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

27 of every 100 drug pairs have DDIs Law et al. (2014). The DDIs would put patients who are treated with
28 multiple drugs in an unsafe situation Leape et al. (1995); Businaro (2013); Karbownik et al. (2017); Mulroy
29 et al. (2017). Understanding DDI is the first step in drug combinations, which becomes one of the most
30 promising solutions for the treatment of multifactorial complex diseases Zhao et al. (2011). Therefore,
31 there is an urgent need for screening and analysis of DDIs before clinical co-medications are administered.
32 However, traditional DDI identification approaches (e.g., testing Cytochrome P450 Veith et al. (2009) or
33 transporter-associated interactions Huang et al. (2007)) face challenges, such as high costs, long duration,
34 animal welfare considerations Zhang et al. (2015), the very limited number of participants in the trial, and
35 the great number of drug combinations under screening in clinical trials. As a result, only a few DDIs have
36 been identified during drug development production (usually in the clinical trial phase). Some of them have
37 been reported after drugs approved, and many have been found in post-marketing surveillance.

38 Computational approaches are a promising alternative to discovering potential DDIs on a large scale,
39 and they have gained attention from academy and industry recently Wiśniowska and Polak (2016); Zhou
40 et al. (2016). Data mining-based computational approaches have been developed to detect DDIs from
41 various sources Zhang et al. (2015) , such as scientific literature Bui et al. (2014); Zhang et al. (2016b)
42 , electronic medical records Yamanishi et al. (2008) , and the Adverse Event Reporting System of FDA
43 (http://www.fda.gov). These approaches rely on post-market clinical evidence. So, they cannot provide
44 alerts of potential DDIs before clinical medications are administered. In contrast, machine learning-
45 based computational approaches (e.g. Naïve Similarity-Based Approach Vilar et al. (2014) , Network
46 Recommendation-Based Zhang et al. (2015) , Classification-Based Cheng and Zhao (2014)) can provide
47 such alerts by utilizing pre-marketed or post-marketed drug attributes, such as drug features or similarities
48 Pahikkala et al. (2015). These methods use different drug features to predict DDIs, such as chemical
49 structures Vilar et al. (2014) , targets Luo et al. (2014) , hierarchical classification codes Cheng and Zhao
50 (2014) , side effects, and off-label side effects Zhang et al. (2015); Shi et al. (2017).

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

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

67 Most of these existing machine learning approaches are designed to predict the typical two-class problem,
68 which only indicates how likely a pair of drugs is a DDI. However, two interacting drugs may change
69 their own pharmacological behaviors or effects (e.g., increasing or decreasing serum concentration) in
70 vivo. For example, the serum concentration of Flunisolide (DrugBank Id: DB00180) decreases when it is

71 taken with Mitotane (DrugBank Id: DB00648), whereas its serum concentration increases when taken with
72 Roxithromycin (DrugBank Id: DB00778). For short, the first case is degressive DDI, and the second case
73 is enhancive DDI, which contains drug changes in terms of pharmacological effects. It is more important to
74 know exactly whether the interaction increases or decreases the drug's pharmaceutical behaviors, especially
75 when making optimal patient care, establishing drug dosage, designing prophylactic drug therapy, or
76 finding the resistance to therapy with a drug Koch-Weser (1981).

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

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

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

96 It should be noted that the proposed method of this research is a recommender-based on deep neural
97 networks and has no structural similarities with matrix factorization methods. The only reason for
98 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

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

2.2 Problem formulation

107 Without loss of generality, let $D = \{d_i\}$, $i = 1, 2, .., m$ be a set of m approved drugs. Their interactions
108 can be accordingly represented as an $m \times m$ symmetric interaction matrix $A_{m \times m} = \{a_{ij}\}$. For the
109 conventional DDIs, $a_{ij} = 1$ if d_i interacts with d_j, and $a_{ij} = 0$ otherwise. For the comprehensive DDIs,
110 $a_{ij} \in \{-1, 0, +1\}$. Again, if d_i and d_j do not interact with each other, $a_{ij} = 0$. When there is an enhancive
111 DDI or a degressive DDI between d_i and d_j , $a_{ij} = +1$ or $a_{ij} = -1$ respectively.

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

116   ## 2.3 Data preparing

116   Matrices). F_{str} and F_{se} are, respectively, the feature matrix of the chemical structure and the feature
117   matrix of off-label side effects.

117   Since the new drugs are isolated nodes in the interaction network, we cannot infer their possible interaction
120   from topological information alone. Therefore, additional information (such as chemical structure or off-
123   label side effects) is needed, which is called a drug feature in terms of machine learning. First, we prepare

123   ### 2.3.1 Similarity matrix calculation

124   A common method of calculating similarity called Cosine Similarity is used in machine prediction articles
125   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
126   x_j, Cosine Similarity between x_i and x_j is defined as follows:

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

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

128   It was observed that the values of the feature matrices are discrete, and also the dimensions of the matrices
129   are large. The chemical structure and the off-label side effect have 881 and 9149 dimensions, respectively.
130   On the other hand, machine learning algorithms do not work properly with high-dimensional data and
131   discrete data. As a result, they do not get good results on these kinds of data. Therefore, by exploiting
132   the cosine similarity, that was described above, drug similarity matrices based on chemical structure and
133   off-label side effects are calculated. These matrices are S_{str} and S_{se}, respectively. The dimensions of these

134   ### 2.3.2 Integration drug similarity matrices

134   two matrices are $m \times m$, where $a_{i,j}$ is an element of similarity matrices that shows similarity value between
135   drug i and drug j. Each element $a_{i,j}$ has a value between zero and one.

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

146   In this section, using the similarity network fusion method that has described above, similarity matrices of
147   the chemical structure and the off-label side effect of drugs were integrated. The output of this integration is
148   a new similarity matrix of S_{snf} with dimensions of 568×568, and elements of S_{snf} have a value between
149   zero and one. To integrate the network similarity, the package of SNFPy is used, which is implemented in
150   Python and is available at Ross Markello (2018).
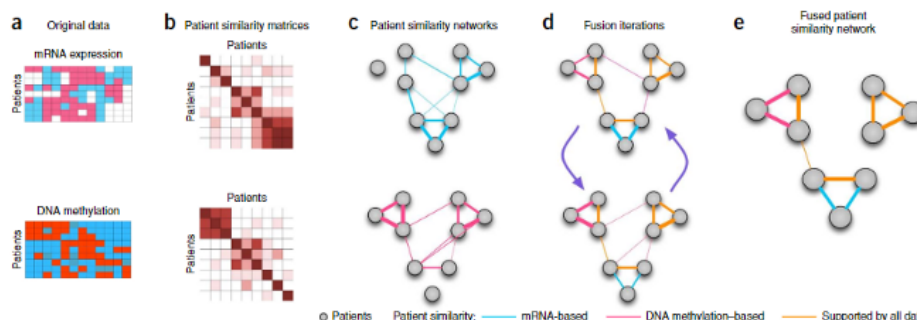
Figure 1. SNF processes Wang et al. (2014): A detailed example of SNF steps. (a) An example representation of chemical structure feature and off-label side effect feature for the same set of drugs. (b) Drug-drug similarity matrices for each feature type. (c) Drug-drug similarity networks, equivalent 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.

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.
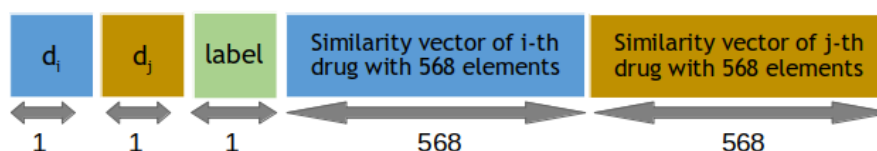
Figure 2. Matrix header of B

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,

170 it does indicate that no interaction has yet been found for this pair of drugs. In the following, we present a
171 method for detecting pairs of non-interacting drugs. Then we use these pairs of drugs as zero-labeled data
172 in the next training.

2.5   Selecting and training model on known interactions

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

2.6   Selecting model

179 We separated rows of matrix B that contain positive and negative interactions. The new matrix contains
180 42,702 pairs of drugs with degressive and enhancive interactions. This data was used to train and find a
181 more suitable model and found a stronger model among many models with different network structures.
182 The final model was a deep neural network that used convolutional and fully connected layers. The features
183 of all interactions (+1 and -1) contain 1136 features. We first divide these features into 10 equal parts. Then,
184 in a 10-cycle for, each period we consider 1 part as testing data and the other 9 parts as a set of training
185 data. We select different models and train the model in the 10-fold CV with 90 of the data. Then we test the
186 model on the remaining 10 percent of the data. In the separating process, pairs of drugs dual are considered.
187 Since the (d_i, d_j) and (d_j, d_i) pairs of drugs are not biologically different from each other, in the separation
188 of training and testing data, necessarily a pair of drugs and their duals are in the same group. This work
189 prevents machine fraud.

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

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

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

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

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

199 Convolution layers using a Flatten layer Connects to fully connected layers. The function of this layer is
200 to transform a two-dimensional matrix into a one-dimensional vector. The output of this input layer of the
201 first layer is fully connected. Also, between fully connected 64 and 16 nodes, we used one Dropout layer
202 Srivastava et al. (2014) with a wast value of 0.2. This value indicates that the network in this layer does not
203 randomly consider 20 percent of the features. This layer is used to prevent over-fitting of the model and
204 forces the model to extract and use more features with more confidence for prediction. If some of them
205 are removed, the algorithm's prediction power either doesn't decrease or doesn't rely on a few specific
206 features.
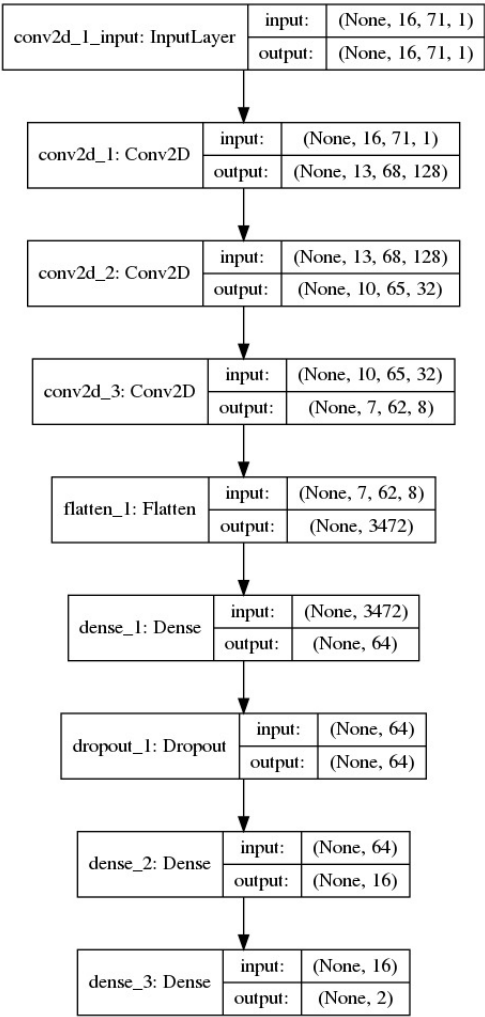
Figure 3. The arrangement of the neural network layers for detecting possible zeros

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

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

213   1) We used Tensorflow Abadi et al. (2016) (version 1.14.0) and KERAS Chollet et al. (2015) (version
214 2.2.5) packages to implement the neural network.

215   2) The categorical-cross entropy loss function was considered an objective function for the neural network,
216 which is generally used to train a classification network Ghosal et al. (1997); Toda and Okura (2012); Seen
217 (2012).

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

220   4) The number of epochs was considered 5.

```
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

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

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

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

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

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

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

236   2) Variance: The confidence interval for the reported values with a reliability coefficient above 95 percent
237   was narrow and close to each other. Out of four reported confidence interval values, three values were
238   less than ± 0.002, and only for the degressive interaction, the AUPR was in the range of ± 0.005. The low
239   amount of variance obtained from the model shows that the proposed model is robust.

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

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

Detection of drug pairs without possible interaction

245   In the previous step, a high-precision, robust, and accurate model has been presented to detect drug pairs'
246   potential interactions for both degressive and enhancive. Therefore, this model has the ability to detect
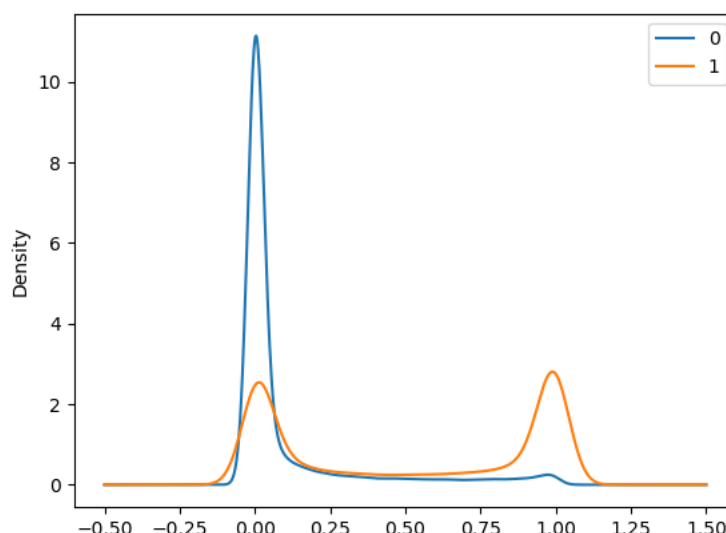
Figure 5. probability density distribution diagram of Degressive and enhancive. In this figure, 0 is the same as the -1label, and 1 is the same as $+1$.

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.

---

247 non-interactions (real zeros) as follows. If drug pairs are unlikely to interact, then those drug pairs are
248 likely to be real zeros.

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

255 **2.7   Selecting and training model on known and unknown interactions**

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

259   First, the B matrix rows, which contain the +1 and -1 interactions, are separated according to the
260 previously detailed procedure and placed in 10 parts. Then, 30,000 non-interacting candidate drug pairs
261 were randomly selected from 65,000 drug pairs. In the chosen drug pairs, the drug pairs and the duals of
262 them must be non-interaction candidates. The zeros group is randomly divided into 10 parts, so each drug
263 pairs and the dual are in the same batch. Then 10 parts of zeros are merged with 10 parts of pre-prepared
264 +1s and -1s.

---

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

2.8   Selecting final model

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

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

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

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

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.

---

3   RESULT AND DISCUSSION

3.1   Assessment

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

290   The whole data set is divided into k equal parts. The k-1 part is used as a training data set, and the model
291   is built based on it, and the testing is performed with the remaining part. The procedure is repeated k
292   times so that each of the k parts is used only once for testing, and each time accuracy is calculated for the
293   constructed model. In this method, the average of predicting accuracy in all rounds of K-fold CV is taken
294   as the final accuracy Classifier. The most common value for k in scientific literary is 5 or 10. Obviously,
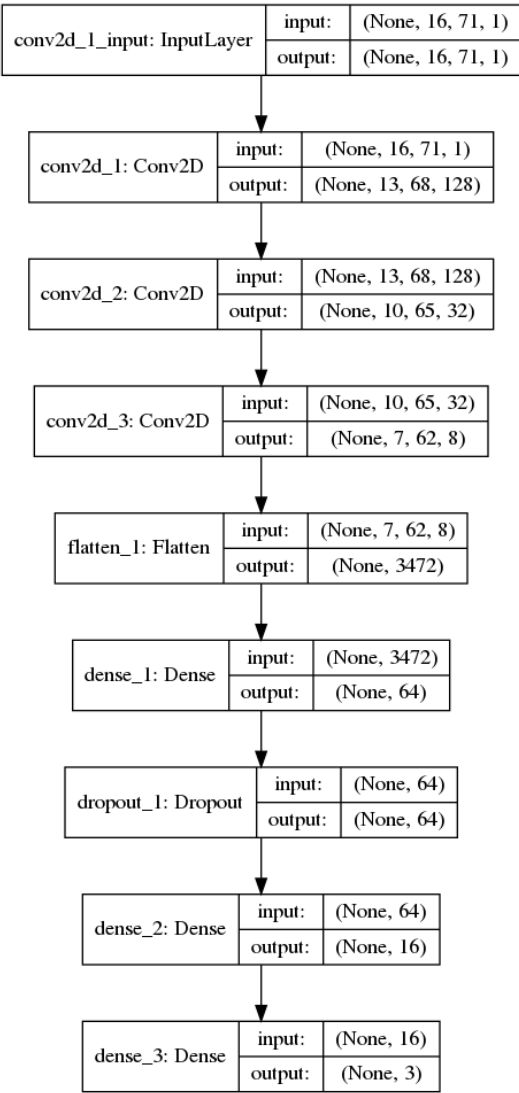

Figure 6. Arrangement of neural network layers SNF-CNN Predict triple-class interaction. Non-interaction (0), degressive interaction (-1) and enhancive interaction (+1)

295   the larger the value of k, the more reliable the calculated accuracy classifier, and the more comprehensive
296   the knowledge obtained, and the longer the classifier's testing time, which is the most important problem.
297   Each setting and each data set has its own validation. In this approach, according to the type of problem
298   and the methods, we used two types of 10-fold CV to divide the data into two sets of testing and training,
299   which are:

3.1.1   First case: 10-fold CV without unknown interactions

301   In this case, we randomly select 90 percent of the enhancive and degressive interactions. For the testing
302   set, we consider the remaining 10 percent of the enhancive and degressive interactions. In the first case of
303   the testing procedure, the model was selected, and some hyper-parameters, such as the number of epochs,
304   were determined.Figure 7 It shows the training process for the selected model. As expected, the model's
305   accuracy is strict on ascending training data, but there are ups and downs for testing data after Epoch 5. In
306   the loss function graph, by the end of epoch 5, as the epochs increase, the loss function's value on training
307   and testing procedure decreases. After epoch 5, the trend of training data continues, but the testing trend

308 is reversed. In other words, over feet occurs. Therefore, based on the graphs, the appropriate number of epochs in this step was considered 5.
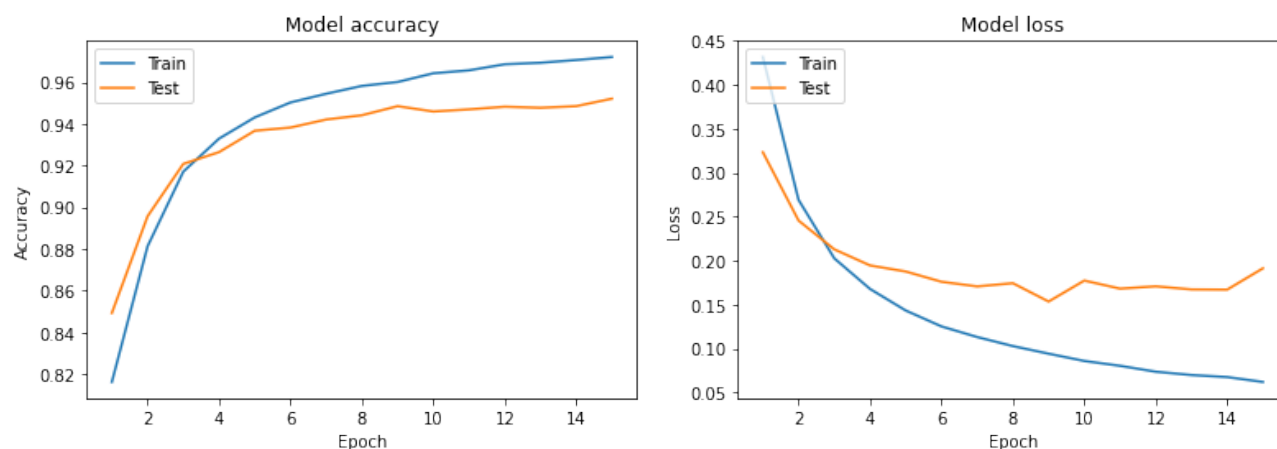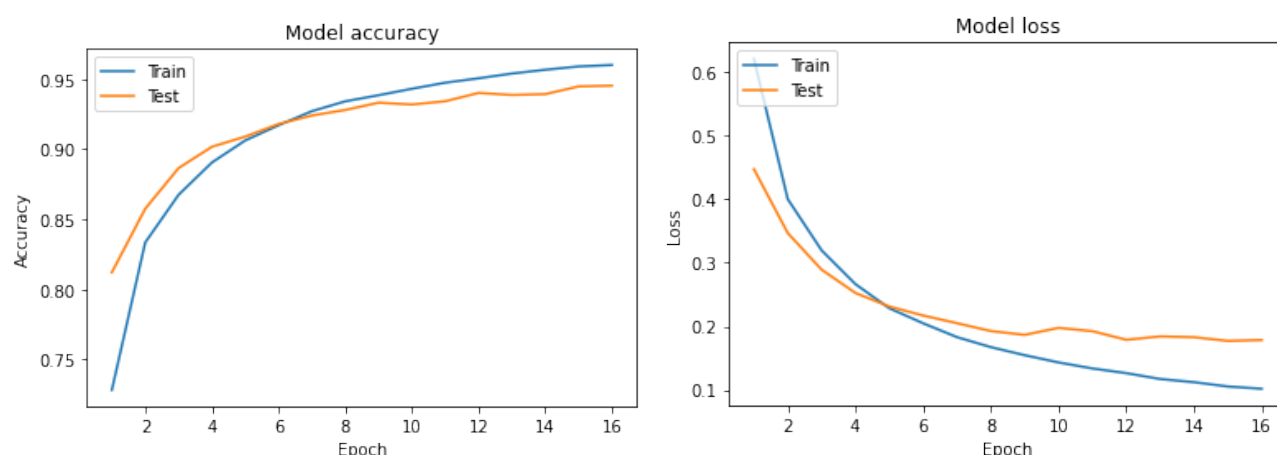

Figure 7. Accuracy and loss function for the binary model: The right figure shows the model's accuracy on training and validation data during 15 epochs, and the left figure shows the loss function values at different epochs.

309

310 ### 3.1.2   Second case: 10-fold CV with unknown interactions

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

316 It shows the training process. The process of the accuracy of the model on training data increases steadily
317 with the increase of epochs. Still, the model after epoch 9 reduces a constant and decreases the accuracy a
318 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

319 In this study, we classify drug pairs according to the type of interaction or non-interaction. So in order to
320 compare the method performance with other existing methods, from four measurement criteria, F- measure,
321 accuracy, Area Under Roc Curve(AUC), and Area Under Precision-Recall Curve(AUPR) Used. To define
322 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.

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

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

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

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

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

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

5   COMPARISON OF RESULTS

336   Based on the validation procedure described in Section **??** , the interaction type detection model was
337   selected and trained. Then, the final three-class model was presented by recognizing the most probable
338   non-interactions. So we tested the SNF-CNN model on data to check the evaluation, robustness, and
339   efficiency of the CV. Results SNF-CNN and other methods for comparison are presented and discussed
340   in this section. Before comparing the different methods, an example of the results of neural network
341   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

342   Table 2 an example of a model implementation result is the ability of the model in terms of precision,
343   recall and F-measure Indicates the type of interactions. According to table 2 the precision of the model in
344   detecting enhancive and degressive interactions is 95 percent and 94 percent, while recall is 99 percent
345   and 83 percent, respectively. the F-measure is also 97 percent and 88 percent that the higher ability of the
346   model to detect degressive interactions comes from a higher number of these types of interactions. The
347   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

348   Also in the table 3 three-Classes interaction classification is displayed. In this implementation, the
349 accuracy of the model for detecting degressive interaction, non-interaction, and enhancive interaction is
350 95 percent, 96 percent, and 88 percent, respectively. The recall is 97 percent, 95 percent, and 84 percent,
351 respectively, and finally F-measure It is 96, 96, and 86 percent. In comparison, the model power in the
352 three-classes decreases slightly to the two-classes, which can be due to two reasons.

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

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

356   For the above reasons, some reduction in the power of the three-classes model in detection was not
357 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

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

	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

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

6 CONCLUSIONS

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

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

380   The first is named as false-positive drug pairs, which are precisely labeled as DDIs in DrugBank version
381   4, but in the current version, they are correctly labeled as non-DDIs. For example, the old version of
382   DrugBank records that Apraclonidine (a Sympathomimetic drug used in glaucoma therapy) increases the
383   atrioventricular blocking activities of Alprenolol and Bevantolol, while the newer version removes it.

384   The second one is false-negative drug Pairs, which are wrongly labeled as non-DDIs in DrugBank version
385   4 but in the current version, they are reported as DDIs.For example, the pair of Valrubicin and Cyclosporine
386   as well as the pair of Ergocalciferol and Calcitriol in the newer version of DrugBank reports, Valrubicin
387   (bladder cancer treatment drug) Increases the activity of the nephrotoxic drug Cyclosporine(A drug that
388   suppresses the immune system with a special action on T-lymphocytes), while the combined therapy of
389   Calcitriol and Ergocalciferol increases the risk or severity of adverse effects in the multiple-drug therapy.

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

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

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

402   ## 6.1   Figures

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

409   ### 6.1.1   Permission to Reuse and Copyright

410   Figures, tables, and images will be published under a Creative Commons CC-BY licence and
411   permission must be obtained for use of copyrighted material from other sources (including re-
412   published/adapted/modified/partial figures and images from the internet). It is the responsibility of the
413   authors to acquire the licenses, to follow any citation instructions requested by third-party rights holders,
414   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

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

REFERENCES

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

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

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

453 Cheng, F. and Zhao, Z. (2014). Machine learning-based prediction of drug–drug interactions by integrating
454   drug phenotypic, therapeutic, chemical, and genomic properties. *Journal of the American Medical*
455   *Informatics Association* 21, e278–e286

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

457 Fu, L. and Peng, Q. (2017). A deep ensemble model to predict mirna-disease association. *Scientific reports*
458   7, 1–13

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

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

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

467 Huang, S.-M., Temple, R., Throckmorton, D., and Lesko, L. (2007). Drug interaction studies: study design,
468   data analysis, and implications for dosing and labeling. *Clinical Pharmacology & Therapeutics* 81,
469   298–304

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

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

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

477 Koch-Weser, J. (1981). Serum drug concentrations in clinical perspective. *Therapeutic drug monitoring* 3,
478   3–16

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

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

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

486 Luo, H., Zhang, P., Huang, H., Huang, J., Kao, E., Shi, L., et al. (2014). Ddi-cpi, a server that predicts
487   drug–drug interactions through implementing the chemical–protein interactome. *Nucleic acids research*
488   42, W46–W52

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

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

492 Olayan, R. S., Ashoor, H., and Bajic, V. B. (2018). Ddr: efficient computational method to predict
493   drug–target interactions using graph mining and machine learning approaches. *Bioinformatics* 34,
494   1164–1173

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

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

499 [Dataset] Ross Markello (2018). snfpy 0.2.2

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

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

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

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

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

513 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a
514   simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15,
515   1929–1958

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

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

520 Veith, H., Southall, N., Huang, R., James, T., Fayne, D., Artemenko, N., et al. (2009). Comprehensive
521   characterization of cytochrome p450 isozyme selectivity across chemical libraries. *Nature biotechnology*
522   27, 1050–1055

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

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

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

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

531 Wiśniowska, B. and Polak, S. (2016). The role of interaction model in simulation of drug interactions and
532   qt prolongation. *Current pharmacology reports* 2, 339–344
533 Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M. (2008). Prediction of drug–target
534   interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24, i232–i240
535 Yu, H., Mao, K.-T., Shi, J.-Y., Huang, H., Chen, Z., Dong, K., et al. (2018). Predicting and understanding
536   comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC systems biology*
537   12, 14
538 Zhang, P., Wang, F., Hu, J., and Sorrentino, R. (2015). Label propagation prediction of drug-drug
539   interactions based on clinical side effects. *Scientific reports* 5, 1–10
540 Zhang, W., Chen, Y., Li, D., and Yue, X. (2018). Manifold regularized matrix factorization for drug-drug
541   interaction prediction. *Journal of biomedical informatics* 88, 90–97
542 Zhang, W., Chen, Y., Tu, S., Liu, F., and Qu, Q. (2016a). Drug side effect prediction through linear
543   neighborhoods and multiple data source integration. In *2016 IEEE international conference on*
544   *bioinformatics and biomedicine (BIBM)* (IEEE), 427–434
545 Zhang, Y., Wu, H.-Y., Xu, J., Wang, J., Soysal, E., Li, L., et al. (2016b). Leveraging syntactic and semantic
546   graph kernels to extract pharmacokinetic drug drug interactions from biomedical literature. *BMC systems*
547   *biology* 10, 67
548 Zhao, X.-M., Iskar, M., Zeller, G., Kuhn, M., Van Noort, V., and Bork, P. (2011). Prediction of drug
549   combinations by integrating molecular and pharmacological data. *PLoS Comput Biol* 7, e1002323
550 Zhou, D., Bui, K., Sostek, M., and Al-Huniti, N. (2016). Simulation and prediction of the drug-
551   drug interaction potential of naloxegol by physiologically based pharmacokinetic modeling. *CPT:*
552   *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.