SNF-CNN: Predicting Comprehensive Drug-Drug Interaction via Similarity Network Fusion and Convolutional Neural Networks

M.Amin Khodamoradi [1,4]*, Bahareh Levian [2,3] and Changiz Eslahchi [2,3], Maria Marques [4], Ricardo Jardim-Gonçalves [1]

1-Universidade NOVA de Lisboa, NOVA School of Science and Technology (FCT NOVA), Caparica, Portugal
2- Department of Computer Sciences, Shahid Beheshti University, Tehran, Iran.
3-School of Bioinformatics, IPM - Institute for Research in Fundamental Sciences, Tehran, Iran.
4 - Center of Technology and Systems (UNINOVA-CTS) and Associated Lab of Intelligent Systems (LASI), 2829-516, Caparica, Portugal

Abstract

Background: Drug-drug interactions (DDIs) often lead to unexpected and adverse reactions, underscoring the critical need for their identification before market entry. However, preclinical detection of DDIs remains resource-intensive. Computational methods have demonstrated the ability to predict potential DDIs on a large scale using premarket drug properties. While many models focus on determining whether drugs interact, they frequently overlook the enhancive (positive) and depressive (negative) alterations of pharmacological effects. Furthermore, these complex DDIs are not random and may stem from the structural features of the DDI graph. Uncovering such relationships is essential for understanding the mechanisms behind DDI occurrences. Predicting comprehensive DDIs and elucidating structural patterns in the DDI graph provide crucial guidance for prescribing multiple drugs.

Results: In this study, we treat a set of comprehensive DDIs as a signed network and introduce the Similarity Network Fusion and Convolutional Neural Networks (SNF-CNN) model for predicting the enhancive or depressive effects of drug pairs. SNF-CNN demonstrates strong performance in depressive DDI prediction (AUC = 0.975, AUPR = 0.967), enhancive DDI prediction (AUC = 0.969, AUPR = 0.822), and Unknown DDI prediction (AUC = 0.971, AUPR = 0.948). Comparative analysis against three state-of-the-art methods on our dataset highlights the superiority of SNF-CNN.

Conclusions: This innovative approach excels not only in predicting comprehensive DDIs but also in accurately forecasting non-DDIs.

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 might be unexpectedly influenced by each other [1]. This kind of influence is termed Drug-Drug interaction (DDI), which may 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 DrugBank[1], on average, 15 out of every 100 drug pairs have known DDIs [2]. The DDIs may put patients, who are treated with multiple drugs,

[1] https://go.drugbank.com/

in an unsafe situation [3] [4] [5]. Understanding DDI is the first step in drug combinations, which becomes one of the most promising solutions for the treatment of multifactorial complex diseases [6]. 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 [7] or transporter-associated interactions [8]) face challenges, such as high costs, long duration, animal welfare considerations [9], a very limited number of participants in trials, and a 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 were approved, and many have been found in post-marketing surveillance [10].

DDIs can be significantly affected by a patient's medical history [a] and genetics [b]. To facilitate the link between these aspects, the Smart4Health project[2] developed two platforms: one personal, containing health information from the citizen (Citizen Health Data Platform – CHDP) such as medical conditions, allergies and intolerances, medication use, as well as genetic data, and one deidentified, containing data donated for research by the citizen (Research Platform – RP). While CHDP adopts HL7 FHIR[3] to structure collected data, RP follows OMOP CDM[4] to convey data coming from CHDP and make it reusable by third-party research infrastructures (e.g., ELIXIR[5]). The concept of use entails the possibility of a citizen to collect and aggregate data generated from interactions with medical institutions (e.g., medication prescriptions, laboratory results, discharge letters, etc.) into one single, interoperable HER. This data may also include genetic data if available. At citizen discretion, this data can be donated to the RP. In the specific case of data related to medication intake and genetic data, these are linked to drug exposure and outcome data within the OMOP CDM. This mechanism has the potential to facilitate data collection and contribute to ensuring the quality of the collected data. In addition, having the citizen at the center of this process may accelerate and expand the identification of DDIs enabling a more comprehensive understanding of their mechanisms.

Computational approaches are a promising alternative to discovering potential DDIs on a large scale, and they have gained attention from academics and industry recently [11][12]. Data mining-based computational approaches have been developed to detect DDIs from various sources [9], such as scientific literature [13] [14], electronic medical records [15], and the Adverse Event Reporting System of the Food and Drug Administration (FDA[6]). These approaches rely on post-market clinical evidence, thus, 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 [9], 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 [9] [20].

A Dependency-based Convolutional Neural Network (DCNN) was proposed for drug-drug interaction extraction by Liu et al. in 2016 [21]. DCNN is a text-mining approach that predicts DDIs based on unstructured biomedical literature and the existing knowledge bases. It applies convolution layers on

[2] www.smart4health.eu
[3] https://hl7.org/fhir/
[4] https://www.ohdsi.org/data-standardization/
[5] https://elixir-europe.org/
[6] http://www.fda.gov

word sequences as well as dependency parsing trees of candidate DDIs for adjacent words. DeepDDI has been 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 the prediction of potential causal mechanisms 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 analyses [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 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].

Although the occurrence of both enhancive and degressive DDIs is not random[32] [33], 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 since 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 for high-order drug interaction.

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

Model of TMFUF [32] is proposed by Shi 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, assigns every drug to a drug community. As a result, some correlations are observed between drug communities and the numbers of enhancive, degressive, sum, and difference of DDIs for each drug.

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

All three introduced algorithms use 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 for safer prescriptions.

In this study, we commence by elucidating the data preparation process and introducing a recommendation system designed to discern pairs of non-interacting drugs with high precision. Following this, we present a groundbreaking algorithm that integrates drug similarities and leverages deep learning recommendation systems to predict Drug-Drug Interactions (DDIs) within a comprehensive three-class model. Termed 'Predicting Comprehensive Drug-Drug Interaction via Similarity Network Fusion and Convolutional Neural Networks' (SNF-CNN), this algorithm strives to identify unknown and potential DDIs that have not yet been detected. The intrinsic features of off-label side effects and the chemical structure of drugs embedded in our approach provide valuable insights for uncovering hidden potential DDIs within the existing DDI network. To harness both similarity features effectively, we capitalize on the power of Similarity Network Fusion (SNF).

This approach diverges from conventional methods by relying on deep neural networks, specifically a convolutional neural network, and distinctly differs from matrix factorization techniques. While we briefly acknowledge these alternative methods, it is crucial to emphasize that our work represents a unique exploration within the realm of three-class data, setting it apart from existing studies.

Methods

Dataset and features

This study utilizes the dataset introduced by Yu et al. in 2018 [33], comprising 568 approved small-molecule drugs. Each drug within the dataset exhibits at least one interaction with other drugs, resulting in a total of 21,351 Drug-Drug Interactions (DDIs). Notably, these interactions are further categorized into 16,757 enhancive DDIs and 4,594 degressive DDIs.

Each drug in the dataset is uniquely characterized by two feature vectors:

1. An 881-dimensional feature vector (Fstr), derived from PubChem chemical structure descriptors.
2. A 9149-dimensional feature vector (Fse), based on off-label side effects sourced from the OFFSIDES database [37].

The elements in these vectors are binary, assigned 1 if the corresponding side effect or chemical structure is reported or observed and 0 otherwise.

This dual-feature representation encapsulates both the structural attributes and off-label side effects of each drug, forming the foundational elements for subsequent analyses in our investigation.

Problem formulation

Let $D = \{d_i\}, i = 1, 2, ..., m$, represent a set of m-approved drugs. Each drug d in D is denoted by a p-dimensional feature vector $f_i = [f_1, f_2, ..., f_k, ..., f_p]$, where $f_k = 1$ indicates the presence of the k-th specific chemical structure fragment or occurrence of an off-label side effect, and $f_k = 0$ otherwise. Given that each drug has two feature vectors representing chemical structure and off-label side effects, two feature matrices F are constructed with dimensions $m \times p$ (where the magnitude of p depends on the type of feature). The matrices F_{str} and F_{se} correspond to the feature matrices of chemical structure and off-label side effects, respectively DDI can be represented by an $m \times m$ symmetric interaction matrix $A_{m \times m} =$

$\{a_{ij}\}$. For conventional binary DDIs, $a_{ij} = 1$ if d_i interacts with d_j, and $a_{ij} = 0$ otherwise. In the case of comprehensive DDIs, the matrix includes three values ($a_{ij} \in \{-1, 0, +1\}$). Similar to the binary formulation, if d_i and d_j do not interact, $a_{ij} = 0$. However, if 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

The feature matrices' values are discrete, and their dimensions 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 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 [38] [39]. If we name feature vectors of the drug of d_i and d_j as x_i and x_j, Cosine Similarity between x_i and x_j is defined as follows:

$$S_{cos}(x_i, x_j) = \frac{x_i \cdot x_j}{(||x_i||\,||x_j||)} \qquad (1)$$

Where $||\ .\ ||$ is the Euclidean Norm and $x_i \cdot x_j$ is the 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 effects, 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. Figure 1 is a detailed example of SNF [23] 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 are 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. The edge color indicates which data type has contributed to the given similarity.

In this section, similarity matrices of the chemical structure and the off-label side effects of drugs were integrated via the 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 [40].

Input matrix format

At this stage, a matrix forms with 1139 columns and 322056 rows. In Figure 2, shows the input data header including 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 the *j-th* drug from the S_{snf} matrix with 568 elements.

The data contains 568 drugs. 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 ($(568 \times 568) - 568 = 322056$). According to the explanation, a matrix with dimensions of 322056×1139 is formed to input into our model, which is called the *B* matrix.

SNF processes [23]: 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 are 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. The edge color indicates which data type has contributed to the given similarity.

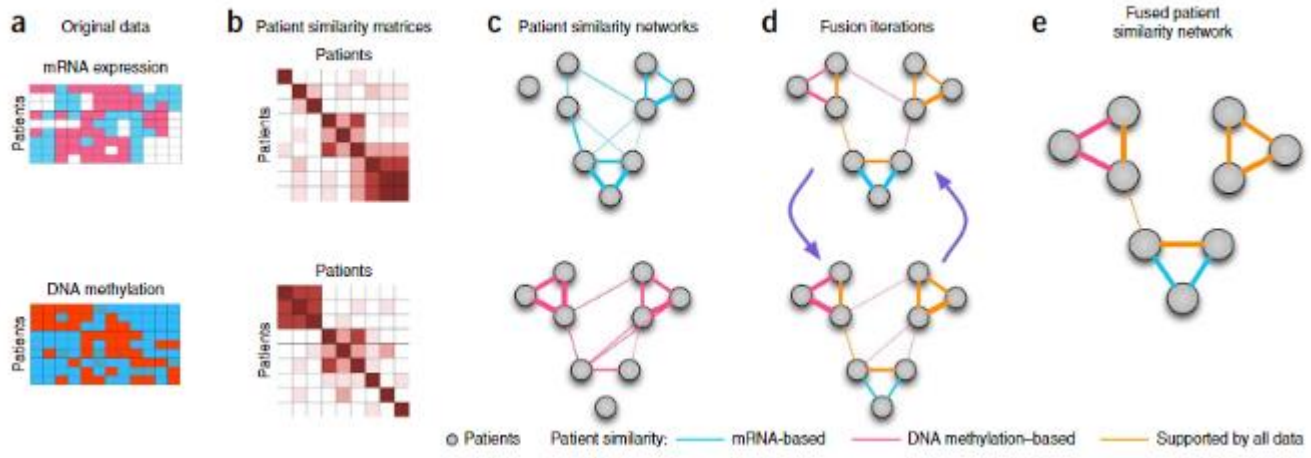

Figure 1 SNF processes [23]: A detailed example of SNF steps. (a) Chemical structure and off-label side effect features of drugs. (b) Drug-drug similarity matrices for each feature type. (c) Drug-drug similarity networks are equivalent to the matrix in the previous step. Nodes are drugs, and edges

represent drug pairwise similarities. (d) Network fusion by SNF iterative updates. (e) The final result of SNF.

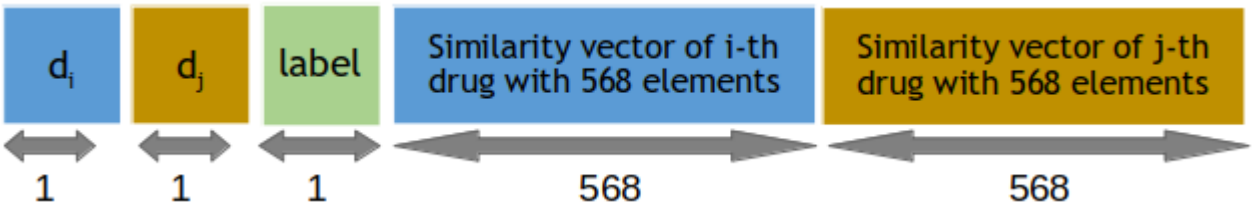

Figure 2 Matrix scheme of tabular input data (table of B)

Devising of Recommender System

In the earlier stages, the data was meticulously prepared to accommodate a variety of learning machines, encompassing those utilizing deep learning techniques. While positive and negative DDIs come with distinct labels, the zero label does not imply the lack of interaction between a drug pair. Rather, it signifies that no interaction has been identified for that particular drug pair. In the following sections, we delineate a method for pinpointing pairs of non-interacting drugs. These drug pairs are subsequently employed as zero-labeled data in the subsequent training phase.

Evaluation process

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 been 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, the model has been built based on them, and the test has been performed with the remaining part. This procedure has repeated K times so that each of the K parts is used only once for testing and each time a resolution metric is 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 literature is 5 or 10. The more detailed validation in the K-fold CV the more reliable the classifier accuracy, the more comprehensive the obtained knowledge, and also the more time-consuming the validation process.

Table 1 Confusion matrix for interaction type and relevant evaluation metrics. True Positive (TP): Drug pairs, correctly, classified as enhancive interaction, False Positive (FP): Drug pairs, incorrectly, classified as enhancive interaction incorrectly, False Negative (FN): Drug pairs, incorrectly, classified

as degressive interaction, True Negative (TN): Drug pairs, correctly, classified as degressive interaction.

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

If consider degressive interaction as a Negative (N) and enhancive interaction as a 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. Expressed as:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

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

$$Precision_{0(enh/deg)} = \frac{TP}{TP+FP} \quad (3)$$

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

$$Recall_{(enh/deg)} = \frac{TP}{TP+FN} \quad (4)$$

Precision and recall have a trade-off; thus, improving one of them may lead to a reduction in another. Therefore, utilizing the F-measure is more reasonable.
F-measure: The geometric mean of precision and recall. Expressed as:

$$F-measure_{(enh/deg)} = \frac{2 \times Precision_{(enh/deg)} \times Recall_{(enh/deg)}}{Precision_{(enh/deg)} + Recall_{(enh/deg)}} \quad (5)$$

Since the values of precision, recall, and F-measure are 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, which is the area under the precision-recall curve. These criteria indicate the efficiency of methods independent of the threshold value. In cases where the fraction of negative samples and positive samples are not equal, and AUPR is the fairer criterion for evaluation.

Selecting and training models 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. High resolution in detecting these zeros can help provide a more accurate and confident three-class model.

Selecting model

We partitioned the rows of matrix *B* to isolate instances of positive and negative interactions, resulting in a new matrix comprising 42,702 drug pairs manifesting degressive and enhancive interactions. This curated dataset served as the foundation for model training and selection, where we identified a robust deep neural network model incorporating convolutional and fully connected layers from a pool of diverse network structures.

The interaction data, denoted by both positive and negative labels (+1 and -1), is characterized by feature vectors comprising 1136 elements. Diverse models underwent consideration, and the chosen model underwent rigorous training in a 10-fold Cross-Validation (CV) procedure, employing 90 percent of the data for training. Subsequently, the model's performance was evaluated on the remaining 10 percent of the data. Throughout the data split process, the dual drug pairs duals were considered. Acknowledging that drug pairs (d_i, d_j) and (d_j, d_i) lack biological distinctions, these pairs were consistently grouped in both training and testing sets, ensuring methodological integrity and preventing potential biases that could yield 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 [41], which is defined as the positive part of its argument:

$$ReLU(x) = max\{x, 0\} \qquad (6)$$

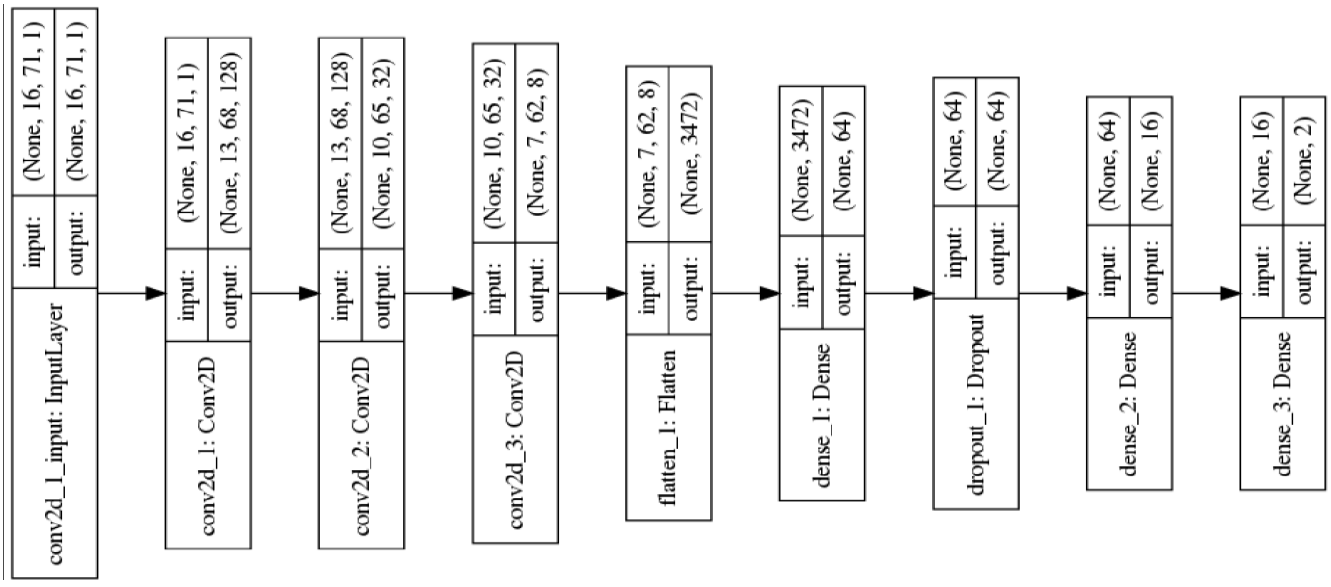

Figure 3 *The arrangement of the neural network layers for detecting possible zeros*

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 [42], which is calculated as follows:

$$Sigmoid(x) = \frac{1}{(1 + e^{-x})} \quad (7)$$

Convolution layers using a Flatten layer connect 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 [43] with a waste 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 17×16 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
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

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

1. We used Tensorflow [44] (version 1.14.0) and KERAS [45] (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 [46] [47] [48].
3. ADAM optimization [49] 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. A 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. In 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 in the training and testing procedure decreases. After epoch 5, the trend of training data continues, but the testing trend is reversed. In other words, over feet occurs. Therefore, based on the graphs, the appropriate number of epochs in this step was considered 5.

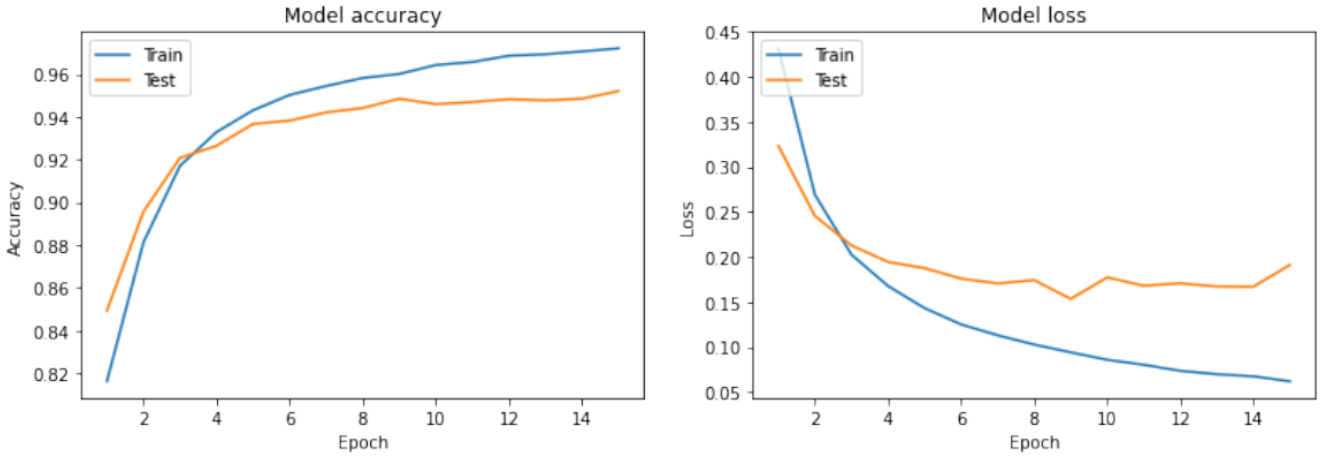

Figure 5 ***Accuracy and loss function for the binary model:*** *The right figure shows the model's accuracy on training and validation data, during 15 epochs, and the left figure shows the loss function values at different epochs.*

Hyperparameters Optimization

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

1) Model overfitting: If hyper-parameters 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 work 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 hyper-parameters 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.

Reliability evidence 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 enhancive interactions. These results indicate the high resolution and detection power of the selected model. The selected model resolution results are presented in Table 2 that 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 the implemented model which shows the ability of the model in terms of precision, recall, and F-measure Indicating 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 and 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 slight Subscriptions. The Pseudocode 1 shows the step-by-step model selection process.

Pseudocode 1 Model selection process

	Algorithm 1 Model selection pseudocode
In/out	Input: +1 and -1 drug pairs features Output: +1 and -1 diagnostic model
1	Apply a 10-fold CV to the features of +1 and -1 drug pairs.
2	Select the right model.
3	Test the model results in a 10-fold CV.

4	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.

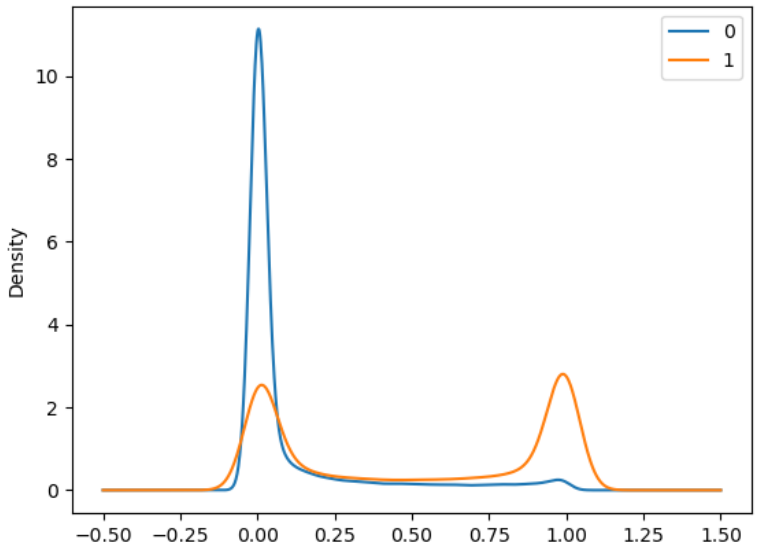

Figure 6 Probability density distribution diagram of degressive and enhancive. Here, 0 is the same as the −1 label, and 1 is the same as +1.

Therefore, this model can 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. 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 the 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 pair 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, and is ready to use in the training and testing of the final recommender system.

Three-class model training trend

The same model structure of none-interaction detection is used for a three-class scenario. All interaction data (enhancive, degressive, and zeros of the first step) are divided into ten equal parts for a 10-fold CV validation procedure, while each fold contains ten percent of each class. Figure 7 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.

The proposed SNF-CNN method's general process is presented in the form of Pseudocode 2, which includes the steps of preparation, model selection, real zero detection, and the presentation of a comprehensive recommender system.

Pseudocode 2 Model selection steps for SNF-CNN

	Algorithm 2 Final model selection (SNF-CNN) pseudocode
In/out	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 the step 5 model, and train it.
7	Predict unknown drug pairs.

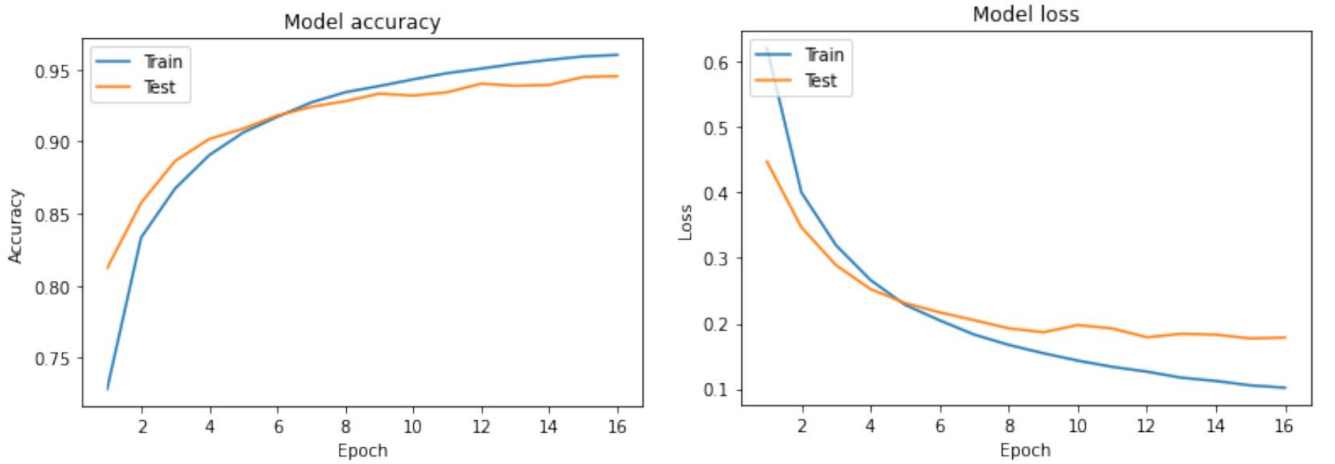

Figure 7 Accuracy and loss function diagrams for the triple model: The right figure shows the accuracy of the model on training and validation data during 16 epochs, and the left figure shows the loss function values at different epochs.

Results and discussion

Each setting of model and features and each data set needs its validation process. According to the type of problem and the methods, we use two types of 10-fold CV to select the most proper model also to validate the results of the chosen model. The metrics which are used in validation are described in the following.

Evaluation criteria for the final three-class model

In this study, we classify drug pairs into three classes, according to the type of interaction or non-interaction, so to compare the method performance with other existing methods, four measurement criteria, F- measure, accuracy, Area Under Roc Curve (AUC), and Area Under Precision-Recall curve (AUPR) are used. To define these criteria, we should use a confusion matrix as Table 3 demonstrates.

Table 3 shows the confusion matrix for the triple class case. In this mode, we have to redefine the definitions of TP, TF, TN, and TP, as well as accuracy, precision, recall, and F- measure.

	Actual Enhancive	***Actual Degressive***	***Actual Non-interaction***
Predicted Enhancive	$cell_{1:}$ T_{Enh}	$cell_{2:}$ F_{Enh}	$Cell_{3:}$ F_{Enh}
Predicted Degressive	$cell_{4:}$ F_{Deg}	$cell_{5:}$ T_{Deg}	$cell_{6:}$ F_{Deg}
Predicted Non-interaction	$cell_{7:}$ $F_{Non-Int}$	$cell_{8}$ $:F_{Non-Int}$	$cell_{9:}$ $T_{Non-Int}$

Table 3 The confusion matrix and relevant evaluation index for predicting triple classes whose cell names start with T (True). The main diagonal amounts of the matrix show correct predictions for each class. The other cells show drug pairs that have been classified by mistake and whose cell names start with F (False).

Accuracy$_{(enh/deg/nonInt)}$: The fraction of all correct predictions (TPs) to all predictions.

$$Acurracy = \frac{All\ corroctly\ predicted\ \ (cell1 + cell5 + cell9)}{Sumation\ all\ nine\ cells\ (\sum_{i=1}^{9} cell_i)} \qquad (8)$$

Based on Table 3 TP, TF, TN, and TP will be defined as Table 4 for triple-classes scenario.

The other three evaluation criteria for each class are defined as follows:

Precision: The ratio of correct predicted ($TP_{Enh/Deg/Non-Int}$) interactions among all predicted (Enh/Deg/Non-Int) interactions.

$$Precision_{\ (Enh/Deg/Non-Int)} = \frac{TP_{(Enh/Deg/Non-Int)}}{TP_{(Enh/Deg/Non-Int)} + FP_{(Enh/Deg/Non-Int)}} \qquad (9)$$

Enhancive	Degressive	Non-interaction
TP = $cell_1$	TP = $cell_5$	TP = $cell_9$
FP = $cell_2$ + $cell_3$	FP = $cell_4$ + $cell_6$	FP = $cell_7$ + $cell_8$
FN = $cell_4$ + $cell_7$	FN = $cell_2$ + $cell_8$	FN = $cell_3$ + $cell_6$
TN = $cell_5$ + $cell_6$ + $cell_8$ + $cell_9$	TN = $cell_1$ + $cell_3$ + $cell_7$ + $cell_9$	TN = $cell_1$ + $cell_2$ + $cell_4$ + $cell_5$

Table 4 The new definition of TP, TN, FP, and FN is in the triple class mod.

Recall: The ratio of correct predicted ($TP_{Enh/Deg/Non-Int}$) interactions among all true (Enh/Deg/Non-Int) interactions.

$$Recall_{(Enh/Deg/Non-Int)} = \frac{TP_{(Enh/Deg/Non-Int)}}{TP_{(Enh/Deg/Non-Int)} + FN_{(Enh/Deg/Non-Int)}} \quad (10)$$

Precision and recall for each class have a trade-off; Therefore, the F-measure can show the resolution ability of the model in each class.

F-measure: The geometric mean of precision and recall in triple-classes:

$$F - measure_{(Enh/Deg/Non-Int)} = \frac{2 \times Precision_{(Enh/Deg/Non-Int)} \times Recall_{(Enh/Deg/Non-Int)}}{Precision_{(Enh/Deg/Non-Int)} + Recall_{(Enh/Deg/Non-Int)}} \quad (11)$$

Also, we evaluated methods via modified AUC, and AUPR for the three-class model.

Comparison of results

Based on the validation procedure, which is described in the Section, the binary interaction type detection model is devised and trained. Then, the final three-class model is presented which took the most probable non-interactions as zeros. The SNF-CNN model is evaluated in 10- a fold CV to check the robustness, and efficiency of the SNF-CNN. Results of SNF-CNN and other methods for comparison are presented and discussed in this section.

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

Table 5 Three-Classes Interaction Classification Report

Table 5 three-class interaction classification is displayed. In this implementation, the precision of the model in detecting degressive interactions, non-interactions, and enhancive interactions are 95%, 96%, and 88%, respectively. The recalls are 97%, 95%, and 84%, respectively, and finally, F-measures are 96%, 96%, and 86%. The model power in the three-classes mode decreases slightly in comparison to the two-classes which can be due to two reasons.

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

2) The suggested non-interactions or zeros are not necessarily real and are not pharmacologically proven, so there is a possibility of some disturbance.

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

	AUC	AUPR
Degressive	0.9747 ± 0.0033	0.9666 ± 0.0045
Enhancive	0.9686 ± 0.0028	0.8221 ± 0.0184
Non-interaction	0.9714 ± 0.0040	0.9480 ± 0.0083

Table 6 Results of the SNF-CNN algorithm in predicting triple-classes based on AUC and AUPR criteria and their confidence interval

Since the previous triple-classes, DDI models reported AUC and AUPR for comparison, SNF-CNN results which are shown in *Table 6* are reported based on these two criteria. Also, the margin of error with the 95% confidence interval is reported in *Table 6*. The algorithm results have a small margin in the 10-fold CV which shows the robustness and reliability of the proposed algorithm.

	AUC	AUPR
SNF-CNN	0.971	0.912
BRSNMF *[35]*	0.805	0.644
Semi-NMF *[33]*	0.796	0.579
TMFUF *[32]*	0.842	0.526

Table 7 Comparison of the results of three-classes prediction algorithms based on criteria AUC and AUPR

In *Table 7*, results of SNF-CNN averaged for the three classes and compared with other existing three-classes algorithms. According to *Table 7*, the proposed algorithm has a high difference compared to other superior algorithms with the problem of ternary and has been able to challenge other algorithms.

Conclusions

Existing machine learning approaches can detect potential interactions of a drug by using large-scale data before taking the drug. However, they cannot predict comprehensive three-class DDIs, including degressive and enhancive interactions. It is more informative to clarify if a drug pair has an enhancive DDI a degressive DDI or even a non-DDI than recognizing whether a drug pair whether or not has a DDI. Without considering the pharmacological changes caused by DDIs, most existing approaches only report two-classes predictions. In addition, not only the occurrence of degressive and enhancive DDIs is not random, but also it represents a piece of information about the major behaviors of those two drugs. None of the existing approaches investigates and considers this intrinsic important property of interactions when treating complex diseases *[34]*.

In this work, after representing comprehensive DDI data and drug features, we used the template of recommender systems to design a novel algorithm. Although the prediction obtained by our algorithm was inspiring, overall performance could still be improved. Incorrectly predicted DDIs were investigated to prove the algorithm in practice we checked case-by-case of the model prediction, in the latest version of the DrugBank database. Observations and investigations led to the discovery of three reasons for wrong predictions, all of them related to differences between versions 4 and 5 of the DrugBank:

1) Removed interactions. Data was precisely labeled in DrugBanck version 4 is no longer labeled as such in version 5. For example, the DrugBank version 4 records that Apraclonidine (DB00964) (also known as iopidine, is a sympathomimetic used in glaucoma therapy.) increases the atrioventricular blocking activities of Alprenolol (DB00866) and Bevantolol (DB01295), while version 5 removes them. This represents a problem for new research as data shows no interaction.

2) **Drug Pairs labeled as non-DDIs. Some drug pairs labeled as non-DDIs in** DrugBank version 4 drug are reported as DDIs in version 5. For example, the pair of Valrubicin (DB00385) and Cyclosporine (DB00091), as well as the pair of Ergocalciferol (DB00153) and Calcitriol (DB00136) in the newer version of DrugBank, reports, Valrubicin (bladder cancer treatment drug) Increases the activity of the nephrotoxic drug Cyclosporine (A drug that suppresses the immune system with a special action on T-lymphocytes), while the combined therapy of Calcitriol and Ergocalciferol increases the risk or severity of adverse effects in the multiple-drug therapy.

3) **Altering DDIs' types.** Labeled as enhancive DDIs in DrugBank version 4, are labeled as degressive DDIs in version 5, and vice versa.

We expect that the SNF-CNN approach will be able to achieve better results in DDI prediction with a better dataset that has less missed or false information about drug pairs. For future research, it is recommended that drugs and their features always be collected from the latest version of DrugBank.

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

As future work, authors are also evaluating the possibility of extending the research by combining the results here presented with the results achieved within the Smart4Health project regarding pharmacogenomics for personalized health [c] to study DDI mechanisms in specific patient profiles and contribute to the development of personalized treatment schemes.

Appendix

Funding

This work was partially funded by the European Union's Horizon 2020 research and innovation program in the scope of the Smart4Health under grant agreement No 826117 and by the Portuguese FCT program, Center of Technology and Systems (CTS) UIDB/00066/2020 / UIDP/00066/2020.

Availability of data and materials

The code and data are available on the GitHub page of SNF-CNN code and data (https://github.com/aminkhod/DDI-Project)

Author Details

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.

References

[1] W. LC and H. TG, "Predicting in vivo drug interactions from in vitro drug discovery data. Nat Rev Drug Discov," *Nature reviews. Drug discovery,* vol. 4, no. 10, pp. 825-833, 2005.

[2] L. V, K. C, D. Y, J. T, G. AC, L. Y, M. A, A. D, W. M, N. V, T. A, G. G, L. C, A. S, D. ZT, H. B, Z. Y and W. DS, "DrugBank 4.0: shedding new light on drug metabolism," *Nucleic acids research,* vol. 42, no. Database issue (2014), pp. 1091-1097, 2014.

[3] L. LL, B. DW, C. DJ, C. J, D. HJ, G. T, H. R, I. J, L. N and e. a. Laffel G, "Systems analysis of adverse drug events. ADE Prevention Study Group," *JAMA,* vol. 274, no. 1, pp. 35-43, 1995.

[4] K. A, S. E, S. K, G. T, W. A and G. E, "Pharmacokinetic drug-drug interaction between erlotinib and paracetamol: A potential risk for clinical practice," *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences,* vol. 102, pp. 55-62, 2017.

[5] E. Mulroy, J. Highton and S. Jordan, "Giant cell arteritis treatment failure resulting from probable steroid/antiepileptic drug-drug interaction," The New Zealand Medical Journal (Online), November 2017. [Online]. Available: https://journal.nzma.org.nz/journal-articles/giant-cell-arteritis-treatment-failure-resulting-from-probable-steroid-antiepileptic-drug-drug-interaction. [Accessed 13 December 2023].

[6] Z. Xing-Ming, I. Murat, Z. Georg, K. Michael, v. N. Vera and B. Peer, "Prediction of drug combinations by integrating molecular and pharmacological data," *PLOS Computational Biology,* vol. 7, no. 12, pp. 1-7, 12 2011.

[7] V. H, S. N, H. R, J. T, F. D, A. N, S. M, I. J, A. CP, L. DG and A. DS, "Comprehensive characterization of cytochrome p450 isozyme selectivity across chemical libraries," *Nat Biotechnol,* vol. 27, no. 11, pp. 1050-5, 2009.

[8] H. SM, T. R, T. DC and L. LJ, "Drug interaction studies: study design, data analysis, and implications for dosing and labeling," *Clinical pharmacology and therapeutics,* vol. 81, no. 2, pp. 298-304, 2007.

[9] P. Zhang, F. Wang, J. Hu and R. Sorrentino, "Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects," *Scientific Reports,* vol. 5, no. 1, 2015.

[10] Karim, M. R. a. Cochez, M. a. Jares, J. B. a. Uddin, M. a. Beyan, O. a. Decker and Stefan, "Drug-Drug Interaction Prediction Based on Knowledge Graph Embeddings and Convolutional-LSTM Network," in *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, New York, NY, USA, 2019.

[11] B. Wiśniowska and S. Polak, "The role of interaction model in simulation of drug interactions and qt prolongation," *Current pharmacology reports,* vol. 2, no. 6, pp. 339-344, 2016.

[12] D. Zhou, K. Bui, M. Sostek and N. Al-Huniti, "Simulation and prediction of the drug-drug interaction potential of naloxegol by physiologically based pharmacokinetic modeling," *CPT: pharmacometrics & systems pharmacology,* vol. 5, no. 5, pp. 250-257, 2016.

[13] B. QC, S. PM, v. M. EM and K. JA, "A novel feature-based approach to extract drug–drug," *Bioinformatics (Oxford, England),* vol. 30, no. 23, pp. 3365-3371, 2014.

[14] Z. Y, W. HY, X. J, W. J, S. E, L. L and X. H, "Leveraging syntactic and semantic graph kernels to extract pharmacokinetic drug-drug interactions from biomedical literature," *BMC systems biology,* vol. 10, no. 8, p. Suppl 3(Suppl 3):67, 2016.

[15] Y. Y, A. M, G. A, H. W and K. M, "Prediction of drug–target interaction networks from the integration of chemical and genomic spaces," *Bioinformatics,* vol. 24, no. 13, pp. i232-i240, 2008.

[16] V. S, U. E, S. L, L. T, H. G, F. C and T. NP, "Similarity-based modeling in large-scale prediction of drug-drug interactions," *Nature protocols,* vol. 9, no. 9, pp. 2147-63, 2014.

[17] F. Cheng and Z. Zhongming, "Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties," *Journal of the American Medical Informatics Association,* vol. 21, no. e2, p. e278–e286, 2014.

[18] P. T, A. A, P. S, S. S, S. A, T. J and A. T, "Toward more realistic drug–target interaction predictions," *Brief Bioinform,* vol. 16, no. 2, pp. 325-37, 2015.

[19] L. H, Z. P, H. H, H. J, K. E, S. L, H. L and Y. L, "DDI-CPI, a server that predicts drug-drug interactions through implementing the chemical-protein interactome," *Nucleic acids research,* vol. 42, no. 2014, pp. W46-52, 2014.

[20] H. Huang, J.-X. Li, P. Lei, Y.-N. Zhang and S.-M. Yiu, "Predicting comprehensive drug-drug interactions for new drugs via triple matrix factorization," in *Bioinformatics and Biomedical Engineering*, Cham, 2017.

[21] S. Liu, K. Chen, Q. Chen and B. Tang, "Dependency-based convolutional neural network for drug-drug interaction extraction," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016.

[22] "Deep learning improves prediction of drug–drug and drug–food interactions," in *Proceedings of the National Academy of Sciences of the United States of America*, 2018.

[23] W. B, M. AM, D. F, F. M, T. Z, B. M, H.-K. B and G. A, "Similarity network fusion for aggregating data types on a genomic scale," *Nature methods,* vol. 11, no. 3, pp. 333-7, 2014.

[24] O. RS, A. H and B. VB, "DDR: efficient computational method to predict drug-target interactions using graph mining and machine learning approaches," *Bioinformatics (Oxford, England),* vol. 34, no. 7, pp. 1164-1173, 2018.

[25] T. Z, G. M, W. C, X. L, W. L and Z. Y, "Constructing an integrated gene similarity network for the identification of disease genes," *Journal of biomedical semantics,* vol. 8, no. 9, pp. (Suppl 1), 32, 2017.

[26] K. YA, C. DY and P. TM, "Understanding Genotype-Phenotype Effects in Cancer via Network Approaches," *PLoS computational biology,* vol. 12, no. 3, p. e1004747, 2016.

[27] W. Y, L. T, X. D, S. H, Z. C, M. YY and W. Z, "Predicting DNA Methylation State of CpG Dinucleotide Using Genome Topological Features and Deep Networks," *Scientific reports,* vol. 6, no. 1, p. 19598, 2016.

[28] Q.-R. HUANG, F. HU, S. HUANG, H.-X. LI, Y.-H. YUAN, G.-X. PAN and W.-J. ZHANG, "Effect of long-term fertilization on organic carbon and nitrogen in a subtropical paddy soil," *Pedosphere,* vol. 19, no. 6, pp. 727-734, 2009.

[29] L. Fu and Q. Peng, "A deep ensemble model to predict miRNA-disease association," *Scientific Reports,* vol. 7, no. 1, 2017.

[30] P. X, F. YX, Y. J and S. HB, "IPMiner: hidden ncRNA-protein interaction sequential pattern mining with stacked autoencoder for accurate computational prediction," *BMC Genomics,* vol. 17, no. 9, p. 582, 2016.

[31] J. Koch-Weser, "Serum drug concentrations in clinical perspective," *Therapeutic drug monitoring,* vol. 3, no. 1, pp. 3-16, 1981.

[32] S. JY, H. H, L. JX, L. P, Z. YN, D. K and Y. SM, "TMFUF: a triple matrix factorization-based unified framework for predicting comprehensive drug-drug interactions of new drugs," *BMC Bioinformatics,* vol. 20, no. 11, p. 19(Suppl 14):411, 2018.

[33] Y. H, M. KT, S. JY, H. H, C. Z, D. K and Y. SM, "Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization," *BMC systems biology,* vol. 12, no. 4, p. (Suppl 1):14, 2018.

[34] C. M, K. N, B. E, L.-F. J and A. BB, "Efficient measurement and factorization of high-order drug interactions in mycobacterium tuberculosis," *Science advances,* vol. 3, no. 10, p. e1701881, 2017.

[35] S. JY, M. KT, Y. H and Y. SM, "Detecting drug communities and predicting comprehensive drug-drug interactions via balance regularized semi-nonnegative matrix factorization," *Journal of cheminformatics,* vol. 11, no. 1, p. 28, 2019.

[36] L. A. G. Camacho and S. N. Alves-Souza, "Social network data to alleviate cold-start in recommender system: A systematic review," *Information Processing & Management,* vol. 54, no. 4, pp. 529-544, 2018.

[37] T. NP, Y. PP, D. R and A. RB, "Data-driven prediction of drug effects and interactions," *Science translational medicine,* vol. 6, no. 4, pp. 125-31, 14 Mar 2012.

[38] W. Zhang, Y. Chen, S. Tu, F. Liu and Q. Qu, "Drug side effect prediction through linear neighborhoods and multiple data source integration," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016.

[39] W. Zhang, Y. Chen, D. Li and X. Yue, "Manifold regularized matrix factorization for drug-drug interaction prediction," *Journal of Biomedical Informatics,* vol. 88, pp. 90-97, 2018.

[40] R. Markello, "snfpy 0.2.2," 3 March 2020. [Online]. Available: https://pypi.org/project/snfpy/. [Accessed 18 December 2023].

[41] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning*, Haifa, Israel, 2010.

[42] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath and B. Kingsbury, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," in *IEEE Signal Processing Magazine*, 2012.

[43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.,* vol. 15, no. 1, p. 1929–1958, 2014.

[44] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray and B. Steiner, "TensorFlow: A System for Large-Scale Machine Learning," in *USENIX Association*, Savannah, GA, USA, 2016.

[45] F. C. et.al., "Keras," keras-team, 28 3 2015. [Online]. Available: https://github.com/keras-team/keras. [Accessed 19 12 2023].

[46] T. Ghosal, V. Edithal, A. Ekbal, P. Bhattacharyya, S. S. S. K. Chivukula and G. Tsatsaronis, "Is your document novel? Let attention guide you. An attention-based model for document-level novelty detection," *Natural Language Engineering,* vol. 27, pp. 427 - 454, 2020.

[47] Y. Toda and F. Okura., "How Convolutional Neural Networks Diagnose Plant Disease," *Plant phenomics (Washington, D.C.),* vol. 2019 9237136, no. 3, 2019.

[48] G. Kim, B. Park and A. Kim, "1-Day Learning, 1-Year Localization: Long-Term LiDAR Localization Using Scan Context Image," *IEEE Robotics and Automation Letters,* vol. 4, no. 2, pp. 1948-1955, 2019.

[49] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations, {ICLR} 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, San Diego, CA, USA, 2015.