Drug-Drug Interaction Prediction by Deep Learning Approach

Mohammad.Amin Khodamoradi [1,2], Bahareh Levian [1,2] and Changiz Eslahchi [1,2,*]

[1] *Department of Computer Science, Faculty of Mathematical Science,Shahid Beheshti University, City Tehran , Country Iran*
[2] *Laboratory Eslahchi, School of Bioinformatics, IPM - Institute for Research in Fundamental Sciences, City Tehran , Country Iran*

Correspondence*:
Changiz Eslahchi
eslahchi.ch@gmail.com

ABSTRACT

Drug-drug interactions (DDIs) always cause unexpected and even adverse drug reactions. It is important to identify DDIs before drugs are used in the market.However, preclinical identification of DDIs requires much money and time. Computational approaches have exhibited their abilities to predict potential DDIs on a large scale by utilizing premarket drug properties. Nevertheless, most of them only predict whether or not one drug interacts with another, but neglect their enhancive (positive) and depressive (negative) changes of pharmacological effects. Moreover, these comprehensive DDIs do not occur at random, and derived from the structural features of the graph of DDIs. Revealing such a relationship is very important, because it is able to help understand how DDIs occur. Both the prediction of comprehensive DDIs and the discovery of structural relationship among them play an important guidance when making a co-prescription.

In this work, treating a set of comprehensive DDIs as a signed network, we design a novel model (SNF-CNN) for the prediction of enhancive and degressive DDIs based on similarity network fusion and convolutional neural networks. SNF-CNN achieves the depressive DDI prediction ($AUC = 0/9747 \pm 0/0033$ and $AUPR = 0/9666 \pm 0/0045$), enhancive DDI prediction ($AUC = 0/9686 \pm 0/0028$ and $AUPR = 0/8221 \pm 0/0184$) and the Unknown DDI prediction ($AUC = 0/9714 \pm 0/0040$ and $AUPR = 0/9480 \pm 0/0083$). Compared with three state-of-the-art approaches, SNF-CNN shows it superiority.

This new approach is not only able to predict comprehensive DDI, but also predicts non-DDI.

Keywords: Drug-Drug Interaction, Drug Similarity, Drug Similarity Integration, Feature Selection, Recommender System

1 INTRODUCTION

When two or more drugs are taken together, drugs' effects or behaviors are unexpectedly influenced by each other Wienkers and Heath (2005). This kind of influence is termed as Drug-Drug interaction (DDI), which would reduce drug efficacy, increase unexpected toxicity, or induce other adverse drug reactions between the co-prescribed drugs. As the number of approved drugs increases, the number of drug-unidentified DDIs is rapidly increasing, such that among approved small molecular drugs in Drug Bank, on average, 15 out of every 100 drug pairs have DDIs Law et al. (2014). The DDIs would put patients who are treated with multiple drugs in an unsafe situation Leape et al. (1995); Businaro (2013); Karbownik et al. (2017); Mulroy et al. (2017). Understanding DDI is the first step in drug combinations, which becomes one of the most promising solutions for the treatment of multifactorial complex diseases Zhao et al. (2011). Therefore, there is an urgent need for screening and analysis of DDIs before clinical co-medications are administered. However, traditional DDI identification approaches (e.g., testing Cytochrome P450 Veith et al. (2009) or transporter-associated interactions Huang et al. (2007)) face challenges, such as high costs, long duration, animal welfare considerations Zhang et al. (2015), the very limited number of participants in the trial, and the great number of drug combinations under screening in clinical trials. As a result, only a few DDIs have been identified during drug development production (usually in the clinical trial phase). Some of them have been reported after drugs approved, and many have been found in post-marketing surveillance.

Computational approaches are a promising alternative to discovering potential DDIs on a large scale, and they have gained attention from academy and industry recently Wiśniowska and Polak (2016); Zhou et al. (2016). Data mining-based computational approaches have been developed to detect DDIs from various sources Zhang et al. (2015) , such as scientific literature Bui et al. (2014); Zhang et al. (2016b) , electronic medical records Yamanishi et al. (2008) , and the Adverse Event Reporting System of FDA (http://www.fda.gov). These approaches rely on post-market clinical evidence. So, they cannot provide alerts of potential DDIs before clinical medications are administered. In contrast, machine learning-based computational approaches (e.g. Naïve Similarity-Based Approach Vilar et al. (2014) , Network Recommendation-Based Zhang et al. (2015) , Classification-Based Cheng and Zhao (2014)) can provide such alerts by utilizing pre-marketed or post-marketed drug attributes, such as drug features or similarities Pahikkala et al. (2015). These methods use different drug features to predict DDIs, such as chemical structures Vilar et al. (2014) , targets Luo et al. (2014) , hierarchical classification codes Cheng and Zhao (2014) , side effects, and off-label side effects Zhang et al. (2015); Shi et al. (2017).

A Dependency-based Convolutional Neural Network (DCNN) has proposed for drug-drug interaction extraction at paper of Liu et al. (2016). DCNN is a text-mining approach which predicts DDIs based on unstructured biomedical literature and the existing knowledge bases. It applies convolution layers on word sequences as well as dependency parsing trees of candidate DDIs for adjacent words. DeepDDI has proposed by Ryu et al. (2018), which is a combination of the structural similarity profile generation pipeline and Deep Neural Network (DNN). DeepDDI predicts DDIs from chemical structures and names of drugs in pairs. It has various implications for adverse drug events such as prediction of potential causal mechanism and using them for output sentences.

Although previous methods had great advances, more prediction accuracy is still needed. Exploiting more similarities may help to make more advances in this way. Similarity Network Fusion (SNF) Wang et al. (2014) is a competent method to integrate various similarities, which is used in numerous biological contexts Olayan et al. (2018); Tian et al. (2017); Kim et al. (2016). The neural network is a strongly developed approach that provides satisfactory solutions, especially for large datasets and nonlinear analyzes Wang et al. (2016), which is widely used in critical problems Huang et al. (2009); Fu and Peng (2017); Pan et al. (2016). We developed a method to overcome this issue via similarity fusion and Convolutional Neural Network.

Most of these existing machine learning approaches are designed to predict the typical two-class problem, which only indicates how likely a pair of drugs is a DDI. However, two interacting drugs may change their own pharmacological behaviors or effects (e.g., increasing or decreasing serum concentration) in vivo. For example, the serum concentration of Flunisolide (DrugBank Id: DB00180) decreases when it is taken with Mitotane (DrugBank Id: DB00648), whereas its serum concentration increases when taken with Roxithromycin (DrugBank Id: DB00778). For short, the first case is degressive DDI, and the second case is enhancive DDI, which contains drug changes in terms of pharmacological effects. It is more important to know exactly whether the interaction increases or decreases the drug's pharmaceutical behaviors, especially when making optimal patient care, establishing drug dosage, designing prophylactic drug therapy, or finding the resistance to therapy with a drug Koch-Weser (1981).

On the other hand, the occurrence of both enhancive and degressive DDIs is not random, but most current approaches have not yet exploited this structural property and have been developed only for conventional two-classes DDIs. Furthermore, revealing such a structural relationship is very important because it can help us understand how DDIs occur. It is one of the most important steps for treating complex diseases and guides physicians in preparing safer prescriptions to high-order drug interaction. The proposed algorithms for predicting three-classes DDIs are introduced in the following. And how they work are briefly described. All three introduced algorithms use matrix factorization methods, which is a network recommender-based approach. The matrix factorization approach, with slightly modifying, is a suitable solution for the subject of predicting DDI that has received much attention from researchers.

In this paper, we firstly introduce data and features. Then, a novel algorithm (SNF-CNN) based on the integration of drug similarities and deep learning recommendation systems for predicting DDI is presented in a comprehensive three-class model. This algorithm is called Predicting Comprehensive Drug-Drug Interaction via Similarity Network Fusion and Convolutional Neural Networks.

The paper is organized as follows. In the first section, the data preparation process is explained. The recommendation system is then designed and trained on enhancive and degressive, which detects pairs of non-interacting drugs with high probability. Next, the previous recommender system, based on a convolutional neural network, is trained on incremental and decremental interaction data without interaction (detected in the previous step).In section Results and Discussions, we investigate the results of SNF-CNN in the 10-fold cross-validation (CV) process.

It should be noted that the proposed method of this research is a recommender-based on deep neural networks and has no structural similarities with matrix factorization methods. The only reason for mentioning these methods is the limited number of articles that have used three-class data in their work.

2 METHODS

2.1 Dataset and features

In this study, we use the data set presented in paper of Yu et al. (2018). This set contains 568 approved small molecule drugs, each of them has at least one interaction with the other drugs in the set. In total, the interactions between these 568 drugs contain 21,351 DDIs, including 16,757 enhancive DDIs and 4,594 degressive DDIs.In addition, each drug represented as an 881-dimensional feature vector F_{str} based on PubChem chemical structure descriptor and also a 9149-dimensional feature vector F_{se} according to the off-label side effects provided by OFFSIDES.

2.2 Problem formulation

Without loss of generality, let $D = \{d_i\}$, $i = 1, 2, .., m$ be a set of m approved drugs. Their interactions can be accordingly represented as an $m \times m$ symmetric interaction matrix $A_{m\times m} = \{a_{ij}\}$. For the conventional DDIs, $a_{ij} = 1$ if d_i interacts with d_j, and $a_{ij} = 0$ otherwise. For the comprehensive DDIs,

$a_{ij} \in \{-1, 0, +1\}$. Again, if d_i and d_j do not interact with each other, $a_{ij} = 0$. When there is an enhancive DDI or a degressive DDI between d_i and d_j , $a_{ij} = +1$ or $a_{ij} = -1$ respectively.

In addition, each drug d_i in the D is represented as a p-dimension feature vector $f_i = [f_1, f_2, .., f_k, .., f_p]$, which $f_k = 1$ indicates the K-th specific chemical structure fragment or occurs an off-label side effect, and $f_k = 0$ otherwise. Because each drug has two chemical structure feature vectors and off-label side effects, there are two feature matrices of F with dimensions of $m \times p$ (amount of p depends on kind of feature). Matrices of F_{str} and F_{se} are, respectively, the feature matrix of the chemical structure and the feature matrix of off-label side effects.

2.3 Data preparing

Since the new drugs are isolated nodes in the interaction network, we cannot infer their possible interaction from topological information alone. Therefore, additional information (such as chemical structure or off-label side effects) is needed, which is called a drug feature in terms of machine learning. First, we prepare the features based on our model, and then we teach a deep learning model of interaction prediction.

2.3.1 Similarity matrix calculation

A common method of calculating similarity called Cosine Similarity is used in machine learning articles such as Articles Zhang et al. (2016a, 2018). If we name feature vectors of the drug of d_i and d_j as x_i and x_j,Cosine Similarity between x_i and x_j is defined as follows:

$$S_{Cos}(x_i, x_j) = \frac{x_i.x_j}{||x_i||_2||x_j||_2} \tag{1}$$

Where $||.||_2$ is the Euclidean Norm and $x_i.x_j$ is inner product of two vectors.

It was observed that the values of the feature matrices are discrete, and also the dimensions of the matrices are large. The chemical structure and the off-label side effect have 881 and 9149 dimensions, respectively. On the other hand, machine learning algorithms do not work properly with high-dimensional data and discrete data. As a result, they do not get good results on these kinds of data. Therefore, by exploiting the cosine similarity, that was described above, drug similarity matrices based on chemical structure and off-label side effects are calculated. These matrices are S_{str} and S_{se}, respectively. The dimensions of these two matrices are $m \times m$, where $s_{i,j}$ is an element of similarity matrices that shows similarity value between drugs of di and d_j. Each element of S has a continuous value between zero and one.

2.3.2 Integration drug similarity matrices

Similarity Network Fusion (SNF)Wang et al. (2014) is a new computational method for data integration. Briefly, SNF combines many different types of features (such as chemical structure and off-label side effect, and more - clinical data, questionnaires, image data, etc.) for a given set of samples (e.g., drugs). SNF first constructs a sample similarity network for each of the data types and then iteratively integrates these networks using a novel network fusion method. Working in the sample network space allows SNF to avoid dealing with different scales, collection bias, and noise in different data types. Integrating data in a non-linear fashion allows SNF to take advantage of the common and complementary information in different data types. Figure 1 is a good visualization of SNF processes that has been used in our method structure.

In this section, using the similarity network fusion method that has described above, similarity matrices of the chemical structure and the off-label side effect of drugs were integrated. The output of this integration is

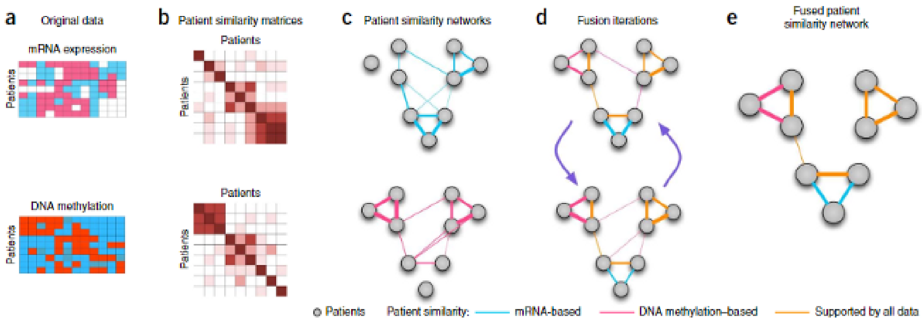

Figure 1. SNF processes Wang et al. (2014): A detailed example of SNF steps. (a) An example representation of chemical structure feature and off-label side effect feature for the same set of drugs. (b) Drug-drug similarity matrices for each feature type. (c) Drug-drug similarity networks, equivalent to the drug-drug data. Nodes represent drugs, and edges represent drug pairwise similarities. (d) Network fusion by SNF iteratively updates each of the networks with information from the other networks, making them more similar with each step. (e) The iterative network fusion results in convergence to the final fused network. Edge color indicates which data type has contributed to the given similarity.

a new similarity matrix of S_{snf} with dimensions of 568×568, and elements of S_{snf} have a value between zero and one.To integrate the network similarity, the package of SNFPy is used, which is implemented in Python and is available at Ross Markello (2018).

2.3.3 Input matrix format

At this stage, a matrix forms with 1139 columns and 322056 rows, which consists of the following columns:

1) Drug pairs: Name of the drug i-th and the name of the drug j-th.

2) Type of interaction: degressive (-1), enhancive (+1), and unknown (0).

3) The similarity vector of i-th drug from the S_{snf} matrix with 568 elements.

4) The similarity vector of j-th drug from the S_{snf} matrix with 568 elements.

Figure of 2 shows the header of matrix.

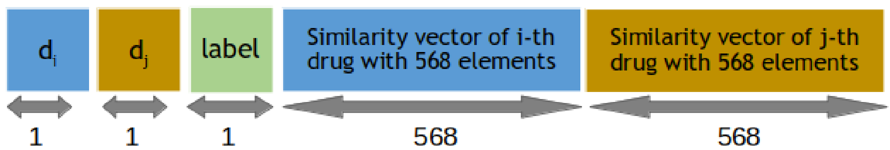

Figure 2. Matrix header of B

We have 568 drugs, and the interaction of each drug with itself is meaningless. On the other hand, the drug pairs of (d_i, d_j) and (d_j, d_i) have the same label, while the corresponding similarity vectors of drugs in the drug pairs have been displaced. So, these drug pairs are dual. Both of them in the data augment the training data, which increases the model's ability to have a better prediction. As a result, the resulting matrix has 322056 data samples or rows $((568 \times 568) - 568 = 322056)$. According to the explanations provided, a matrix with dimensions of 322056×1139 is formed to input into our model, which is called B.

2.4 Devising of Recommender System

In the previous steps, data was prepared to input any learning machine, including deep learning machines. But before presenting the model and inputting the data into the machine, one important point must be considered. As mentioned before, in this approach and other approaches, the positive and negative data have a specific label. While the zero label does not mean that there is no interaction between a pair of drugs, it does indicate that no interaction has yet been found for this pair of drugs. In the following, we present a method for detecting pairs of non-interacting drugs. Then we use these pairs of drugs as zero-labeled data in the next training.

2.5 Selecting and training model on known interactions

To solve this problem, it is necessary to provide a model that detects non-interaction with high accuracy and confidence. Therefore, we design a model based on deep learning that predicts the possible non-interaction drug pairs and then use it to design a three-class model. Obviously, high accuracy in detecting these zeros can help provide a more accurate and confident three-class model.

2.6 Selecting model

We separated rows of matrix B that contain positive and negative interactions. The new matrix contains 42,702 pairs of drugs with degressive and enhancive interactions. This data was used to train and find a more suitable model and found a stronger model among many models with different network structures. The final model was a deep neural network that used convolutional and fully connected layers. The features of all interactions (+1 and -1) contain 1136 features. We first divide these features into 10 equal parts. Then, in a 10-cycle for, each period we consider 1 part as testing data and the other 9 parts as a set of training data. We select different models and train the model in the 10-fold CV with 90 of the data. Then we test the model on the remaining 10 percent of the data. In the separating process, pairs of drugs dual are considered. Since the (d_i, d_j) and (d_j, d_i) pairs of drugs are not biologically different from each other, in the separation of training and testing data, necessarily a pair of drugs and their duals are in the same group. This work prevents machine fraud.

After testing the different structures, we have modeled the final deep neural network shown in Figure 3.

This network has three layers of two-dimensional convolution. In the following, there are three fully connected convolution layers. The last layer has two outputs for predicting degressive or enhancive interaction. Convolution layers have 4-dimensions square filters with a Stride of 1. Each convolution layer also has a Rectified Linear Units (ReLU) activation function Nair and Hinton (2010), which is defined as the positive part of its argument:

$$ReLU(x) = max\{x, 0\} \tag{2}$$

The number of convolution filters is 128, 32, and 8, respectively. All connected layers have 64, 16, and 2 nodes, respectively. The first two layers have the activation function of ReLU, and the last layer with 2 nodes has a Sigmoid activation function Hinton et al. (2012), which is calculated as follows:

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

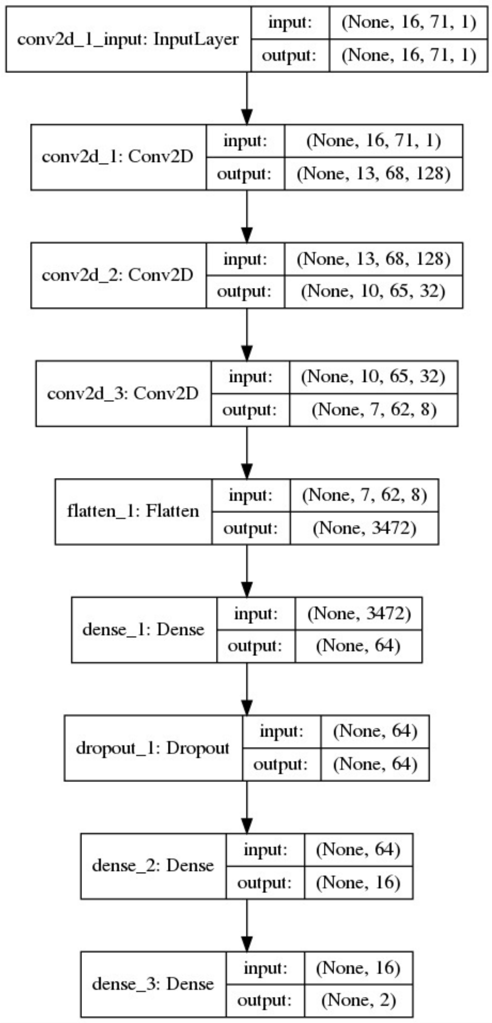

Figure 3. The arrangement of the neural network layers for detecting possible zeros

Convolution layers using a Flatten layer Connects to fully connected layers. The function of this layer is to transform a two-dimensional matrix into a one-dimensional vector. The output of this input layer of the first layer is fully connected. Also, between fully connected 64 and 16 nodes, we used one Dropout layer Srivastava et al. (2014) with a wast value of 0.2. This value indicates that the network in this layer does not randomly consider 20 percent of the features. This layer is used to prevent over-fitting of the model and forces the model to extract and use more features with more confidence for prediction. If some of them are removed, the algorithm's prediction power either doesn't decrease or doesn't rely on a few specific features.

Our studies and trials have shown that two-dimensional convolution layers work better than their one-dimensional counterparts because in this case, the filters can detect more drug similarities, and it is possible to extract more powerful Features. Therefore, the 1136-dimension feature vectors are transformed into matrices with dimensions of 71 16 times. Figure 4 shows the number of learnable weights for each layer. Also, the total number of weights is calculated, which indicates the general complexity of the model.

The following settings are used in the construction of the convolution neural network:

```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 13, 68, 128)       2176
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 10, 65, 32)        65568
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 7, 62, 8)          4104
_________________________________________________________________
flatten_1 (Flatten)          (None, 3472)              0
_________________________________________________________________
dense_1 (Dense)              (None, 64)                222272
_________________________________________________________________
dropout_1 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_2 (Dense)              (None, 16)                1040
_________________________________________________________________
dense_3 (Dense)              (None, 2)                 34
=================================================================
Total params: 295,194
Trainable params: 295,194
Non-trainable params: 0
```

Figure 4. Learnable parameters of two-class Neural Networks

1) We used Tensorflow Abadi et al. (2016) (version 1.14.0) and KERAS Chollet et al. (2015) (version 2.2.5) packages to implement the neural network.

2) The categorical-cross entropy loss function was considered an objective function for the neural network, which is generally used to train a classification network Ghosal et al. (1997); Toda and Okura (2012); Seen (2012).

3) ADAM optimization Kingma and Ba (2014) was used to manipulate the neural network weights to find a promising optimal (minimum) state of the loss function.

4) The number of epochs was considered 5.

5) Learning rate of 1.0×10^{-5} was used.

Keep in mind that this network's hyperparameters are not optimized, and the specified parameters are not necessarily at their best. There are two reasons for not optimizing hyperparameters:

1) Model overfitting: If hyperparameters changed to the best values, it is expected that the model will get better results on the present data, but there is no guarantee that the extracted features by the model are significant and works well when used in new cases. In this case, the so-called model is over-fitted and will be a negative point for the model.

2) Robustness: Optimal hyperparameters give better results for the present data, but different drug similarities may be used in the future, or new data may be collected, and the present results may not be repeated. In this case, the model loses its robustness and will not be accepted in the pharmaceutical and pharmacological community.

Finally, we examine the results of the proposed model in the 10-fold CV from three views:

1)Accuracy:In a 10-fold CV,the model obtained $AUC = 0.97, AUPR = 0.93$ for degressive interactions , and $AUC = 0.97, AUPR = 0.99$ for enhansive interactions. These results indicate the high accuracy and detection power of the model.

2) Variance: The confidence interval for the reported values with a reliability coefficient above 95 percent was narrow and close to each other. Out of four reported confidence interval values, three values were

less than ± 0.002, and only for the degressive interaction, the AUPR was in the range of ± 0.005. The low amount of variance obtained from the model shows that the proposed model is robust.

3) Resolution capability: By plotting the output probability distribution diagram, as shown in Figure 5, it is clear that values +1 and -1 are well separated, and probability distribution degressive and enhancive have slightly Subscriptions.

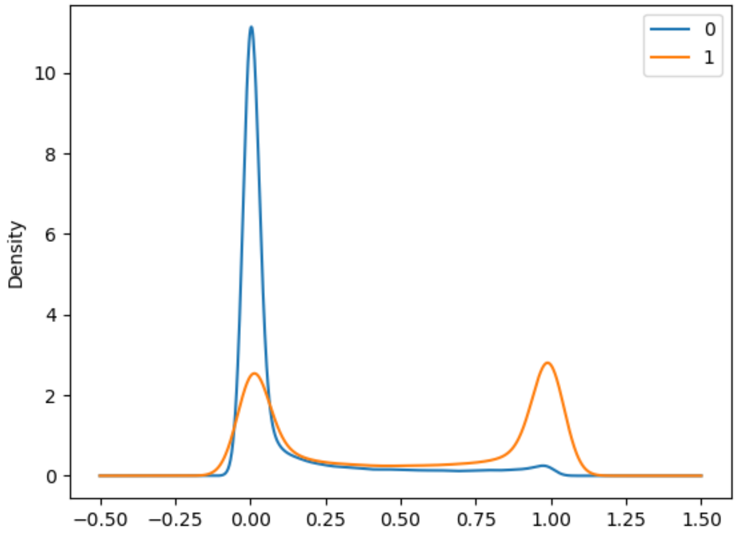

Figure 5. probability density distribution diagram of Degressive and enhancive. In this figure, 0 is the same as the -1label, and 1 is the same as $+1$.

The Suducode 1 shows the step-by-step model selection process.

Algorithm 1 Model selection suducode

Input: +1 and -1 drug pairs features
Output: +1 and -1 diagnostic model

1: Apply 10-fold CV to the features of +1 and -1 drug pairs.
2: Select the right model.
3: Test the model results in 10-fold CV.
4: If 3 is correct then select the model, else go to 2.

Detection of drug pairs without possible interaction

In the previous step, a high-precision, robust, and accurate model has been presented to detect drug pairs' potential interactions for both degressive and enhancive. Therefore, this model has the ability to detect non-interactions (real zeros) as follows. If drug pairs are unlikely to interact, then those drug pairs are likely to be real zeros.

According to this hypothesis, the model was used to predict all unknown drug pairs (zeros). Unknown drug pairs include 270,000 drug pairs. We consider drug pairs as non-interacting drug pairs in the model's output if the enhancive and degressive probability are less than 0.4 and 0.4. Among the unlabeled data, about 65,000 drug pairs had these conditions. These drug pairs are candidates for non-interaction. Due to the model's high accuracy, the low variance of results, and the model's high resolution, we consider these pairs non-interaction drug pairs.

2.7 Selecting and training model on known and unknown interactions

This section uses known data and potential non-interaction candidates to form a data set. Here, we use the non-interaction candidate drug pairs as real zeros. The recommender system presented in Section 2.6 is also used for the final model.

First, the B matrix rows, which contain the +1 and -1 interactions, are separated according to the previously detailed procedure and placed in 10 parts. Then, 30,000 non-interacting candidate drug pairs were randomly selected from 65,000 drug pairs. In the chosen drug pairs, the drug pairs and the duals of them must be non-interaction candidates. The zeros group is randomly divided into 10 parts, so each drug pairs and the dual are in the same batch. Then 10 parts of zeros are merged with 10 parts of pre-prepared +1s and -1s.

The data set contains approximately 72,702 drug pairs, divided into relatively equal parts, is ready to use in the training and testing of the final recommender system.

2.8 Selecting final model

The final model is almost the same as that described in Section 2.6. That means it has three Conv-layer with 128, 32, and 8 filters. Then, as before, three fully connected Conv layers were used. The difference was that the number of nodes changed from 64, 16 and 2 to 64, 16, and 3 in each layer, respectively. Specifically, this model gives three possible outputs for the three modes of enhancive interaction, non-interaction, and depressive interaction. Also, the number of epochs was considered 9. The deep neural network model for predicting interaction is shown in Figure 6. At this stage, the new model was not chosen because:

1) The power of this model for relatively accurate detection of enhancive and degressive interactions has been proven.

2) The zeros used in this section are just suggested and have not been approved by the Pharmacology Laboratory. Until the writing of this article, a comprehensive database for non-interaction cases has not been made public. If the model selection is made again, a model may be selected that is not necessarily valid in real-world application and hard to accept.

Due to the above reasons, the zeros recommender system is used to comprehensive drug-drug interactions prediction by changing the number of outputs from 2 to 3 as the input data. The proposed SNF-CNN method's general process is presented in the form of pseudocode 2, which includes the steps of preparation, model selection, real zero detection, and the presentation of a comprehensive recommender system.

3 RESULT AND DISCUSSION

3.1 Assessment

K-Fold cross-validation (CV) is a well-proven approach to verify the algorithms' accuracy in machine learning. To demonstrate that new drugs have no interaction and avoid overly optimistic predictions, the CV equation C must be carefully designed. For drug pairs with no known interactions, the K-fold CV scheme seeks to assess the task of anticipating new types of potential interactions between them and drugs that have known interactions. The production of test and experimental samples is as follows:

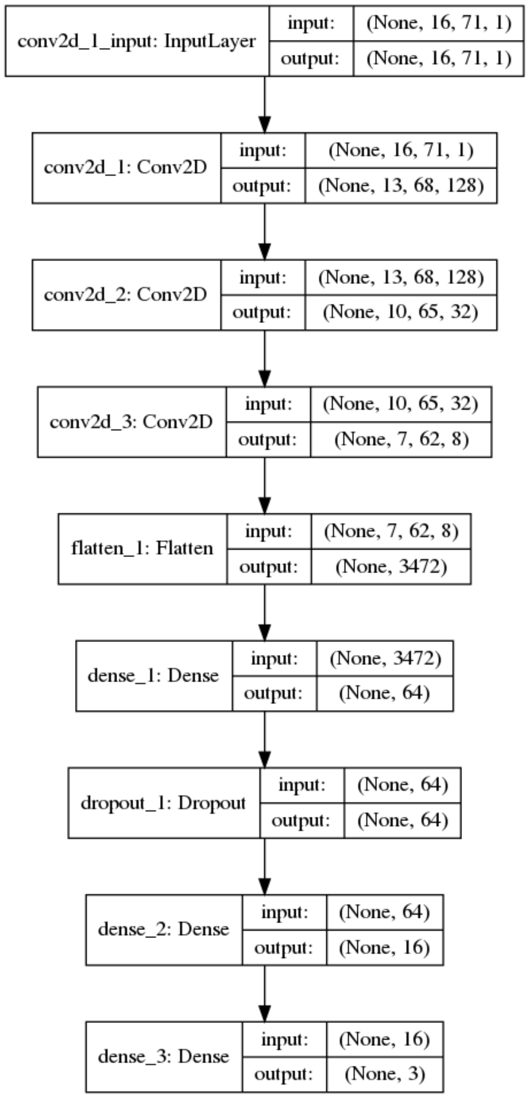

Figure 6. Arrangement of neural network layers SNF-CNN Predict triple-class interaction. Non-interaction (0), degressive interaction (-1) and enhancive interaction (+1)

Algorithm 2 Final model selection(SNF-CNN) suducode

Input: Drug pairs features(+1,-1,real 0)
Output: Diagnostic model for interaction and non-interaction

1: Calculate drug similarity matrices with the cosine method.
2: Integrate drug similarity matrices with the similarity network fusion(SNF) method.
3: Built the input matrix of the model.
4: Select the fit known interactions model and train it.
5: Predict probable zeros by using step 4.
6: Select the fit known interactions and zeros of step 5 model , and train it.
7: Predict on unknown drug pairs.

The whole data set is divided into k equal parts. The k-1 part is used as a training data set, and the model is built based on it, and the testing is performed with the remaining part. The procedure is repeated k times so that each of the k parts is used only once for testing, and each time accuracy is calculated for the constructed model. In this method, the average of predicting accuracy in all rounds of K-fold CV is taken as the final accuracy Classifier. The most common value for k in scientific literary is 5 or 10. Obviously, the larger the value of k, the more reliable the calculated accuracy classifier, and the more comprehensive the knowledge obtained, and the longer the classifier's testing time, which is the most important problem. Each setting and each data set has its own validation. In this approach, according to the type of problem and the methods, we used two types of 10-fold CV to divide the data into two sets of testing and training, which are:

3.1.1 First case: 10-fold CV without unknown interactions

In this case, we randomly select 90 percent of the enhancive and degressive interactions. For the testing set, we consider the remaining 10 percent of the enhancive and degressive interactions. In the first case of the testing procedure, the model was selected, and some hyper-parameters, such as the number of epochs, were determined.Figure 7 It shows the training process for the selected model. As expected, the model's accuracy is strict on ascending training data, but there are ups and downs for testing data after Epoch 5. In the loss function graph, by the end of epoch 5, as the epochs increase, the loss function's value on training and testing procedure decreases. After epoch 5, the trend of training data continues, but the testing trend is reversed. In other words, over feet occurs. Therefore, based on the graphs, the appropriate number of epochs in this step was considered 5.

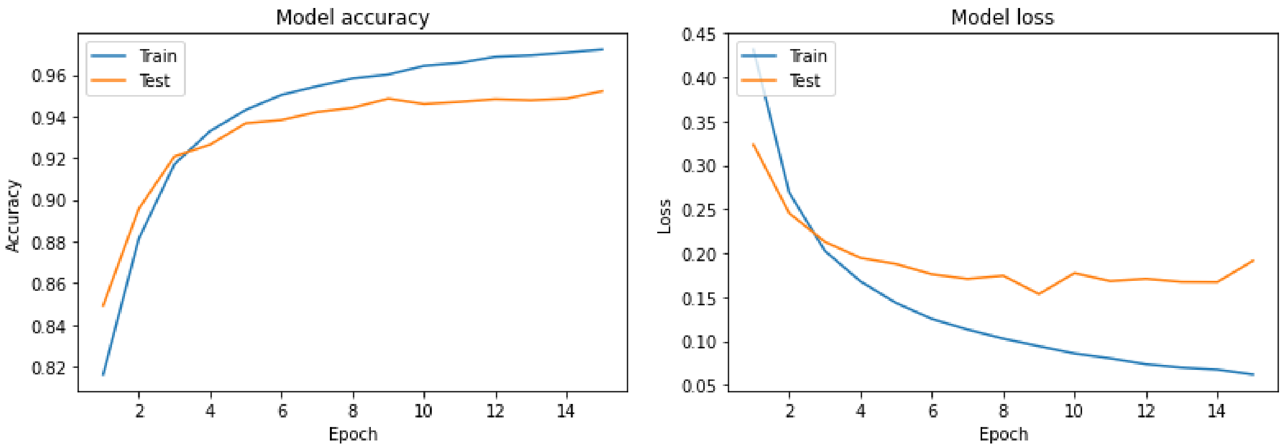

Figure 7. Accuracy and loss function for the binary model: The right figure shows the model's accuracy on training and validation data during 15 epochs, and the left figure shows the loss function values at different epochs.

3.1.2 Second case: 10-fold CV with unknown interactions

In this case, we divide the set of all interactions (enhancive, degressive, and zeros of the first step) into 10 equal parts. We consider one part of the testing set and the other 9 parts as the training data set. Divide all the zeros in the previous step into 10 parts and add a 1 to 9 ratio to both testing and training sets. In the second case, the previous model's 10-fold CV procedure was trained with the least changes to predict the three classes. Besides, hyper-parameter, the number of epochs was determined. Figure 8

It shows the training process. The process of the accuracy of the model on training data increases steadily with the increase of epochs. Still, the model after epoch 9 reduces a constant and decreases the accuracy a little for testing data.

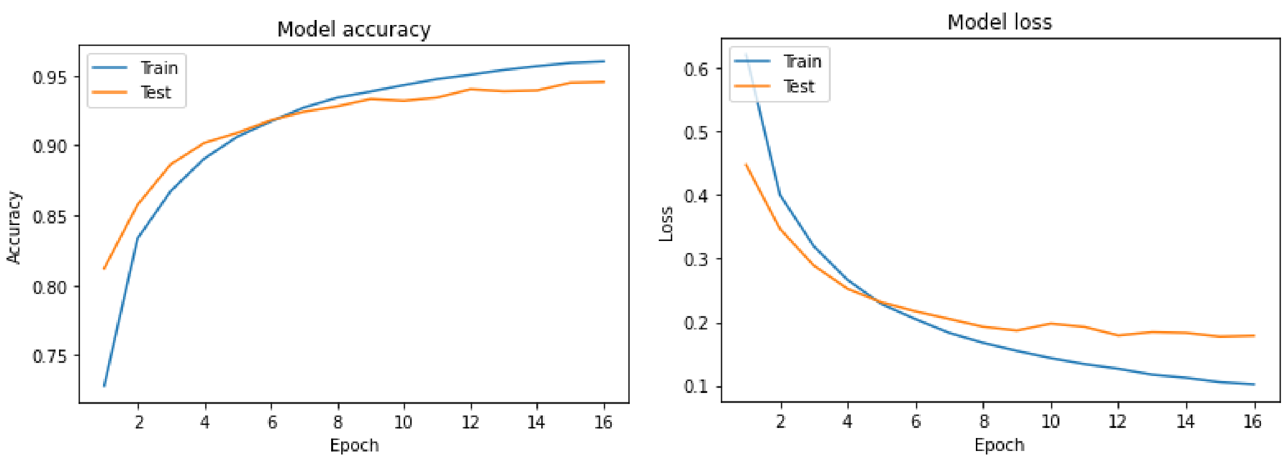

Figure 8. Accuracy and loss function diagrams for the triple model: The right figure shows the accuracy of the model on training and validation data during 16 epochs, and the left figure shows the loss function values at different epochs.

4 EVALUATION CRITERIA

In this study, we classify drug pairs according to the type of interaction or non-interaction. So in order to compare the method performance with other existing methods, from four measurement criteria, F- measure, accuracy, Area Under Roc Curve(AUC), and Area Under Precision-Recall Curve(AUPR) Used. To define these criteria, at first, introduced four counting criteria in table **??**.

		Actual Condition		
	Total Smpelse	Actual Positive	Actual Negative	
output of classifier	Classify Positive	TP	FP	PPV(Precision)
	Classify Negative	FN	TN	
		TPR(Recall)	TNR(Specifcity)	ACC F-measure MCC

Table 1. The confusion matrix and relevant evaluation index.True Positive (TP): The number of residues classified as interacting correctly, False Positive (FP): The number of residues classified as interacting correctly incorrectly, False Negative (FN): The number of residues classified as non-interacting incorrectly, True Negative (TN): The number of residues classified as non-interacting correctly.

By using the table 1, four evaluation criteria are defined in the following order: Acurracy:The fraction of actual predicted interactions (TP and TN) to all predicted interactions.

$$\text{Acurracy} = \frac{\text{TP} + \text{TN}}{\text{all predicted interactions}}$$

Precision:The fraction of correct predicted interactions among all predicted interactions.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall:The fraction of correct predicted interactions among all true interactions.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Precision and recall have a trade-off; thus, improving one of them may lead to a reduction in another.Therefore, utilizing F-measure, is more reasonable.

F-measure:the geometric mean of precision and recall.

$$F - measure = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

It should be noted that if the interaction of two drugs is assigned to zero, it denotes that no evidence of their interaction has been found yet; thus, they may interact with each other. So we cannot identify TN and FP pairs correctly. The training process requires both interaction and non-interaction samples. Therefore, some of the zero assigned pairs are considered as non-interactive pairs in the training model. So every method may have some FP in its evaluations. This leads to a reduction in calculated precision and F-measure, while the real values of precision and F-measure may be higher.

Since the values of precision, recall, and F-measure is dependent on the value of the threshold, we also evaluate methods via AUC which is the area under the receiver operating characteristic (ROC) curve, and AUPR, that is the area under the precision-recall curve. These criteria indicate the efficiency of methods independent of the threshold value. In cases that the fraction of negative samples and positive samples are not equal, AUPR is the fairer criterion for evaluation.

5 COMPARISON OF RESULTS

Based on the validation procedure described in Section **??** , the interaction type detection model was selected and trained. Then, the final three-class model was presented by recognizing the most probable non-interactions. So we tested the SNF-CNN model on data to check the evaluation, robustness, and efficiency of the CV. Results SNF-CNN and other methods for comparison are presented and discussed in this section. Before comparing the different methods, an example of the results of neural network implementation is presented to identify the type of degressive and enhancive interactions.

	Precision	Recall	F-measure	Support
Degressive	0.94	0.83	0.88	850
Enhancive	0.95	0.99	0.97	3052
Accuracy			0.95	3902
Macro Avg	0.95	0.91	0.93	3902
Weighted Avg	0.95	0.95	0.95	3902

Table 2. Interaction type classification report

Table 2 an example of a model implementation result is the ability of the model in terms of precision, recall and F-measure Indicates the type of interactions. According to table 2 the precision of the model in detecting enhancive and degressive interactions is 95 percent and 94 percent, while recall is 99 percent and 83 percent, respectively. the F-measure is also 97 percent and 88 percent that the higher ability of the model to detect degressive interactions comes from a higher number of these types of interactions. The ratio of degressive interaction to enhancive interaction is approximately 4 to 1.

	Precision	Recall	F-measure	Support
Enhancive	0.88	0.84	0.86	850
Non-intraction	0.96	0.95	0.96	3000
Degressive	0.95	0.97	0.96	3052
Accuracy			0.95	6902
Macro Avg	0.93	0.92	0.93	6902
Weighted Avg	0.95	0.95	0.95	6902

Table 3. Three-Classes interaction classification report

Also in the table 3 three-Classes interaction classification is displayed. In this implementation, the accuracy of the model for detecting degressive interaction, non-interaction, and enhancive interaction is 95 percent, 96 percent, and 88 percent, respectively. The recall is 97 percent, 95 percent, and 84 percent, respectively, and finally F-measure It is 96, 96, and 86 percent. In comparison, the model power in the three-classes decreases slightly to the two-classes, which can be due to two reasons.

1) The problem of three-classes is more difficult than two-classes.

2) Zeros or non-interaction are not necessarily real and are not pharmacologically proven, so there is a possibility of some disturbance.

For the above reasons, some reduction in the power of the three-classes model in detection was not unexpected.

	AUC	AUPR
degressive	0.9747 ± 0.0033	0.9666 ± 0.0045
enhancive	0.9686 ± 0.0028	0.8221 ± 0.0184
non-interaction	0.9714 ± 0.0040	0.9480 ± 0.0083

Table 4. Results of SNF-CNN algorithm in predicting three-classes based on AUC and AUPR criteria and their confidence interval

Since the previous algorithms in the three-classes prediction of DDI used AUC and AUPR therefore, the results of the proposed algorithm are based on these two criteria according to table 4.Also in table 4 for the algorithm presented in this research, high and low intervals are reported with 95 percent confidence, which shows that the results of the algorithm have slightly changed in the 10- fold CV and the proposed algorithm is robust and reliable.

In the table 5 results of SNF-CNN averaged for the three classes and compared with other existing three-classes algorithms. According to the table 5 the proposed algorithm has a high difference compared to other superior algorithms with the problem of ternary and has been able to challenge other algorithms.

	AUC	AUPR
SNF-CNN	0.971	0.912
BRSNMFShi et al. (2019)	0.645	0.346
Semi-NMF Yu et al. (2018)	0.796	0.579
TMFUFShi et al. (2018)	0.842	0.526

Table 5. Comparison of the results of three-classes prediction algorithms based on criteria AUC and AUPR

6 CONCLUSIONS

Existing computational approaches are able to present potential large-scale interactions before using drugs on the market. However, they cannot predict comprehensive interactions, including degressive and enhancive interactions. It is more useful to know whether or not a drug pair is enhancive DDI or a degressive DDI than to know whether or not a drug pair is DDI. Without considering the pharmacological changes caused by DDIs, most existing approaches only report two-classes prediction. In addition, the occurrence of degressive and enhancive DDIs is not random, but none of the existing approaches investigates and leverages this intrinsic important property of interactions when treating complex diseases (including treatment with three or more drugs).

In this work, after representing a comprehensive DDI network, we used the structure of recommender systems to design a novel algorithm. Although the prediction obtained by the new algorithm is inspiring, overall performance can still be improved. For this reason, we investigate those incorrectly predicted DDIs. After checking them case-by-case, and in order to prove the algorithm in practice, check the prediction performance of the algorithm in the latest version of the DrugBank database. Observations and investigations led to the discovery of three reasons for wrong predictions:

The first is named as false-positive drug pairs, which are precisely labeled as DDIs in DrugBank version 4, but in the current version, they are correctly labeled as non-DDIs. For example, the old version of DrugBank records that Apraclonidine (a Sympathomimetic drug used in glaucoma therapy) increases the atrioventricular blocking activities of Alprenolol and Bevantolol, while the newer version removes it.

The second one is false-negative drug Pairs, which are wrongly labeled as non-DDIs in DrugBank version 4 but in the current version, they are reported as DDIs.For example, the pair of Valrubicin and Cyclosporine as well as the pair of Ergocalciferol and Calcitriol in the newer version of DrugBank reports, Valrubicin (bladder cancer treatment drug) Increases the activity of the nephrotoxic drug Cyclosporine(A drug that suppresses the immune system with a special action on T-lymphocytes), while the combined therapy of Calcitriol and Ergocalciferol increases the risk or severity of adverse effects in the multiple-drug therapy.

The last one refers to change Pair DDIs, which are labeled as enhancive DDIs in DrugBank version 4, but in the current version, are labeled as degressive DDIs, and vice versa.

It is anticipated that the SNF-CNN approach, will be able to achieve better DDI prediction by the better datasets with fewer of both false-positive and false-negative drug pairs. For future work, it is recommended that the dataset always be collected from the latest version of DrugBank.

Three-classes data is an attempt to improve expression and problem solving over two-classes data. However, three-classes data does not have sufficient biological significance and provides limited biological information. This means that predicting the type of DDI can be useful, but it is not clear at what stage of the pharmacokinetic or pharmacodynamic stages this DDI occurred. Therefore, it is suggested to collect datasets with degressive and enhancive labels from each of the pharmacokinetic and pharmacodynamic steps. In this case, more meaningful models in terms of pharmacology and machine learning are taught. The resulting models are more important to pharmacists and pharmacists and will be more useful.

6.1 Figures

Frontiers requires figures to be submitted individually, in the same order as they are referred to in the manuscript. Figures will then be automatically embedded at the bottom of the submitted manuscript. Kindly ensure that each table and figure is mentioned in the text and in numerical order. Figures must be of sufficient resolution for publication see here for examples and minimum requirements. Figures which are not according to the guidelines will cause substantial delay during the production process. Please see here for full figure guidelines. Cite figures with subfigures as figure 10B.

6.1.1 Permission to Reuse and Copyright

Figures, tables, and images will be published under a Creative Commons CC-BY licence and permission must be obtained for use of copyrighted material from other sources (including re-published/adapted/modified/partial figures and images from the internet). It is the responsibility of the authors to acquire the licenses, to follow any citation instructions requested by third-party rights holders, and cover any supplementary charges.

6.2 Tables

Tables should be inserted at the end of the manuscript. Please build your table directly in LaTeX.Tables provided as jpeg/tiff files will not be accepted. Please note that very large tables (covering several pages) cannot be included in the final PDF for reasons of space. These tables will be published as Supplementary Material on the online article page at the time of acceptance. The author will be notified during the typesetting of the final article if this is the case.

7 NOMENCLATURE

7.1 Resource Identification Initiative

To take part in the Resource Identification Initiative, please use the corresponding catalog number and RRID in your current manuscript. For more information about the project and for steps on how to search for an RRID, please click here.

7.2 Life Science Identifiers

Life Science Identifiers (LSIDs) for ZOOBANK registered names or nomenclatural acts should be listed in the manuscript before the keywords. For more information on LSIDs please see Inclusion of Zoological Nomenclature section of the guidelines.

8 ADDITIONAL REQUIREMENTS

For additional requirements for specific article types and further information please refer to Author Guidelines.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

The Author Contributions section is mandatory for all articles, including articles by sole authors. If an appropriate statement is not provided on submission, a standard one will be inserted during the production process. The Author Contributions statement must describe the contributions of individual authors referred to by their initials and, in doing so, all authors agree to be accountable for the content of the work. Please see here for full authorship criteria.

FUNDING

Details of all funding sources should be provided, including grant numbers if applicable. Please ensure to add all necessary funding information, as after publication this is no longer possible.

ACKNOWLEDGMENTS

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

SUPPLEMENTAL DATA

Supplementary Material should be uploaded separately on submission, if there are Supplementary Figures, please include the caption in the same file as the figure. LaTeX Supplementary Material templates can be found in the Frontiers LaTeX folder.

DATA AVAILABILITY STATEMENT

the code and data is available at GitHub page of SNF-CNN code and data.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*. 265–283

Bui, Q.-C., Sloot, P. M., Van Mulligen, E. M., and Kors, J. A. (2014). A novel feature-based approach to extract drug–drug interactions from biomedical text. *Bioinformatics* 30, 3365–3371

Businaro, R. (2013). Why we need an efficient and careful pharmacovigilance? *Journal of pharmacovigilance*

Cheng, F. and Zhao, Z. (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, e278–e286

[Dataset] Chollet, F. et al. (2015). Keras

Fu, L. and Peng, Q. (2017). A deep ensemble model to predict mirna-disease association. *Scientific reports* 7, 1–13

Ghosal, T., Edithal, V., Ekbal, A., Bhattacharyya, P., Chivukula, S. S. S. K., and Tsatsaronis, G. (1997). Is your document novel? let attention guide you. an attention based model for document level novelty detection. *Natural Language Engineering* 1, 1–38

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine* 29, 82–97

Huang, Q.-R., Hu, F., Huang, S., Li, H.-X., Yuan, Y.-H., Pan, G.-X., et al. (2009). Effect of long-term fertilization on organic carbon and nitrogen in a subtropical paddy soil. *Pedosphere* 19, 727–734

Huang, S.-M., Temple, R., Throckmorton, D., and Lesko, L. (2007). Drug interaction studies: study design, data analysis, and implications for dosing and labeling. *Clinical Pharmacology & Therapeutics* 81, 298–304

Karbownik, A., Szałek, E., Sobańska, K., Grabowski, T., Wolc, A., and Grześkowiak, E. (2017). Pharmacokinetic drug-drug interaction between erlotinib and paracetamol: a potential risk for clinical practice. *European Journal of Pharmaceutical Sciences* 102, 55–62

Kim, Y.-A., Cho, D.-Y., and Przytycka, T. M. (2016). Understanding genotype-phenotype effects in cancer via network approaches. *PLoS computational biology* 12, e1004747

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

Koch-Weser, J. (1981). Serum drug concentrations in clinical perspective. *Therapeutic drug monitoring* 3, 3–16

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., et al. (2014). Drugbank 4.0: shedding new light on drug metabolism. *Nucleic acids research* 42, D1091–D1097

Leape, L. L., Bates, D. W., Cullen, D. J., Cooper, J., Demonaco, H. J., Gallivan, T., et al. (1995). Systems analysis of adverse drug events. *Jama* 274, 35–43

Liu, S., Chen, K., Chen, Q., and Tang, B. (2016). Dependency-based convolutional neural network for drug-drug interaction extraction. In *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)* (IEEE), 1074–1080

Luo, H., Zhang, P., Huang, H., Huang, J., Kao, E., Shi, L., et al. (2014). Ddi-cpi, a server that predicts drug–drug interactions through implementing the chemical–protein interactome. *Nucleic acids research* 42, W46–W52

Mulroy, E., Highton, J., and Jordan, S. (2017). Giant cell arteritis treatment failure resulting from probable steroid/antiepileptic drug-drug interaction. *The New Zealand Medical Journal (Online)* 130, 102

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

Olayan, R. S., Ashoor, H., and Bajic, V. B. (2018). Ddr: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. *Bioinformatics* 34, 1164–1173

Pahikkala, T., Airola, A., Pietilä, S., Shakyawar, S., Szwajda, A., Tang, J., et al. (2015). Toward more realistic drug–target interaction predictions. *Briefings in bioinformatics* 16, 325–337

Pan, X., Fan, Y.-X., Yan, J., and Shen, H.-B. (2016). Ipminer: hidden ncrna-protein interaction sequential pattern mining with stacked autoencoder for accurate computational prediction. *BMC genomics* 17, 582

[Dataset] Ross Markello (2018). snfpy 0.2.2

Ryu, J. Y., Kim, H. U., and Lee, S. Y. (2018). Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the National Academy of Sciences* 115, E4304–E4311

Seen, S. (2012). 1-day learning, 1-year localization: Long-term lidar localization using scan context image. *Database* 2012, 01–15

Shi, J.-Y., Huang, H., Li, J.-X., Lei, P., Zhang, Y.-N., Dong, K., et al. (2018). Tmfuf: a triple matrix factorization-based unified framework for predicting comprehensive drug-drug interactions of new drugs. *BMC bioinformatics* 19, 27–37

Shi, J.-Y., Huang, H., Li, J.-X., Lei, P., Zhang, Y.-N., and Yiu, S.-M. (2017). Predicting comprehensive drug-drug interactions for new drugs via triple matrix factorization. In *International Conference on Bioinformatics and Biomedical Engineering* (Springer), 108–117

Shi, J.-Y., Mao, K.-T., Yu, H., and Yiu, S.-M. (2019). Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization. *Journal of cheminformatics* 11, 1–16

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

Tian, Z., Guo, M., Wang, C., Xing, L., Wang, L., and Zhang, Y. (2017). Constructing an integrated gene similarity network for the identification of disease genes. *Journal of biomedical semantics* 8, 32

Toda, Y. and Okura, F. (2012). Research article how convolutional neural networks diagnose plant disease. *Challenge (LSVRC)* , 20

Veith, H., Southall, N., Huang, R., James, T., Fayne, D., Artemenko, N., et al. (2009). Comprehensive characterization of cytochrome p450 isozyme selectivity across chemical libraries. *Nature biotechnology* 27, 1050–1055

Vilar, S., Uriarte, E., Santana, L., Lorberbaum, T., Hripcsak, G., Friedman, C., et al. (2014). Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nature protocols* 9, 2147

Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., et al. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nature methods* 11, 333

Wang, Y., Liu, T., Xu, D., Shi, H., Zhang, C., Mo, Y.-Y., et al. (2016). Predicting dna methylation state of cpg dinucleotide using genome topological features and deep networks. *Scientific reports* 6, 19598

Wienkers, L. C. and Heath, T. G. (2005). Predicting in vivo drug interactions from in vitro drug discovery data. *Nature reviews Drug discovery* 4, 825–833

Wiśniowska, B. and Polak, S. (2016). The role of interaction model in simulation of drug interactions and qt prolongation. *Current pharmacology reports* 2, 339–344

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M. (2008). Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24, i232–i240

Yu, H., Mao, K.-T., Shi, J.-Y., Huang, H., Chen, Z., Dong, K., et al. (2018). Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC systems biology* 12, 14

Zhang, P., Wang, F., Hu, J., and Sorrentino, R. (2015). Label propagation prediction of drug-drug interactions based on clinical side effects. *Scientific reports* 5, 1–10

Zhang, W., Chen, Y., Li, D., and Yue, X. (2018). Manifold regularized matrix factorization for drug-drug interaction prediction. *Journal of biomedical informatics* 88, 90–97

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

Zhang, Y., Wu, H.-Y., Xu, J., Wang, J., Soysal, E., Li, L., et al. (2016b). Leveraging syntactic and semantic graph kernels to extract pharmacokinetic drug drug interactions from biomedical literature. *BMC systems biology* 10, 67

Zhao, X.-M., Iskar, M., Zeller, G., Kuhn, M., Van Noort, V., and Bork, P. (2011). Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput Biol* 7, e1002323

Zhou, D., Bui, K., Sostek, M., and Al-Huniti, N. (2016). Simulation and prediction of the drug-drug interaction potential of naloxegol by physiologically based pharmacokinetic modeling. *CPT: pharmacometrics & systems pharmacology* 5, 250–257

FIGURE CAPTIONS

Figure 9. Enter the caption for your figure here. Repeat as necessary for each of your figures

Figure 10. This is a figure with sub figures, **(A)** is one logo, **(B)** is a different logo.