the model is able to produce a same result for a same drug combination, regardless of the order in which the drugs in the combination are input to the model. Note that in order invariance, the term "order" does not refer to cardinality but to a notion of ordering.

The problem that we try to solve is defined as follows:

Problem definition: Given a set of drug combinations and their ADR labels, build a classification model of cardinality invariance and order invariance that is able to predict the ADR labels for new drug combinations of arbitrary cardinalities.

In this manuscript, we are only concerned with one specific ADR, that is, myopathy [4]. Therefore, the classification model is a binary classifier. However, multi-class classifier can be extended from our models, and will be investigated in our future research.

4 METHODS

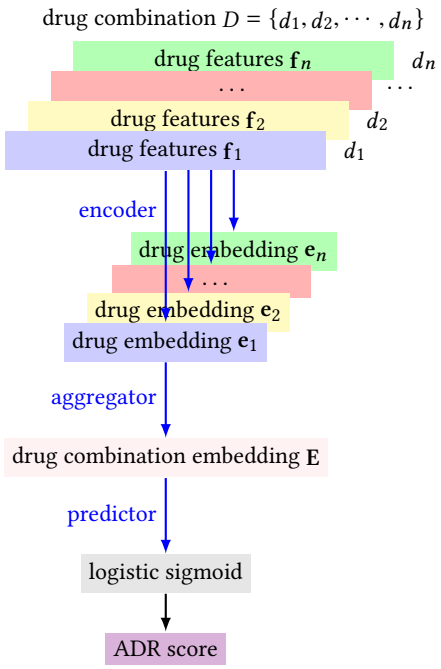

Figure 1: D^{3}I Architecture

We develop a new deep model to conduct cardinality- and order-invariant high-order DDI prediction. This model is refereed to as **D**eep **DDI** model and denoted as D^{3}I. D^{3}I has the following three key components:

- An encoder, which encodes each of the drugs in an input drug combination into a latent representation (i.e., embedding);
- An aggregator, which learns a single, high-level representation/embedding for the drug combination from the representations/embeddings of its component drugs; and
- A predictor, which predicts the likelihood of ADR labels using the drug-combination representation/embedding.

Figure 1 presents the architecture of D^{3}I. The novelty of D^{3}I is that its aggregator is able to deal with arbitrary number of drug embeddings in drug combinations regardless of drug input orders. Meanwhile, the single drug embeddings and the drug-combination embeddings could enable additional insights on the drug properties and relations in inducing ADRs.

Note that in this manuscript, only myopathy is considered as the ADR of interest. That is, we predict if a drug combination will induce myopathy or not. The reason for myopathy as the interested ADR is that it has been better studied [9] than other side effects, particularly in terms of the underlying mechanisms and the ground-truth myopathy-inducing single drugs. Even though, our D^{3}I is effortlessly applicable to other specific, single ADRs, and can be easily extended to the prediction of multiple, specific ADRs (by learning multiple outputs) and to the prediction of general ADRs (i.e., whether there will be ADRs or not; not specific to a certain type of ADR).

4.1 D^{3}I Encoder

The D^{3}I encoder learns and represents signals that could be pertinent to ADR prediction from each drug in the input drug combination. For a drug combination $D = \{d_1, d_2, \cdots, d_n\}$ of n drugs, the encoder g_e learns an embedding $\mathbf{e}_i$ for each drug d_i from its feature vector $\mathbf{f}_i$ as follows:

$$\mathbf{e}_i = g_e(\mathbf{f}_i), \tag{1}$$

where $\mathbf{e}_i \in \mathbb{R}^{1\times k}$, $\mathbf{f}_i \in \mathbb{R}^{1\times m}$ and typically $k < m$. We use an n_e-layer neural network (NN) as g_e, that is,

$$g_e(\mathbf{f}) = g_{n_e}(\cdots(g_2(g_1(\mathbf{f})))), \tag{2}$$

with each layer parameterized by a weighting matrix W_j^e ($j = 1, \cdots, n_e$) of appropriate dimensions. The input drug features will be discussed later in Section 5. Note that the encoder applies on each individual drug in the input drug combination independently, and thus it is order invariant. For input $D = \{d_1, d_2, \cdots, d_n\}$, the output from the encoder is denoted as $\mathbf{e}(D) = [\mathbf{e}_1; \mathbf{e}_2; \cdots; \mathbf{e}_n]$, that is, $\mathbf{e}(D)$ is an $n \times k$ matrix.

4.2 D^{3}I Aggregator

The D^{3}I aggregator learns one embedding for the input drug combination from its individual drug embeddings out of the encoder. We adopt three aggregation strategies: 1). max pooling, 2). mean pooling and 3). aggregation with attentions, respectively, in the D^{3}I aggregator.

4.2.1 Max Pooling. In the max pooling strategy, we calculate the drug-combination embedding, denoted as $\mathbf{E}_D$, for $D = \{d_1, d_2, \cdots, d_n\}$ as follows:

$$\mathbf{E}_D = \max(\mathbf{e}(D)) = [\max_{\forall i}\{\mathbf{e}_{i,1}\}, \max_{\forall i}\{\mathbf{e}_{i,2}\}, \cdots, \max_{\forall i}\{\mathbf{e}_{i,k}\}], \tag{3}$$

where max is an element-wise operator that selects the maximum value in each dimension in all the drug embeddings $\mathbf{e}(D) = [\mathbf{e}_1; \mathbf{e}_2; \cdots; \mathbf{e}_n]$. The max pooling is trivially cardinality invariant and order invariant due to the max function used. It is expected that drugs contribute differently in their interactions and induced ADRs, and their respective contributions could be represented in their maximum values in their embeddings through learning and the max pooling. $\mathtt{D^3I}$ with the max pooling strategy is denoted as $\mathtt{D^3I}_{\mathtt{max}}$.

4.2.2 Mean Pooling. In the mean pooling strategy, we calculate the drug-combination embedding $\mathbf{E}_D$ as follows:

$$\mathbf{E}_D = \text{mean}(\mathbf{e}(D)) = [\text{avg}_{\forall i}\{\mathbf{e}_{i,1}\}, \text{avg}_{\forall i}\{\mathbf{e}_{i,2}\}, \cdots, \text{avg}_{\forall i}\{\mathbf{e}_{i,k}\}], \tag{4}$$

where the avg operator calculates the average value in each dimension in all the drug embeddings $\mathbf{e}(D) = [\mathbf{e}_1; \mathbf{e}_2; \cdots; \mathbf{e}_n]$. The mean pooling is also trivially cardinality invariant and order invariant. It intends to average the information from each involved drug in representing a drug combination. $\mathtt{D^3I}$ with the mean pool strategy is denoted as $\mathtt{D^3I}_{\mathtt{mean}}$.

4.2.3 Self-Attention. Inspired by the recent work in deep multi-instance learning [17], we propose to use a weighted sum of drug embeddings to learn a single embedding of a drug combination. For a drug combination $D = \{d_1, d_2, \cdots, d_n\}$, the embedding of D is calculated as follows:

$$\mathbf{E}_D = \sum_{i=1}^{n} a_i \mathbf{e}_i, \tag{5}$$

where a_i is a weight on $\mathbf{e}_i$. To allow the drug embeddings to determine their own importance in the drug-combination embedding, a_i is also calculated as a function of $\mathbf{e}_i$ as follows,

$$a_i = \text{softmax}(\mathbf{w} \tanh(V\mathbf{e}_i^\mathsf{T})),$$

where $V \in \mathbb{R}^{l \times k}$ and $\mathbf{w} \in \mathbb{R}^{1 \times l}$ are two parameters that will be learned, and softmax(x) is the softmax function defined as follows:

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}, \tag{6}$$

and thus $\sum_i a_i = \sum_i(\text{softmax}(\mathbf{w}^\mathsf{T} \tanh(V\mathbf{e}_i^\mathsf{T}))) = 1$; and the hyperbolic tangent function $\tanh(\cdot)$ is used to introduce element-wise non-linearity. The attention mechanism as in Equation 5 is order invariant simply because the sum operation in Equation 5 is order invariant. It is also cardinality invariant because of the normalization in softmax in Equation 6. $\mathtt{D^3I}$ with the self-attention pooling strategy is denoted as $\mathtt{D^3I}_{\mathtt{Att}}$.

4.3 $\mathtt{D^3I}$ Predictor

The $\mathtt{D^3I}$ predictor predicts the probability of a drug combination in inducing ADRs. For a drug combination D, its embedding $\mathbf{E}_D$ is first converted through n_p fully-connected layers with tanh as the activation function, that is,

$$h_e(\mathbf{E}_D) = h_{n_p}(\cdots(h_2(h_1(\mathbf{E}_D)))), \tag{7}$$

with each layer parameterized by a weighting matrix W_j^E ($j = 1, \cdots, n_p$) of appropriate dimensions. Then a sigmoid function is used to do the prediction as follows:

$$p(D) = \frac{1}{1 + \exp(-h_e(\mathbf{E}_D)^\mathsf{T})} \tag{8}$$

where $\mathbf{E}_D$ is the drug-combination embedding of D out of $\mathtt{D^3I}$ aggregator, and $p(D)$ is the probability of D in inducing ADRs ($p(D) \in [0, 1]$).

4.4 Learning Algorithm

In $\mathtt{D^3I}$, we formulate the DDI-induced ADR prediction as a binary classification problem, and learn the $\mathtt{D^3I}$ models by solving the following optimization problem, in which the cross entropy loss is used as the objective:

$$\min_{\mathbf{\Theta}} -\sum_{i=1}^{n} y_i \log p_i + (1 - y_i) \log(1 - p_i), \tag{9}$$

where y_i is the label of the i-th drug combination (positive for ADR inducing and negative otherwise), p_i as calculated in Equation 8 is the probability of i-th drug combination in inducing ADRs, and $\mathbf{\Theta}$ is the set of parameters of the $\mathtt{D^3I}$ model, including the weighting matrices $\{W^e\}$ (in $\mathtt{D^3I}$ encoder as in Section 4.1) and $\{W^E\}$ (in $\mathtt{D^3I}$ predictor as in Section 4.3) among the fully-connected layers. We use the Adam gradient descent algorithm [19] to solve the problem 9. We use batch training, described in Section 2.1 in the supplementary materials [2], to train $\mathtt{D^3I}$ models. All the hyper-parameters are reported in Section 2.2 in the supplementary materials [2].

4.5 Data Availability

The data and the code are made publicly available [1].

5 MATERIALS

We use two datasets in our experiments to test the performance of $\mathtt{D^3I}$. The first dataset is derived from Chiang *et al.* [7], denoted as FEARS. The second dataset is derived from Zhang *et al.* [41], denoted as BMC. The dataset statistics is presented in Table 2.

Table 2: Dataset Statistics

dataset	#d	#D	$\|D\|$	$\|\bar{D}\|$	features
FEARS	826	6,338	2-52	3.6	substructures (FP), targets (TG), side effects (SE), indications (TI)
BMC	548	48,584	2	2	substructures (FP), targets (TG), off-side effects (OSE), indications (TI), enzymes (EM), pathways (PW), transporters (TP)

The columns corresponding to #d, #D, $\|D\|$, $\|\bar{D}\|$ and "features" have the number of drugs, the number of drug combinations, the cardinalities of drug combinations, average cardinaltiy of drug combinations and the drug features in the dataset, respectively.

[1] https://gitlab.com/peng10/d3i

5.1 FEARS Dataset

The FEARS dataset has 6,338 drug combinations from 826 drugs, including 2,981 2-drug combinations, 1,555 3-drug combinations, 652 4-drug combinations, 323 5-drug combinations, 220 6-drug combinations, 157 7-drug combinations and 450 combinations with more than 7 drugs. The maximum number of drugs in a combination is 52, and the average is 3.6. The drug combinations are selected based on their odds ratios [28] of inducing myopathy among a large collection of spontaneous reports to FDA[2]. The detailed description of drug combination selection and dataset construction is available in the Materials section in Chiang *et al.* [7]. We collected 4 types of information for the drugs, including chemical substructure fingerprints (FP), side-effect profiles (SE), therapeutic-indication profiles (TI) and target profiles (TG). Unfortunately, we cannot find all the 4 types of features for each drug. As the result, the size of used data when using different features as input is different, that is, 6,638 combinations with FP, 3,330 combinations with SE, 3,088 combinations with TI, and 5,621 combinations with TG. The detailed description on such features is available in Section 1 in the supplementary materials [2]. Please note that in FEARS, half of the drug combinations induce myopathy (i.e., positive drug combinations) and the rest do not (i.e., negative drug combinations).

5.2 BMC Dataset

The BMC has 48,584 drug pairs from 548 drugs with 9 different types of drug features, including chemical substructures denoted as FP, drug target profiles denoted as TG, transporter profiles denoted as TP, enzymes denoted as EM, pathways denoted as PW, drug indications denoted as TI, side effects denoted as SE, off-side effects denoted as OSE and the drug-drug interaction profiles. We download the drug similarity profiles calculated from the 7 types of features[3]. For more details about the drug features, drug similarity profiles and BMC, please refer to Zhang *et al.* [41]. Note that in BMC, the drug combinations all induce side effects (i.e., positive drug combinations).

5.3 Generation of Drug Feature Vectors

For each dataset, we calculate the pairwise Jaccard similarity coefficients for all the drugs in the dataset using each of the drug features (e.g., TG, TI), and use each row of the similarity matrix as the corresponding feature vector representation of the corresponding drug. Intuitively, the feature vector $\mathbf{f}$ of a drug d presents the similarities between d and all drugs in the same dataset using the corresponding drug features. This feature representation scheme is inspired by the idea in Que and Belkin [25]. It provides an easy framework to mitigate high-dimension features with missing values and integrate multiple types of features.

6 EXPERIMENTAL PROTOCOLS

6.1 Positive and Negative Data Generation

We conduct the experiments under two settings, denoted as TPTN and TPRN, respectively. In TPTN, we use the true positive and true negative drug combinations from the datasets to train and test our models. That is, the positive and negative samples are fixed from the datasets. In TPRN, we only use the positive drug combinations in the datasets and sample corresponding equal-size negative drug combinations for training and testing.

6.1.1 Negative Data Sampling in TPRN. The negative sample generation process is only conducted in the TPRN setting, that is, for a cardinality-k positive drug combination $D = \{d_1, ...d_k\}$, we sample k drugs and construct a corresponding negative drug combination $D' = \{d'_1, ...d'_k\}$ such that D' is not in the positive drug combinations. Drug d' is selected according to the following distribution P,

$$P(d') = \frac{f(d')}{\sum_{i=1}^{n}(f(d'_i))}, \tag{10}$$

where $f(d')$ denotes the frequency of drug d' in training and validation set (see Section 6.2 for details on cross validation). Please note that sampled drug combinations could be false negative, and thus we need to check the sampled combinations against the training and validation set to remove false negative samples.

The reason why we do negative sampling, even though there could be labeled negative drug combinations, is to avoid the situation in which the classification is biased by a confounder from the cardinalities of drug combinations. We noticed that combinations of high cardinalities are more likely to induce side effects, but true negative drug combinations tend to have low cardinalities (will be discussed later in Section 7.3.1). Therefore, a model trained from such negative drug combinations could be biased by the signals in high-cardinality, true positive drug combinations, and the signals in low-cardinality, true negative drug combinations. By doing the negative sampling as above, we introduce negative training instances of high cardinality, and thus force the model to learn non-trivial signals from drug combinations.

6.2 Cross Validation

We conduct 5-fold cross validation in both TPTN and TPRN settings. In the TPTN setting, we randomly split the original datasets into 5 folds of equal size, with all the folds having relatively same number of true positive/true negative drug combinations. We use 3 folds for model training, 1 fold for validation and 1 fold for testing each time. In the TPRN setting, we randomly split the positive drug combinations in the datasets into 5 folds of equal size. Similarly to the first setting, 3 folds are used for training, and the rest 2 folds are used for testing and validation each time. Before training, we sample negative drug combinations for testing and validation set and fix them (i.e., the negative drug combinations will not change during and after training for the testing and validation set). The negative drug combinations of training set are sampled during training on the fly. That is, in each training batch (Section 2.1 in the supplementary materials [2]), we sample negative drug combinations of the same size and order distribution for the positive drug combinations in that batch. The positive drug combinations and sampled negative drug combinations are together used as training data in the batch to train the model. In both settings, we run experiments for 5 times, with 1 fold as the testing set each time, and report results that are averaged out of the five experiments.

[2]https://www.fda.gov/drugs/informationondrugs/ucm135151.htm
[3]https://github.com/zw9977129/drug-drug-interaction/

6.3 Evaluation Metrics

We use accuracy, precision, recall, F1 and Area Under the ROC Curve (AUC) to evaluate the performance of the various methods. We use TP, FN, TN and FP to denote the number of true positive drug combinations, false negative drug combinations, true negative drug combinations and false positive drug combinations in the testing set, respectively. We also use P to denote the number of positive drug combinations (i.e., P = TP + FN) and N to denote the number of negative drug combinations (i.e., N = FP + TN). Thus, accuracy (acc) is defined as follows,

$$\text{acc} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}}, \tag{11}$$

that is, acc is the fraction of all correctly classified drug combinations over all the drug combinations. Precision (pre) is defined as follows,

$$\text{pre} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{12}$$

that is, pre is the fraction of correctly classified positive drug combinations over all the drug combinations that are classified as positive. Recall (rec) is defined as follows,

$$\text{rec} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{13}$$

that is, it's the fraction of correctly classified positive drug combinations over all the positive drug combinations. F1 is defined as follows,

$$\text{F1} = 2 \times \frac{\text{rec} \times \text{pre}}{\text{rec} + \text{pre}}, \tag{14}$$

that is, it's the harmonic mean of the precision and recall. Area Under the ROC Curve (AUC) [14] is the normalized area under the curve of the true-positive rate against the false positive rate over different classification thresholds. For all the 5 metrics, the larger value indicates better classification performance.

7 EXPERIMENTAL RESULTS

We present the experimental results in this section. Additional experimental results including comparison on drug features and model architectures are available in Section 3 in the supplementary materials [2].

7.1 Overall Performance

7.1.1 Overall performance on FEARS. Table 3 presents the best performance of the three methods $\text{D}^3\text{I}_{\text{max}}$, $\text{D}^3\text{I}_{\text{mean}}$ and $\text{D}^3\text{I}_{\text{Att}}$ on FEARS under the two experimental settings TPTN and TPRN. Note that all the results in the table are selected according to the best F1 values, and the other evaluation measurements according to the best F1 values are also presented. Overall, $\text{D}^3\text{I}_{\text{max}}$ achieves the best performance compared to the other two methods with best F1 0.815 and AUC 0.892 in TPTN, and best F1 0.740 and AUC 0.847 in TPRN. $\text{D}^3\text{I}_{\text{mean}}$ ranks as the second with the best F1 0.766 and AUC 0.842 in TPTN, and the best F1 0.704 and corresponding AUC 0.767 (best AUC 0.770) in TPRN. $\text{D}^3\text{I}_{\text{Att}}$ performs the worst with the best F1 0.756 and AUC 0.834 in TPTN, and the best F1 0.672 and AUC 0.760 in TPRN. These results demonstrate the strong capability of $\text{D}^3\text{I}_{\text{max}}$ in predicting ADRs of drug combinations of various orders.

The primary difference among $\text{D}^3\text{I}_{\text{max}}$, $\text{D}^3\text{I}_{\text{mean}}$ and $\text{D}^3\text{I}_{\text{Att}}$ relies on their aggregators. $\text{D}^3\text{I}_{\text{max}}$ utilizes max pooling as in Equation 3 to construct a combination embedding that consists of the strong signals from each dimension of individual drug embeddings. It is very likely that in the combination embedding, different dimensions selected via max() operator are from different drugs, and therefore, non-linearity in aggregation is realized. More importantly, such combination of embedding dimensions from different drugs corresponds to the notation of drug-drug interaction – intuitively, drugs contribute different aspects all together to introduce ADRs.

The TPTN and TPRN settings are different in the negative drug combinations in both training and testing set. In TPTN, the negative drug combinations typically have different cardinalities compared to those of the positive drug combinations. However, in TPRN, our sampling method as described in Section 6.1.1 guarantees same cardinalities for each positive drug combination and its paired, sampled negative drug combination. The overall worse performance in TPRN compared to that in TPTN indicates the difficulty in learning from same-dimensionality positive and negative drug combinations, and the difficulty in learning synergistic interaction signals from high-cardinality drug combinations. However, our methods are still able to achieve F1 0.740 and AUC 0.847 in TPRN, indicating its strong potential in predicting high-order DDIs and induced ADRs. Compared to TPRN, the TPTN setting is closer to the real application scenario (e.g., different cardinality distributions in positive drug combinations and negative drug combinations), and the good performance of our methods demonstrates their strong potential in high-order DDI prediction in real applications.

7.1.2 Overall performance on BMC. Table 4 presents the overall performance of D^3I methods and the comparison with other methods on the BMC dataset. Please note that BMC dataset has only true positive drug combinations of cardinality 2. The results reported in the original paper [41] correspond to very unbalanced testing data (i.e., 9,716 positives, 101,294 negatives). Therefore, the performance of neighbor-based recommender, random walk and matrix perturbation as used in the paper [41] is good in accuracy and AUC, but not in other metrics. In D^3I methods, we conducted negative sampling and thus the testing data are balanced. In terms of precision, recall and F1, D^3I methods significantly outperform others. In particular, in terms of recall, $\text{D}^3\text{I}_{\text{Att}}$ is 4.6% better than random walk (recall 0.803 vs 0.768), which is the best non-D^3I method. Also, in terms of F1, $\text{D}^3\text{I}_{\text{max}}$ is better than matrix perturbation that achieves the best F1 among all the non-D^3I methods (F1 0.720 vs 0.707).

7.1.3 Clustering Analysis. Figure 2, generated using t-SNE [31] method, presents the single drug embeddings generated from $\text{D}^3\text{I}_{\text{max}}$ (TPRN) on the FEARS dataset. In this figure, there are some well-formed clusters (e.g., C_1, C_3 and C_4). Cluster C_1 primarily includes antipsychotic drugs (e.g., amisulpride, aripiprazole, droperidol, perphenazine, pimozide, pipotiazine, risperidone), antidepressants (e.g., amitriptyline, desipramine, trazodone) and drugs for Parkinson treatment (e.g., isuride, ropinirole) and Huntington treatment (e.g., tetrabenazine). Cluster C_3 includes many anti-inflammatory drugs (e.g., acetylsalicylic acid, flurbiprofen, ibuprofen, loxoprofen, naproxen, rofecoxib, salicyclic acid, tenoxicam). In cluster C_4, most of the drugs (e.g., butabarbital, clonazepam, clotiazepam, etizolam, oxazepam, pentobarbital, thiopental) are used to treat tension, anxiety, nervousness, insomnia, seizures and panic disorders. Cluster C_5 represents a group of drugs (e.g., codeine, heroin, oxycodone,

Table 3: Overall Performance on FEARS Dataset

method	TPTN						TPRN					
	feature	acc	pre	rec	F1	AUC	feature	acc	pre	rec	F1	AUC
D^3I_{max}	TG	**0.823**	**0.862**	0.773	**0.815**	**0.892**	TG	**0.762**	**0.813**	0.680	**0.740**	**0.847**
	TG	0.815	0.834	**0.790**	0.811	0.889	SE	0.700	0.689	**0.748**	0.714	0.784
D^3I_{mean}	FP	**0.773**	**0.790**	0.744	**0.766**	**0.842**	TG	0.706	0.708	0.702	**0.704**	0.767
	FP	0.742	0.734	**0.762**	0.747	0.823	TG	**0.707**	**0.717**	0.683	0.699	**0.770**
	TG	0.761	0.768	0.750	0.759	0.833	SE	0.665	0.650	**0.721**	0.683	0.738
D^3I_{Att}	TG	**0.758**	0.768	0.744	**0.756**	**0.834**	TI	**0.703**	**0.750**	0.609	**0.672**	**0.760**
	FP	0.753	**0.772**	0.720	0.745	0.819	FP	0.649	0.647	**0.661**	0.653	0.719
	FP	0.753	0.756	**0.749**	0.752	0.828	SE	0.668	0.675	0.647	0.661	0.737

Columns "acc", "pre", "rec", "F1" and "AUC" correspond to accuracy, precision, recall, F1 and AUC. The best performance for each method under each evaluation metric is **fold**. The best performance over all the methods is underlined.

Table 4: Overall Performance on BMC Dataset (TPRN)

method	feature	acc	pre	rec	F1	AUC
D^3I_{max}	OSE	**0.693**	**0.663**	**0.788**	**0.720**	**0.744**
D^3I_{mean}	OSE	**0.687**	**0.669**	0.742	**0.703**	**0.743**
	TI	0.681	0.659	**0.752**	0.702	0.734
D^3I_{Att}	OSE	**0.670**	0.635	**0.803**	**0.709**	**0.710**
	TI	**0.670**	**0.640**	0.779	0.702	0.707
neighbor recommender	OSE	**0.951**	**0.629**	**0.765**	**0.691**	**0.940**
random walk	TI	**0.952**	**0.641**	**0.768**	**0.699**	**0.941**
matrix perturbation	-	**0.952**	**0.666**	**0.755**	**0.707**	**0.948**

Columns "acc", "pre", "rec", "F1" and "AUC" correspond to accuracy, precision, recall, F1 and AUC. The original paper [41] did not report drug features used in matrix perturbation method. The best performance for each method under each evaluation metric is **fold**. The best performance over all the methods is underlined.

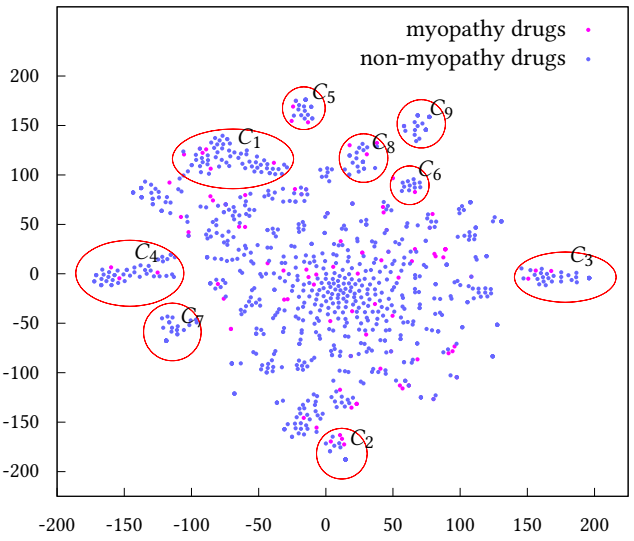

Figure 2: Single Drug Embeddings from D^3I_{max} (FEARS, TPRN)

propoxyphene, sufentanil, tramadol) that are used to treat pains. The above clustering structures among single drug embeddings demonstrate that D^3I methods learn latent representations from single drugs that may conform to domain knowledge.

7.2 Case Study

Table 5 presents some examples that D^3I_{max} is able to correctly predict on FEARS dataset. The first 5 drug combinations have single drugs that induce myopathy on their own (bold), and have various cardinalities. The last 5 drug combinations do not involve any myopathy-inducing single drugs, but all the drugs together in a combination still induce myopathy based on their odds ratios. These results show that D^3I models do not trivially learn from single drugs that induce myopathy, but learn from the synergistic signals from multiple drugs in drug combinations for ADR prediction.

7.3 Comparison over Drug Combination Cardinalities

7.3.1 Performance Comparison over Drug Combinations of Various Cardinalities. Table 6 presents the cardinality distribution of drug combinations in the FEARS dataset. The majority of the drug combinations with ADRs is of order/cardinality 2 or 3. Please note that the total number of drug combinations for different features may be different due to the availability of different features on the drug combinations (Section 5.1).

Table 7 presents the model performance over drug combinations of each cardinality using TG as drug features. In the experiments, all the drug combinations of various cardinalities are used for model training and only drug combinations of each respective cardinality are tested. Table 7 shows that in TPTN setting, interestingly, all the methods share a similar trend in their performance over cardinalities, that is, the F1 values in general increase as the cardinalities increase. However, in TPRN, all the methods tend to achieve their best performance in F1 at drug combination cardinality 3 or 4, and the performance tends to remain similar even when the cardinality

Table 5: Examples of Correctly Predicted Drug Combinations by D^3I_{max} (FEARS, TG, TPRN)

idx	drug combination
1	acetaminophen, alprazolam, **amitriptyline**, amlodipine, anastrozole, azithromycin, baclofen, buprenorphine, calcium, cevimeline, ciprofloxacin, doxycycline, duloxetine, **escitalopram**, estradiol, **fentanyl**, fondaparinux sodium, fulvestrant, furosemide, gabapentin, glucosamine, hydrochlorothiazide, hydrocodone, hydromorphone, ibuprofen, levofloxacin, lidocaine, **methadone**, methocarbamol, metoprolol, montelukast, morphine, moxifloxacin, omeprazole, oxycodone, pamidronate, pantoprazole, pentosan polysulfate, potassium, **pregabalin**, rabeprazole, triamterene, valaciclovir, **valdecoxib**, vitamin c, zoledronate, zolpidem
2	alprazolam, pioglitazone, **rosuvastatin**, **sunitinib**, tamsulosin, valsartan
3	alendronate, cetaminophen, chlorpheniramine, codeine, naproxen, **prednisolone**, **zopiclone**
4	atenolol, **pravastatin**
5	diphenhydramine, hydromorphone, montelukast, omeprazole, razepam, **triamcinolone**
6	gabapentin, haloperidol, morphine, propofol
7	alprazolam, diazepam, diclofenac, dicyclomine, etizolam, losartan, sulpiride
8	atenolol levofloxacin
9	amikacin, amiodarone
10	acetaminophen, alendronate, oxycodone

The drugs that induce myopathy on their own are **bold**.

Table 6: Cardinality Distribution in FEARS (TPTN)

feature		total	2	3	4	5	6	7	≥8
FP	all	6,338	2,981	1,555	652	323	220	157	450
	#pos	3,169	865	841	442	263	195	138	425
	#neg	3,169	2,116	714	210	60	25	19	25
TG	total	5,621	2,809	1,395	544	252	169	132	320
	#pos	2,821	822	795	402	222	158	121	301
	#neg	2,800	1,987	600	142	30	11	11	19

The columns corresponding to "2", "3", ..., "≥8" represent the numbers of drug combinations of cardinality 2, 3, ..., greater than 8. The row of "all" has the total number of drug combinations. The row of "#pos" has the numbers of positive drug combinations. The row of "#neg" has the numbers of negative drug combinations.

increases. In TPTN, as cardinality increases, the true positive drug combinations become more than the true negative drug combinations (Table 6). Therefore, D^3I model training in TPTN is biased by the true positive drug combinations of higher cardinalities, and the true negative drug combinations of lower cardinalities. Consequentially, all D^3I methods in TPTN tend to have better precision and recall performance on drug combinations of higher cardinalities. Please note that D^3I methods are cardinality-invariant and they do not use the cardinality information in prediction. The biased performance in TPTN, although not preferable, actually demonstrates that D^3I methods do learn signals from the multiple drugs in drug combinations.

The strong ability of D^3I methods in learning from multiple drugs in drug combinations is also demonstrated by their performance in TPRN in Table 7. In TPRN, each true positive drug combination will have a corresponding negative drug combinations of same cardinality, and thus the learning of D^3I models will not be biased by the unbalanced distribution between positive and negative drug combinations. In TPRN, D^3I_{max} is able to achieve F1 values above 0.760 for cardinalities higher than 3. In particular, for higher cardinalities, D^3I_{max} achieves even better performance, for example, for cardinality higher than or equal to 8, D^3I_{max} achieves F1 value 0.811.

7.3.2 Performance Comparison over Order-2 Drug Combinations. Table 8 presents the testing results on drug pairs (i.e., drug combinations of cardinality 2) using drug combinations of only cardinality 2 for model training, and using all cardinalities for model training, in D^3I methods. All the experiments are conducted in TPRN setting to avoid biases from imbalanced training data distributions. Table 8 shows that when only drug pairs are used for training (i.e., the first column block in Table 8), the best F1 performance is 0.680, achieved by D^3I_{mean} (with FP as the drug features), and the best AUC performance is 0.765, achieved by D^3I_{max} (with TG as the drug features). However, when drug combinations of all cardinalities are used for training (i.e., the second column block in Table 8), the best F1 performance is 0.685, achieved by D^3I_{max} (with FP as the drug features), and the best AUC performance is 0.786, achieved by D^3I_{max} (with TG as the drug features). The better performance using all-cardinality drug combinations for training demonstrates that D^3I methods do not trivially consider drug combination cardinalities in learning and prediction, but do learn the signals from all drug combinations. In addition, when all-cardinality drug combinations are used for training, D^3I methods are able to capture the more and richer information carried by those drug combinations, and thus better predict drug pairs.

8 CONCLUSIONS AND FUTURE RESEARCH

In this manuscript, we presented our deep learning model D^3I for predicting adverse drug reactions induced by high-order drug-drug interactions. D^3I is able to predict for drug combinations of arbitrary numbers of drugs, and generate meaningful embeddings for single drugs and drug combinations. We tested D^3I on two real datasets, one involving pairwise drug-drug interactions and the other involving high-order drug-drug interactions. Our experimental results demonstrate that D^3I is able to achieve superior performance on high-order drug-drug interaction prediction.

In D^3I, different drug features (e.g., target profiles, side effect profiles) are used independently. Effective integration of such features together may better represent drugs and their properties, and thus enable better performance of deep learning models. In our future work, we will explore feature integration and fusion in D^3I models. In addition, other information may be also highly related to drug-drug interactions and their induced adverse reactions, such as protein pathways and evidences from electronic medical records. We also plan to explore effective methods to integrate such information in D^3I models to further improve D^3I performance. Interpretability and evidence support are important for prediction methods in biomedical applications. A known issue in

Table 7: Performance Comparison on Different Cardinalities in FEARS Dataset (TG)

method	cardinality	TPTN					TPRN				
		acc	pre	rec	F1	AUC	acc	pre	rec	F1	AUC
D^3I_{max}	2	0.810	0.749	0.532	0.621	0.821	0.697	0.752	0.587	0.659	0.786
	3	0.789	0.853	0.763	0.804	0.870	0.774	0.808	0.719	0.760	0.847
	4	0.819	0.879	0.875	0.877	0.848	0.775	0.798	0.737	0.766	0.862
	5	0.913	0.954	0.949	0.950	0.925	0.793	0.883	0.676	0.765	0.885
	6	0.918	0.944	0.971	0.956	0.563	0.799	0.900	0.679	0.770	0.887
	7	0.909	0.913	0.994	0.951	0.716	0.802	0.879	0.704	0.780	0.882
	≥8	0.938	0.941	0.997	0.968	0.709	0.828	0.895	0.743	0.811	0.913
D^3I_{mean}	2	0.754	0.562	0.719	0.631	0.811	0.664	0.642	0.744	0.688	0.732
	3	0.763	0.836	0.726	0.776	0.836	0.723	0.707	0.766	0.734	0.786
	4	0.718	0.901	0.695	0.784	0.818	0.748	0.760	0.724	0.740	0.806
	5	0.777	0.972	0.768	0.857	0.880	0.750	0.812	0.653	0.722	0.837
	6	0.726	0.940	0.758	0.834	0.479	0.724	0.824	0.570	0.672	0.803
	7	0.791	0.928	0.838	0.877	0.716	0.760	0.873	0.611	0.711	0.827
	≥8	0.877	0.951	0.914	0.931	0.735	0.670	0.940	0.364	0.523	0.824
D^3I_{Att}	2	0.744	0.550	0.696	0.614	0.806	0.665	0.674	0.644	0.658	0.726
	3	0.750	0.838	0.698	0.760	0.838	0.651	0.652	0.645	0.648	0.735
	4	0.741	0.913	0.719	0.803	0.816	0.680	0.683	0.670	0.675	0.753
	5	0.786	0.972	0.778	0.864	0.867	0.662	0.693	0.583	0.631	0.761
	6	0.756	0.942	0.789	0.857	0.394	0.651	0.671	0.587	0.622	0.718
	7	0.833	0.930	0.883	0.904	0.701	0.689	0.731	0.595	0.653	0.773
	≥8	0.897	0.956	0.933	0.944	0.677	0.632	0.687	0.508	0.576	0.733

Columns "acc", "pre", "rec", "F1" and "AUC" correspond to accuracy, precision, recall, F1 and AUC. Target profiles (TG) are used as drug features. The best results presented for each drug combination cardinality are selected based on F1.

Table 8: Best Performance on Cardinality-2 Drug Combinations (FEARS, TPRN)

method	training with cardinality-2 drug combinations						training with all-cardinality drug combinations					
	feature	acc	pre	rec	F1	AUC	feature	acc	pre	rec	F1	AUC
D^3I_{max}	FP	0.655	0.637	**0.724**	**0.677**	0.710	FP	0.672	0.659	**0.715**	**0.685**	0.729
	TG	**0.697**	**0.742**	0.610	0.668	**0.765**	TG	**0.697**	**0.752**	0.587	0.659	**0.786**
D^3I_{mean}	FP	0.666	0.652	**0.714**	**0.680**	0.725	TG	**0.651**	0.630	**0.742**	**0.680**	**0.732**
	TG	**0.695**	**0.725**	0.633	0.675	**0.747**	FP	0.635	**0.634**	0.646	0.638	0.697
D^3I_{Att}	TG	**0.689**	0.721	**0.620**	**0.665**	**0.749**	TG	**0.665**	**0.674**	0.644	**0.658**	**0.726**
	TG	0.687	**0.736**	0.589	0.653	0.736	FP	0.617	0.610	**0.656**	0.629	0.687

Columns "acc", "pre", "rec", "F1" and "AUC" correspond to accuracy, precision, recall, F1 and AUC. The best performance for each method under each evaluation metric is **fold**. The best performance over all the methods is underlined.

deep learning is its lack of interpretability by design, and thus it is worthwhile to address the interpretability issues of D^3I (e.g., what each layer learns, what the embeddings represent) in our future research. Mining evidences to support high-order drug-drug interactions and their adverse reactions from literature and electronic medical records is a challenging, related task that we would like to explore in the future.

9 ACKNOWLEDGMENTS

This project was made possible, in part, by support from the National Science Foundation under Grant Number IIS-1855501 and IIS-1827472. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] National Health and Nutrition Examination Survey. http://www.cdc.gov/NCHS/NHANES.htm.
[2] Supplementary Materials. https://u.osu.edu/ning.104/files/2019/07/D3Isup.pdf.
[3] Bissan Al-Lazikani, Udai Banerji, and Paul Workman. 2012. Combinatorial drug therapy for cancer in the post-genomic era. *Nature biotechnology* 30, 7 (2012), 679.
[4] Christine Castro and Mark Gourley. 2012. Diagnosis and treatment of inflammatory myopathy: issues and management. *Therapeutic advances in musculoskeletal disease* 4, 2 (2012), 111–120.

[5] Feixiong Cheng and Zhongming Zhao. 2014. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *Journal of the American Medical Informatics Association* 21, e2 (2014), e278–e286.

[6] Chien-Wei Chiang, Pengyue Zhang, Xueying Wang, Lei Wang, Shijun Zhang, Xia Ning, Li Shen, Sara K Quinney, and Lang Li. 2018. Translational High-Dimensional Drug Interaction Discovery and Validation Using Health Record Databases and Pharmacokinetics Models. *Clinical Pharmacology & Therapeutics* 103, 2 (2018), 287–295.

[7] Wen-Hao Chiang, Shen Li, Li Lang, and Ning Xia. 2019. Drug-drug interaction prediction based on co-medication patterns and graph matching. *arXiv preprint arXiv:1902.08675* (2019).

[8] Sean Ekins and Steven A Wrighton. 2001. Application of in silico approaches to predicting drug–drug interactions. *Journal of pharmacological and toxicological methods* 45, 1 (2001), 65–69.

[9] QiPing Feng, Russell A Wilke, and Tesfaye M Baye. 2012. Individualized risk for statin-induced myopathy: current knowledge, emerging challenges and potential solutions. *Pharmacogenomics* 13, 5 (2012), 579–594.

[10] Assaf Gottlieb, Gideon Y Stein, Yoram Oron, Eytan Ruppin, and Roded Sharan. 2012. INDI: a computational framework for inferring drug interactions and their associated recommendations. *Molecular systems biology* 8, 1 (2012), 592.

[11] Felix Hammann and Juergen Drewe. 2014. Data mining for potential adverse drugâĂŞdrug interactions. *Expert Opinion on Drug Metabolism & Toxicology* 10, 5 (2014), 665–671. arXiv:http://dx.doi.org/10.1517/17425255.2014.894507 PMID: 24588496.

[12] R Harpaz, W DuMouchel, N H Shah, D Madigan, P Ryan, and C Friedman. 2012. Novel Data-Mining Methodologies for Adverse Drug Event Discovery and Analysis. *Clinical Pharmacology & Therapeutics* 91, 6 (2012), 1010–1021.

[13] Rave Harpaz, Krystl Haerian, Herbert S. Chase, and Carol Friedman. 2010. Statistical Mining of Potential Drug Interaction Adverse Effects in FDA's Spontaneous Reporting System. *AMIA Annu Symp Proc* 2010 (2010), 281–285.

[14] Jin Huang and Charles X Ling. 2005. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering* 17, 3 (2005), 299–310.

[15] Jialiang Huang, Chaoqun Niu, Christopher D Green, Lun Yang, Hongkang Mei, and Jing-Dong J Han. 2013. Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. *PLoS computational biology* 9, 3 (2013), e1002998.

[16] Heba Ibrahim, Amr Saad, Amany Abdo, and A Sharaf Eldin. 2016. Mining association patterns of drug-interactions using post marketing FDAâĂŹs spontaneous reporting data. *Journal of biomedical informatics* 60 (2016), 294–308.

[17] Maximilian Ilse, Jakub M. Tomczak, and Max Welling. 2018. Attention-based Deep Multiple Instance Learning. *CoRR* abs/1802.04712 (2018). arXiv:1802.04712

[18] Srinivasan V Iyer, Rave Harpaz, Paea LePendu, Anna Bauer-Mehren, and Nigam H Shah. 2014. Mining clinical text for signals of adverse drug-drug interactions. *Journal of the American Medical Informatics Association* 21, 2 (2014), 353–362.

[19] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). arXiv:1412.6980

[20] Artemy Kolchinsky, Anália Lourenço, Lang Li, and Luis Mateus Rocha. 2013. Evaluation of Linear Classifiers on Articles Containing Pharmacokinetic Evidence of Drug-Drug Interactions. In *Biocomputing 2013: Proceedings of the Pacific Symposium, Kohala Coast, Hawaii, USA, January 3-7, 2013*. 409–420.

[21] Yi Li and Karl J Hale. 2007. Asymmetric total synthesis and formal total synthesis of the antitumor sesquiterpenoid (+)-eremantholide A. *Org Lett* 9, 7 (Mar 2007), 1267–1270.

[22] Heng Luo, Ping Zhang, Hui Huang, Jialiang Huang, Emily Kao, Leming Shi, Lin He, and Lun Yang. 2014. DDI-CPI, a server that predicts drugâĂŞdrug interactions through implementing the chemicalâĂŞprotein interactome. *Nucleic Acids Research* (2014). arXiv:http://nar.oxfordjournals.org/content/early/2014/05/29/nar.gku433.full.pdf+html

[23] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[24] Bethany Percha and Russ B. Altman. 2013. Informatics confronts drug-drug interactions. *Trends in pharmacological sciences* 34, 3 (March 2013), 178–184.

[25] Qichao Que and Mikhail Belkin. 2016. Back to the future: Radial basis function networks revisited. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. 1375–1383.

[26] Sunil Kumar Sahu and Ashish Anand. 2017. Drug-Drug Interaction Extraction from Biomedical Text Using Long Short Term Memory Network. *CoRR* abs/1701.08303 (2017). arXiv:1701.08303

[27] Isabel Segura-Bedmar, Paloma Martínez, and María Herrero Zazo. 2013. Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, Vol. 2. 341–350.

[28] Magdalena Szumilas. 2010. Explaining odds ratios. *Journal of the Canadian academy of child and adolescent psychiatry* 19, 3 (2010), 227.

[29] Nicholas P Tatonetti, JC Denny, SN Murphy, GH Fernald, G Krishnan, V Castro, P Yue, PS Tsau, I Kohane, DM Roden, et al. 2011. Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. *Clinical Pharmacology & Therapeutics* 90, 1 (2011), 133–142.

[30] Elif Tekin, Casey Beppler, Cynthia White, Zhiyuan Mao, Van M Savage, and Pamela J Yeh. 2016. Enhanced identification of synergistic and antagonistic emergent interactions among three or more drugs. *Journal of The Royal Society Interface* 13, 119 (2016), 20160332.

[31] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9 (2008), 2579–2605.

[32] Santiago Vilar, Carol Friedman, and George Hripcsak. 2017. Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media. *Briefings in Bioinformatics* (2017), bbx010.

[33] Santiago Vilar, Eugenio Uriarte, Lourdes Santana, Tal Lorberbaum, George Hripcsak, Carol Friedman, and Nicholas P Tatonetti. 2014. Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nature protocols* 9, 9 (2014), 2147–2163.

[34] Chi-Shiang Wang, Pei-Ju Lin, Ching-Lan Cheng, Shu-Hua Tai, Yea-Huei Kao Yang, and Jung-Hsien Chiang. 2019. Detecting Potential Adverse Drug Reactions Using a Deep Neural Network Model. *Journal of medical Internet research* 21, 2 (2019), e11016.

[35] Xinggang Wang, Yongluan Yan, Peng Tang, Xiang Bai, and Wenyu Liu. 2016. Revisiting Multiple Instance Neural Networks. *arXiv preprint arXiv:1610.02501* (2016).

[36] Xueying Wang, Pengyue Zhang, Chien-Wei Chiang, Hengyi Wu, Li Shen, Xia Ning, Donglin Zeng, Lei Wang, Sara K Quinney, Weixing Feng, et al. 2017. Mixture drug-count response model for the high-dimensional drug combinatory effect on myopathy. *Statistics in medicine* (2017).

[37] Larry C Wienkers and Timothy G Heath. 2005. Predicting in vivo drug interactions from in vitro drug discovery data. *Nature reviews drug discovery* 4, 10 (2005), 825–833.

[38] Su Yan, Xiaoqian Jiang, and Ying Chen. 2013. Text mining driven drug-drug interaction detection. In *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on*. IEEE, 349–354.

[39] Haodong Yang and Christopher C Yang. 2013. Harnessing social media for drug-drug interactions detection. In *Healthcare Informatics (ICHI), 2013 IEEE International Conference on*. IEEE, 22–29.

[40] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. 2017. Deep Sets. *CoRR* abs/1703.06114 (2017). arXiv:1703.06114

[41] Wen Zhang, Yanlin Chen, Feng Liu, Fei Luo, Gang Tian, and Xiaohong Li. 2017. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinformatics* 18, 1 (05 Jan 2017), 18.

[42] Yu Zhang and Dit-Yan Yeung. 2013. Learning High-order Task Relationships in Multi-task Learning. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI '13)*. AAAI Press, 1917–1923.

[43] Xing-Ming Zhao, Murat Iskar, Georg Zeller, Michael Kuhn, Vera Van Noort, and Peer Bork. 2011. Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS computational biology* 7, 12 (2011), e1002323.

[44] Anat Zimmer, Itay Katzir, Erez Dekel, Avraham E Mayo, and Uri Alon. 2016. Prediction of multidimensional drug dose responses based on measurements of drug pairs. *Proceedings of the National Academy of Sciences* 113, 37 (2016), 10442–10447.

[45] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. 2018. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 34, 13 (2018), i457–i466.