Predicting Comprehensive Drug – Drug Interaction via Similarity Network Fusion and Convolutional Neural Networks

M.Amin Khodamoradi 1* , Bahareh Levian 2,3 and Changiz Eslahchi 2,3

1-**Universidade NOVA de Lisboa, NOVA School of Science and Technology (FCT NOVA) / Uninova, Center of Technology and Systems**
2-Department of Computer Science, Faculty of Mathematical Science, Shahid Beheshti University, Tehran, Iran.
3-School of Bioinformatics, IPM - Institute for Research in Fundamental Sciences, Tehran, Iran.

Abstract

Background: Drug-drug interactions (DDIs) usually cause unexpected and even adverse drug reactions. It is important to identify DDIs before using them 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. Prediction of comprehensive DDIs and the discovery of structural relationships in the DDI graph play important guidance when making a prescription of multiple drugs.

Results: 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 or degressive of a pair of drugs based on similarity network fusion and convolutional neural networks. SNF-CNN achieves the depressive DDI prediction (*AUC = 0.975 and AUPR = 0.967*), enhancive DDI prediction (*AUC = 0.969 and AUPR = 0.822*) and the Unknown DDI prediction (*AUC = 0.971 and AUPR = 0.948*). Compared with three state-of-the-art methods on a dataset, SNF-CNN shows superiority.

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

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

Introduction

When two or more drugs are taken together, drugs' effects or behaviors are unexpectedly influenced by each other [1]. 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 known DDIs [2]. The DDIs would put patients, who are treated with multiple drugs, in an unsafe situation [3, 4, 5, 6]. Understanding DDI is the first step in drug combinations, which becomes one of the most promising solutions for the treatment of multi factorial

complex diseases [7]. 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 [8] or transporter-associated interactions [9]) face challenges, such as high costs, long duration, animal welfare considerations [10], 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 [11, 12]. Data mining-based computational approaches have been developed to detect DDIs from various sources [10] , such as scientific literature [13, 14] , electronic medical records [15] , 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 [16] , Network Recommendation-Based [10], Classification-Based [17]) can provide such alerts by utilizing pre-marketed or post-marketed drug attributes, such as drug features or similarities [18]. These methods use different drug features to predict DDIs, such as chemical structures [16] , targets [19] , hierarchical classification codes [17] , side effects, and off-label side effects [10, 20].

A Dependency-based Convolutional Neural Network (DCNN) has proposed for drug-drug interaction extraction by Liu and et al. in 2016 [21]. 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 [22], 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 drug-drug or drug-food constituent 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 problem. Similarity Network Fusion (SNF) [23] is a competent method to integrate various similarities, which is used in numerous biological contexts [24, 25, 26]. The neural network is a strongly developed approach that provides satisfactory solutions, especially for large datasets and nonlinear analyzes [27], which is widely used in critical problems [28, 29, 30].

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 [31].

Besides the occurrence of both enhancive and degressive DDIs is not random [32, 33], however 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 to understand how the DDIs occur. It is one of the most important steps for treating complex diseases [34] and guides physicians in preparing safer prescriptions to high-order drug interaction.

The recent works attempted to investigate two major issues: 1) predicting three-class DDIs instead of two-class prediction, 2) extracting the topological information of drugs in a DDI network.

Model of TMFUF [32] is proposed by Shi and et al. in 2018 which predicts enhancive and degressive DDIs for different predicting scenarios of new drugs (those with no known DDI). Proposed DDINMF model [33] in addition to predicting DDIs, asigns every drugs to drug communities. In result, some corelations are observed between drug communities and the numbers of enhancive, degressive, sum and difference of DDIs for each drug.

These observation shows that not only occurrence of enhancive or degressive DDIs are not randomly but also representes some topological features in DDI network. BRSNMF [35] model is a method based on Semi-NMF to predict the degressive and enhansive DDIs, more accurately, in cold start scenario [36]. This method exploits Drug Binding Protein (DBP) feature to map new drugs (without any known DDIs) with known drugs (drugs which has one DDI at least). Results show that BRSNMF defines drug communities with more moderate sizes by adding a regularization term to Semi-NMF objective function based on weakly balance theorem.

All three introduced algorithms are using matrix factorization methods, which are a network recommender-based approach. The matrix factorization approach, with slight modification, is a suitable solution for the subject of predicting DDI that has received much attention from researchers, but these methods do not work on potential DDIs which are crucially important in giving safer prescriptions.

In this paper, we firstly introduce data and features. Then, a novel algorithm 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 (SNF-CNN). The SNF-CNN trying to recognize unknown and potential DDIs which are not detected yet. The feature’s intrinsic of off-label side effect and chemical structure of drugs may provide information to find hidden potential DDIs in current DDI network. To use both of similarity features, we will take advantage of SNF.

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 DDI prediction results of SNF-CNN in the 10-fold cross-validation (CV) process in compare to other methods.

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.

Methods

Dataset and features

In this study, we use the dataset presented by Yu and et al. in 2018 [33]. This dataset 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} based on the off-label side effects provided by OFFSIDES. If a side effect or chemical structure is reported or has observed for the drug its vector element is one, otherwise is zero.

Problem formulation

Without loss of generality, let $D = \{d_i\}$, $i = 1, 2, .., m$, be a set of m approved drugs. In addition, each drug d i in the D is represented as a p-dimension feature vector $f_i = [f_1, f_2, .., f_k, .., fp]$, 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 feature vectors of the chemical structure and the off-label side effect. There are two feature matrices of F with dimensions of $m \times p$ (magnitude 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.

Drug-Drug interactions can be accordingly represented as an $m \times m$ symmetric interaction matrix $A_{m \times m} = \{a_{ij}\}$. For the conventional binary DDIs, $a_{ij} = 1$ if d i interacts with d_j , and $a_{ij} = 0$ otherwise. Matrix of comprehensive DDIs has three amounts ($a_{ij} \in \{-1, 0, +1\}$). Same as binary formulation, if d_i and d_j do not interact with each other, $a_{ij} = 0$ while there is an enhancive DDI or a degressive DDI between d_i and d_j , $a_{ij} = +1$ or $a_{ij} = -1$ respectively.

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 feature matrix data to be proper input for machine learning methods then we devise and train a deep learning model to predict potential interactions.

Similarity matrix calculation

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 d_i and d_j . Each element of S has a continuous value between zero and one.

A common method of calculating similarity called Cosine Similarity is used in machine learning articles such as [37, 38]. 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:

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

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

Integration drug similarity matrices

Similarity Network Fusion (SNF)[23] 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.

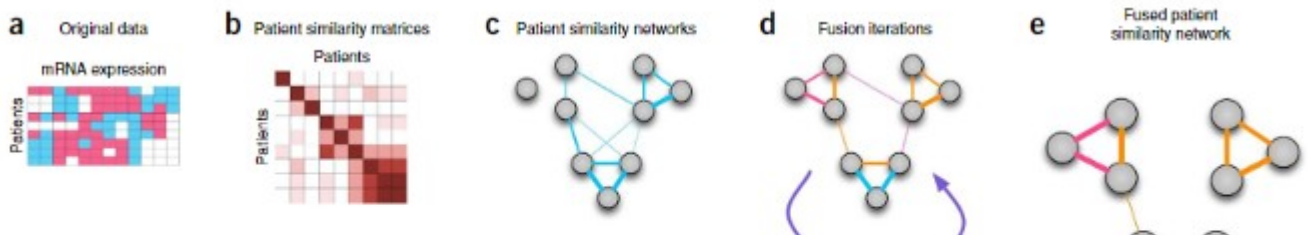
In this section, similarity matrices of the chemical structure and the off-label side effect of drugs were integrated via SNF method. The output of this integration is a new similarity matrix, S_{snf} has dimensions of 568 × 568, and elements of S_{snf} have a value between 0 and 1. To integrate the network similarity, the package of SNFPy is used, which is implemented in Python and is available at [39].

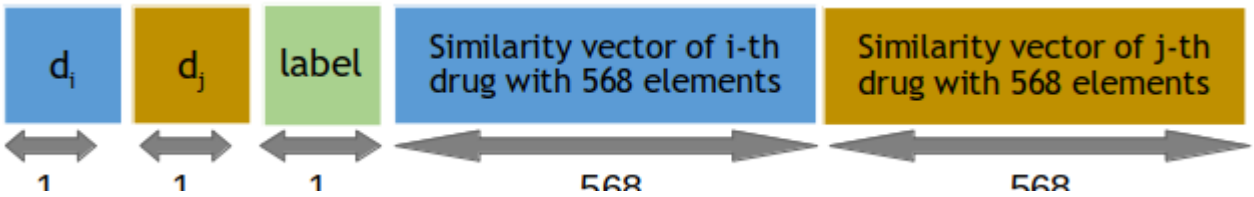
Input matrix format

At this stage, a matrix forms with 1139 columns and 322056 rows. Figure 2 shows the matrix header 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.

The data contains 568 drugs. Obviously, the interaction of a 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 B have been displaced, so these drug pairs are dual. The existence of both dual augments the training data, which increases the model's ability to have a better prediction. As a result, the matrix has 322056 data samples or rows (*(568 × 568) − 568 = 322056*). According to the explanation, a matrix with dimensions of *322056 × 1139* is formed to input into our model, which is called the B matrix.





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. The positive and negative DDIs have specific and real labels while the zero label does not mean that there is no interaction between a drug pair, it just indicates 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.

Assessment

K-Fold cross-validation (CV) is a well-proven approach to verify the algorithms' resolution ability, model selection, and feature engineering in machine learning. To demonstrate that the feature set is

informative enough, the selected model is robust and confident or the method has proper accuracy in comparison with other methods, the CV equation must be carefully designed. The production of test and experimental samples is as follows:
The whole data set has divided into K equal parts with consideration of dual pairs. Since biologically, the (d i , d j) and (d j , d i) drug pairs are the same we expect the model to predict their labels similarly. in the separation of training and testing data, necessarily a drug pair and its dual are in the same group to prevent unfair results. The K-1 parts are used as a training data set, and the model has built based on them, and the test has performed with a remaining part. This procedure has repeated K times so that each of the K parts has used only once for testing, and each time a resolution metric has calculated for the constructed model. In this method, the average prediction resolution metric in all K rounds is taken as the final resolution metric for the classifier. The most common value for K in scientific literary is 5 or 10. Obviously, the more detailed validation in the K-fold CV the more reliable the classifier accuracy, the more comprehensive the obtained knowledge, and the more time-consuming the validation process.

	Actual Enhancive	Actual Degressive
Classified Enhancive	TP	FP
Classified Degressive	FN	TN

Table 1: The confusion matrix for interaction type (Degressive or Enhancive) and relevant evaluation index. True Positive (TP): The number of drug pairs classified as enhancive interaction correctly, False Positive (FP): The number of drug pairs classified as enhancive interaction incorrectly, False Negative (FN): The number of drug pairs classified as degressive interaction incorrectly, True Negative (TN): The number of drug pairs classified as degressive interaction correctly.

If consider degressive interaction as Negative (N) and enhancive interaction as Positive (P) sample then the confusion matrix for interaction type (Degressive or Enhancive) and relevant evaluation index is as shown in Table 1. By using Table 1, four evaluation criteria are defined in the following order:
Accuracy: The fraction of all correct predictions (TP and TN) to all predictions.

$$Acurracy = TP + TN / (TP + FP + TN + FN)$$

Precision: The fraction of correct predicted (enh/deg) interactions among all predicted (enh/deg) interactions.

$$Precision_{(enh/deg)} = TP / (TP + FP)$$

Recall: The fraction of correct predicted (enh/deg) interactions among all true (enh/deg) interactions.

$$Recall_{(enh/deg)} = TP / (TP + FN)$$

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.

$$F - measure_{(enh/deg)} = 2 \times Precision_{(enh/deg)} \times Recall_{(enh/deg)} / (Precision_{(enh/deg)} + Recall_{(enh/deg)})$$

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.

Selecting and training model on known interactions

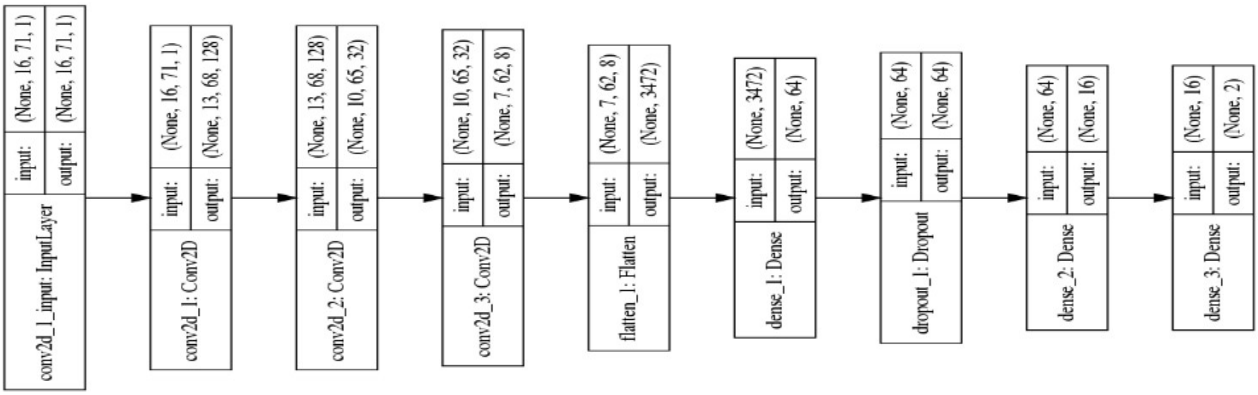
To solve this problem, it is necessary to provide a model that detects non-interaction with high resolution and confidence. Therefore, we design a model based on deep learning that predicts the possible non-interaction drug pairs and then use it to design a three-class model. Obviously, high resolution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

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 dual are in the same group. This prevents unfair results.

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 [40], which is defined as the positive part of its argument:

ReLU (x) = max{x, 0} *(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 [41], which is calculated as follows:

Sigmoid(x) = 1 / (1 + e −x) *(3)*

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 [42] 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.

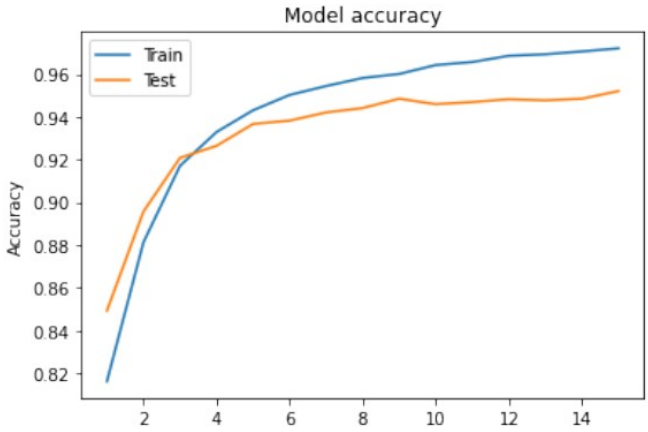
Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 13, 68, 128)	2176
conv2d_2 (Conv2D)	(None, 10, 65, 32)	65568
conv2d_3 (Conv2D)	(None, 7, 62, 8)	4104
flatten_1 (Flatten)	(None, 3472)	0

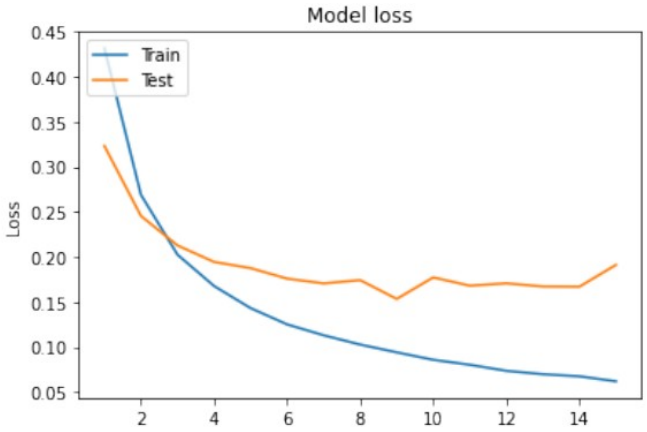
The following settings are used in the construction of the convolution neural network:

1. We used Tensorflow [43] (version 1.14.0) and KERAS [44] (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 [45, 46, 47].
3. ADAM optimization [48] 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.

Two-class model's training trend

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 5, 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.





Hyperparameters Optimization

Keep in mind that 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.

Evidence of reliability of the two-class model

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

1) Model Resolution: 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 resolution and detection power of the selected model. The selected model resolution results are presented in Table 2 which identifies the type of degressive and enhancive interactions.

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

Table 2: Interaction type classification report

Table 2 is an example result of implemented model which shows 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.
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) Separability: By plotting the output probability distribution diagram, as shown in Figure 6, it is clear that values +1 and -1 are well separated, and probability distribution degressive and enhancive have slightly Subscriptions. 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

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

Detecting of non-interaction drug pairs

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.

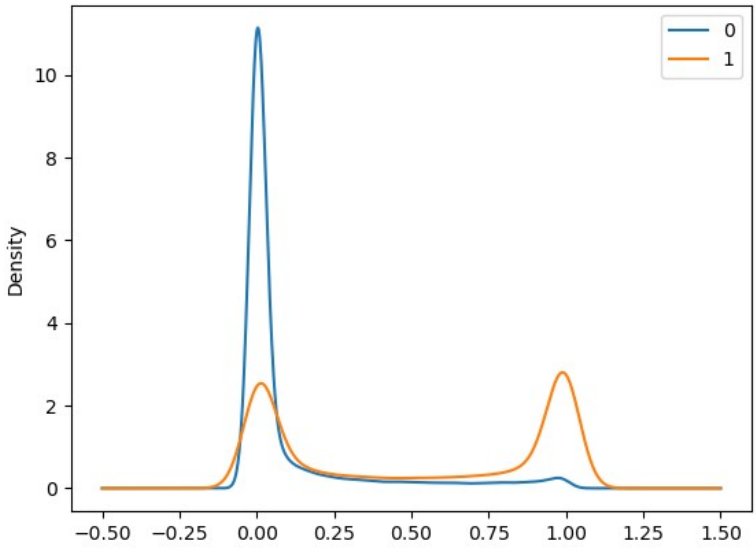


Figure 6: Probability density distribution diagram of Degressive and enhancive. Here, 0 is the same as the −1label, and 1 is the same as +1.

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.

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 Selecting model 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 dual 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.

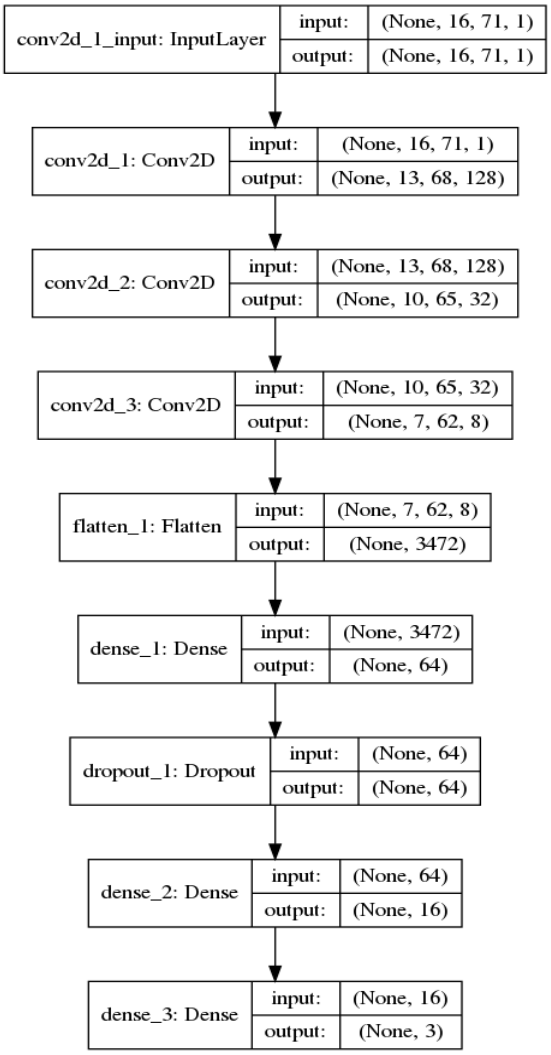


Figure 7: Arrangement of neural network layers SNF-CNN Predict triple-class interaction. Non-interaction (0), degressive interaction (-1) and enhancive interaction (+1)

Selecting final model

The final model is almost the same as that described in Section Selecting model . 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 7. 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.

Three-class model's training trend
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 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.