

Heterogeneous network propagation with forward similarity integration to enhance drug–target association prediction

Piyanut Tangmanussukum¹, Thitipong Kawichai², Apichat Suratanee^{3,4} and Kitiporn Plaimas^{1,5}

¹ Advanced Virtual and Intelligent Computing (AVIC) Center, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

² Department of Mathematics and Computer Science, Academic Division, Chulachomklao Royal Military Academy, Nakhon Nayok, Thailand

³ Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

⁴ Intelligent and Nonlinear Dynamics Innovations Research Center, Science and Technology Research Institute, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

⁵ Omics Science and Bioinformatics Center, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

ABSTRACT

Identification of drug–target interaction (DTI) is a crucial step to reduce time and cost in the drug discovery and development process. Since various biological data are publicly available, DTIs have been identified computationally. To predict DTIs, most existing methods focus on a single similarity measure of drugs and target proteins, whereas some recent methods integrate a particular set of drug and target similarity measures by a single integration function. Therefore, many DTIs are still missing. In this study, we propose heterogeneous network propagation with the forward similarity integration (FSI) algorithm, which systematically selects the optimal integration of multiple similarity measures of drugs and target proteins. Seven drug–drug and nine target–target similarity measures are applied with four distinct integration methods to finally create an optimal heterogeneous network model. Consequently, the optimal model uses the target similarity based on protein sequences and the fused drug similarity, which combines the similarity measures based on chemical structures, the Jaccard scores of drug–disease associations, and the cosine scores of drug–drug interactions. With an accuracy of 99.8%, this model significantly outperforms others that utilize different similarity measures of drugs and target proteins. In addition, the validation of the DTI predictions of this model demonstrates the ability of our method to discover missing potential DTIs.

Subjects Bioinformatics, Computational Biology

Keywords Heterogeneous network, Network propagation, Similarity measures, Drug–target associations, Drug repurposing, Forward selection algorithm

INTRODUCTION

The identification of drug targets is a very important part of drug development. Not only does it help to gain a better understanding of drug mechanisms and side effects, but it also provides opportunities to enhance drug selectivity and enable drug repurposing (*Butina,*

Submitted 17 June 2022
Accepted 14 September 2022
Published 11 October 2022

Corresponding author
Kitiporn Plaimas,
kitiporn.p@chula.ac.th,
kplaimas@gmail.com

Academic editor
Khalid Raza

Additional Information and
Declarations can be found on
page 25

DOI 10.7717/peerj-cs.1124

© Copyright
2022 Tangmanussukum et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Segall & Frankcombe, 2002; Schenone et al., 2013). There are several wet-lab techniques for identifying drug targets, such as biochemical affinity purification and genetic modifications (*Schenone et al., 2013*). However, only a limited number of drug targets were experimentally discovered and validated due to the high costs and time involved in conducting the experimental labs (*Rudrapal, Khairnar & Jadhav, 2020*).

To alleviate the bottleneck problem of drug target identification, computational inference methods were employed to discover potential drug–target interactions (DTIs). Existing computational methods can be broadly assorted into three categories: ligand-based, docking simulation, and chemogenomic methods (*Bhargava, Sharma & Suravajhala, 2021*). The ligand-based methods use similarities between ligands to infer potential DTIs (*Shaikh et al., 2021; Yang et al., 2021*). For docking simulation methods, three-dimensional (3D) structures of target proteins are usually required to simulate molecular docking (*Al-Karmalawy et al., 2021; Alonso, Bliznyuk & Gready, 2006; Cheng et al., 2007; Pinzi & Rastelli, 2019*). Nevertheless, the limited availability of 3D structural data of target proteins is a big obstruction to these methods. In chemogenomic methods, various properties of drugs and targets are used to predict DTIs via machine learning models and network-based models.

Most of the existing chemogenomic methods are based on the similarity-based technique, which infers promising DTIs based on drug–drug and target–target similarities. Typically, chemical structures of drugs and protein sequences of target proteins are used to formulate drug–drug and target–target similarity measures, respectively. For example, *Bleakley & Yamanishi (2009)* developed a bipartite local model (BLM) using similarities based on chemical and genomic data for predicting unknown DTIs. *Wang, Yang & Li, (2013a)* proposed heterogeneous graph-based inference (HGBI), which uses similarity scores based on drug chemical structures and target protein sequences to form a heterogeneous network and predict new drug–target links by using an information propagation algorithm. *Zhang et al. (2022)* proposed an unsupervised clustering model for DTI predictions based on structural similarity of drugs and protein sequences which combines OPTICS and BGMM algorithms using to detect noise and extract significant interaction pairs.

With the rapid growth of high-throughput experiments and information technology, extensive biological data are available in public databases. Many recent methods have created integrated drug–drug and target–target similarities based on various drug-related and target-related data to enhance the performance of models to predict DTIs. In addition, several techniques were used to integrate the multiple domains of these data, such as linear functions (e.g., average, maximum, etc.) (*Cheng et al., 2013; Kim, A-s & Nam, 2019; Thafar et al., 2020*), the similarity network fusion (SNF) (*Rohani & Eslahchi, 2019; Thafar et al., 2020*), and low-dimensional feature learning (e.g., random walk algorithms) (*An & Yu, 2021; Luo et al., 2017*). However, most of these methods integrate all data regardless of optimal dataset selections. In addition, they used only a single integration function without comparing distinct integration functions to identify an optimal function that fits the data. These probably lead to the non-optimal performance of predictions. Moreover, it has been known that network-based methods have been successfully applied in several applications (*Hengphasatporn et al., 2020; Janyasupab, Suratanee & Plaimas, 2021; Kawichai, Suratanee*

Plaimas, 2021; Suratanee, Buaboocha & Plaimas, 2021; Suratanee et al., 2018; Suratanee & Plaimas, 2014; Suratanee & Plaimas, 2015; Suratanee & Plaimas, 2017; Suratanee & Plaimas, 2018; Suratanee & Plaimas, 2020; Suratanee & Plaimas, 2021). Therefore, in this work, we propose a network-based method with the forward similarity integration (FSI) framework to systemically integrate multiple similarities and predict new links between drugs and targets. A schematic diagram illustrating an overview of this work is shown in Fig. 1. Structural data, molecular interaction and phenotypic data of drugs were collected to generate seven drug–drug similarity measures. Genomic data, molecular interaction and functional data of drug target proteins were used to create nine target–target similarity measures. To combine multiple similarity measures, we considered various integration functions, including both linear integration functions (*i.e.*, average, maximum, and minimum) and non-linear integration function (*i.e.*, SNF). The network-based method with the FSI framework was designed to systematically create a heterogeneous network model with an optimal similarity integration. To predict the links of DTIs, we used the best model constructed by the FSI algorithm and performed network propagation. Finally, we validated our link predictions by searching for literature support. The main contribution of this research is to improve the performance of a heterogeneous network model to predict DTIs. The proposed method is a well-suited algorithm to automatically select the best set of similarities and assign suitable weights for network propagation to yield better scores for unknown DTIs. At present, most of existing methods are based on machine learning with various features involved. However, the principle of the heterogeneous network model is a suitable way to directly propagate the links between drugs and targets. It is also an easy task which can be run faster than the machine learning approaches. Therefore, our proposed algorithm is useful and can be used to improve the heterogeneous network model.

MATERIAL AND METHODS

Data collection

First, we downloaded all known DTIs from DrugBank (Wishart et al., 2018). For each drug and target, various aspects of drug and target data from multiple levels of the biological system, including chemical, molecular, phenotypic, and functional data, were collected. The drug and target protein data were accumulated from several public databases, as shown in Table S1.

For the drug data, we downloaded chemical structures and a list of drug–drug interactions—one drug affects the molecular activity of another—from DrugBank (Wishart et al., 2018). Drug–disease association data, in which drugs exert effects on diseases, were collected from the Comparative Toxicogenomics Database (CTD) (Davis et al., 2021). Moreover, we downloaded the data on drug side effects (SEs)—undesirable effects that occur in addition to therapeutic effects—from SIDER (Kuhn et al., 2016). Both drug–disease associations (DDAs) and SEs can represent the phenotypic information of drugs.

In terms of drug target proteins, we collected protein sequences from DrugBank (Wishart et al., 2018). We also used the protein–protein interaction (PPI) network, downloaded from STRING (Szklarczyk et al., 2021), as the molecular interaction data

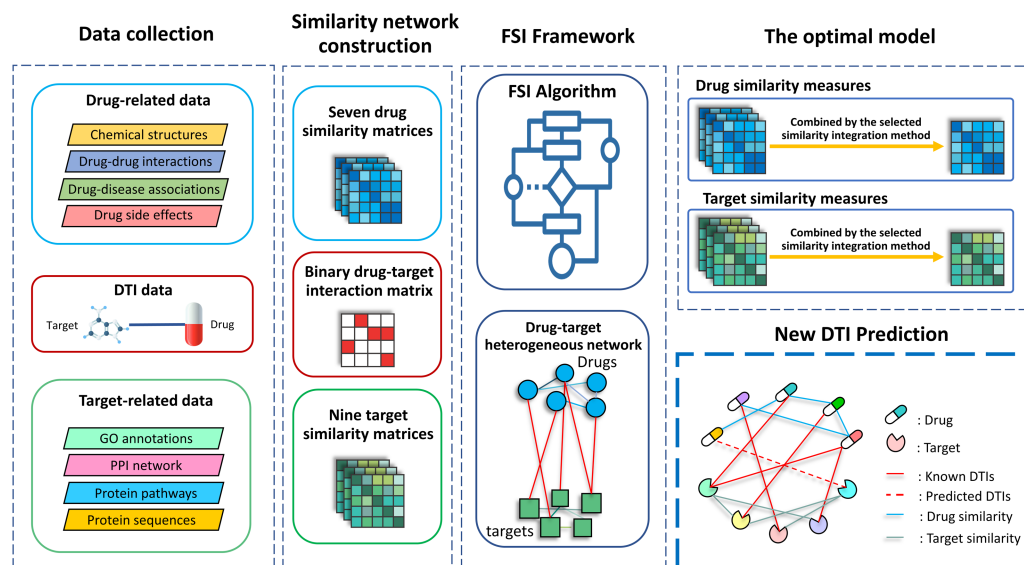


Figure 1 An overview of this study. First, structural data, molecular interaction, and phenotypic data of drugs were collected to generate seven drug–drug similarity measures. Second, genomic data, molecular interaction, and functional data of drug target proteins were used to create nine target–target similarity measures. Third, various integration functions, including both linear integration functions (*i.e.*, average, maximum, and minimum) and non-linear integration functions (*i.e.*, SNF) were applied to combine multiple similarity measures. Then, a heterogeneous network model with an optimal similarity integration was created along with the FSI algorithm framework. The obtained optimal model was then used to predict DTIs and measured the performance of the network propagation model. Finally, our predictions were validated by searching for literature support.

Full-size [DOI: 10.7717/peerjcs.1124/fig-1](https://doi.org/10.7717/peerjcs.1124/fig-1)

of target proteins. We used gene ontology (GO) annotation data from Gene Ontology Annotation (GOA) (Huntley et al., 2009) and used them as the functional information of target proteins. Pathway information related to target proteins was accumulated from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000). Only drugs and target proteins available in all distinct datasets were used. Finally, we obtained 862 drugs, 1,517 target proteins and 3,583 known DTIs.

Drug–drug and target–target similarity measures

Based on four data sets of drugs and four data sets of target proteins, we can formulate seven drug–drug similarity measures and nine target–target similarity measures. The seven drug–drug similarity measures and their defined abbreviations are shown in Table S2. Similarity scores between the two drugs based on their chemical structures (Structures), explained in a simplified molecular-input line-entry system (SMILES) (Weininger, 1988), are computed by the Chemical Development Kit (CDK) (Davis et al., 2021). We used the Tanimoto score (Tanimoto, 1958) to measure structural similarity between two drugs following Eq. (1), where r_i and r_j are binary vectors of the fingerprints of drugs and $\|\cdot\|$ represents the length of a binary vector.

$$S_{\text{Tanimoto}}(r_i, r_j) = \frac{r_i \cdot r_j}{\|r_i\|^2 + \|r_j\|^2 - r_i \cdot r_j}. \quad (1)$$

To measure drug–drug similarity based on DDAs, drug–drug interactions (DDIs), and SEs, we used two well-known similarity indices: the Jaccard index (Jaccard, 1912) and the cosine index (Singhal, 2001). Both similarity indices are frequently applied to measure drug–drug similarity in numerous studies (Gottlieb et al., 2011; Öztürk, Ozkirimli & Özgür, 2016; Perlman et al., 2011; Zhao et al., 2022) due to their simple calculations and similarity measuring efficiencies. With these indices, we obtained six different drug–drug similarity measures based on DDA, DDI, and SE, as shown in Table S2. The Jaccard and cosine similarity indices are defined in Eqs. (2) and (3), where u and v are binary vectors and $\| \cdot \|$ represents the length of a binary vector. For each drug, a value of one in a binary vector represents an existing association with a disease, an existing interaction with another drug, and an existing SE.

$$S_{Jaccard}(u, v) = \frac{u \cdot v}{\|u\|^2 + \|v\|^2 - u \cdot v} \quad (2)$$

$$S_{Cosine}(u, v) = \frac{u \cdot v}{\|u\| \cdot \|v\|}. \quad (3)$$

For target–target similarity measures, we used four data sets of drug target proteins to obtain nine target–target similarity measures. The methods applied to measure the target–target similarities and their defined abbreviations are summarized in Table S3.

Based on the protein sequences of drug targets, we applied the Smith–Waterman algorithm (Smith & Waterman, 1981) and the Needleman–Wunsch algorithm (Needleman & Wunsch, 1970) to produce the scores of local sequence alignments (Seq_Loc) and global sequence alignments (Seq_Glo), respectively. Then, the similarity scores were normalized from 0 to 1 using the formula suggested by Wang, Yang & Li (2013b).

In the STRING (Szklarczyk et al., 2021) database, each protein pair in the PPI network is explained with a confidence score that indicates how much the interaction between two proteins is more likely to be true (Yu & Finley Jr, 2009). To reconstruct a PPI network with more reliable interactions, we kept only the PPIs whose confidence scores were greater than or equal to 0.5. To measure the similarity between the two target proteins, we applied Dijkstra’s algorithm to find the distances of the shortest paths linking the two protein nodes in the PPI network. Then, the distances were transformed to the inverse shortest path similarity (PPI_ISP) scores in Eq. (4), where $S(p, p')$ is the PPI_ISP similarity score between two proteins and $D(p, p')$ is the distance of the shortest path linking between those proteins in the PPI network. According to Perlman et al. (2011), parameters A and b were chosen to be 0.9 and 1, respectively.

$$S(p, p') = Ae^{-bD(p, p')}. \quad (4)$$

In addition to PPI_ISP, we also measured the similarity between two target proteins based on their common neighbors in the PPI network by using the Jaccard index and cosine index, denoted as PPI_Jac and PPI_Cos, respectively.

Based on GO annotations of target proteins, we used GOSemSim (Yu, 2020) in the R package to compute the semantic similarity scores between the two target proteins. The methods used for semantic measurement were the Wang (Yu et al., 2010) and Jiang (Yu et

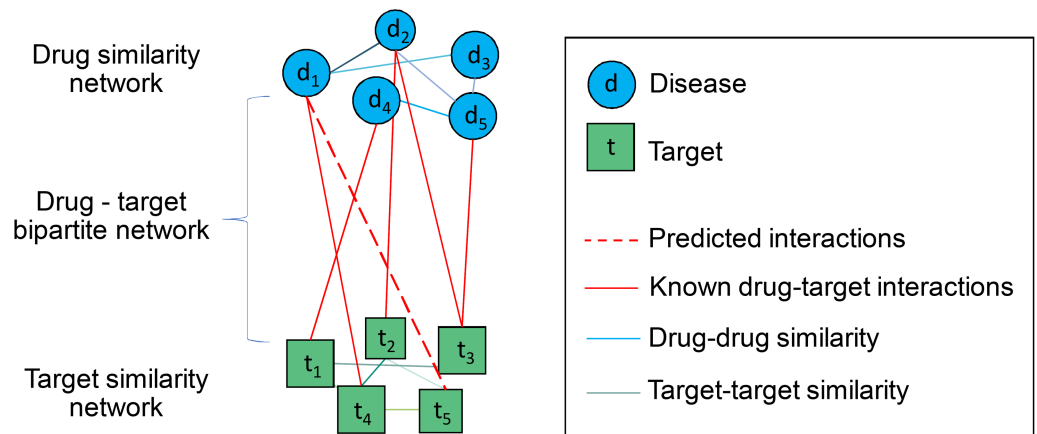


Figure 2 Structure of the drug–target heterogeneous network.

Full-size [DOI: 10.7717/peerjcs.1124/fig-2](https://doi.org/10.7717/peerjcs.1124/fig-2)

et al., 2010) methods. With these two methods, we denoted the target–target similarity scores based on the GO annotations as GO_Wang and GO_Jiang, respectively. Furthermore, we computed the target–target similarity scores based on the list of pathways in which each target protein is involved by using the Jaccard and cosine indices, following Eqs. (2) and (3). According to these two indices, we denoted the similarity measures by PW_Jac and PW_Cos, respectively.

Heterogeneous network propagation

The structure of the drug–target heterogeneous network in this work is illustrated in Fig. 2. This heterogeneous network is composed of three compartmentalized networks: a drug similarity network, a target similarity network, and a drug–target bipartite network.

In the drug similarity network, we used $D = d_1, d_2, d_3, \dots, d_m$ to denote the set of m drug nodes and E_{dd} to denote the set of edges linking between any two drug nodes. Based on a particular drug–drug similarity measure, the similarity scores between two drugs were calculated to represent the edge’s weights W_{dd} . Similarly, let $T = t_1, t_2, t_3, \dots, t_n$ denote the set of n target protein nodes and E_{tt} denote the set of edges in the target similarity network. Based on a defined target–target similarity measure, W_{tt} contains the similarity scores that serve as the edge weights of the target similarity network. In the drug–target bipartite network, let E_{dt} be the set of edges linking between drug and target nodes; W_{dt} denotes by the weights of the edges of the drug–target bipartite network. Each value in W_{dt} can be either one or zero, where one represents the presence of an edge or a DTI and zero represents the absence of an edge or a DTI. The drug–target heterogeneous network (G_{DT}) can be formulated as shown in Eq. (5).

$$G_{DT} = \{D, T\}, \{E_{dd}, E_{tt}, E_{dt}\}, W_{dd}, W_{tt}, W_{dt}. \quad (5)$$

The G_{DT} network is an incomplete graph where there are some missing edges between the drug and target nodes. To infer these missing edges, we employed the heterogeneous

network propagation algorithm, which was broadly utilized in numerous applications, such as the identification of DTIs (Peng, Li & Shang, 2020; Rayhan et al., 2020), the prediction of DDAs (Fiscon & Paci, 2021; He, Yang & Gong, 2020), and other applications (Eswaran et al., 2017; Yang, Zhang & Han, 2019; Yang et al., 2018). In this algorithm, the weights of all networks are considered matrices (i.e., $W_{dd} \in \mathbb{R}^{m \times m}$, $W_{tt} \in \mathbb{R}^{n \times n}$, and $W_{dt} \in \mathbb{R}^{m \times n}$), where m and n are the numbers of drugs and target proteins, respectively. This algorithm iteratively updates the weights in W_{dt} by propagating the weights in the drug similarity network, drug–target network, and target similarity network, as shown in Eq. (6).

$$W_{dt}^{i+1} = \alpha W_{dd} \times W_{dt}^i \times W_{tt} + (1 - \alpha) W_{dt}^0. \quad (6)$$

In this formula, W_{dt}^0 is the matrix of the initial weights of the edges linking drugs and targets, which can be either one or zero following the list of known DTIs. α is a decay factor, that is in the range of 0 to 1. This parameter indicates how much the propagation of the networks' weights affects the newly updated W_{dt} when compared to the effects of W_{dt}^0 . The formulation in Eq. (6) will converge if the W_{dd} and W_{tt} are properly normalized, as shown in Eqs. (7) and (8) (Wang, Yang & Li, 2013b).

$$w(d_i, d_j) = \frac{w(d_i, d_j)}{\sqrt{\sum_{k=1}^m w(d_i, d_k) \sum_{k=1}^m w(d_k, d_j)}} \quad (7)$$

$$w(t_i, t_j) = \frac{w(t_i, t_j)}{\sqrt{\sum_{k=1}^n w(t_i, t_k) \sum_{k=1}^n w(t_k, t_j)}}. \quad (8)$$

Forward similarity integration (FSI)

In the drug–target heterogeneous network, we are interested in integrating multiple similarity measures on the drug and target similarity networks to enhance the performance in predicting DTIs. We propose the FSI algorithm for systematically selecting and combining multiple similarity measures. The implementation of this FSI algorithm and the data table of drug–drug, target–target similarities are provided at https://github.com/piyanuttnk/FSI_framework.

Suppose that there are k different drug–drug similarity measures in set $AD = \{dd_1, dd_2, \dots, dd_k\}$ and l different target–target similarity measures in set $AT = \{tt_1, tt_2, \dots, tt_l\}$. To create an integrated similarity measure by function *IntegrateSim* based on a particular performance measure, the FSI algorithm finds the optimal subsets of drug–drug and target–target similarity measures stepwise, denoted as OD and OT , respectively. The pseudocodes explaining the FSI algorithm are shown in Algorithm 1.

In each step, a drug–drug and a target–target similarity measures are added into sets OD and OT , respectively, to form different heterogeneous network models. These models were compared with their performance (*PRF*) to select the best one, which was added to sets OD and OT . Then, the drug–drug and target–target similarity measures recently added to the optimal subsets are removed from the remaining sets of drug–drug and target–target similarity measures, denoted as RD and RT , respectively. The FSI algorithm iteratively

updates OD and OT by adding a drug–drug and target–target similarity measure until the performance improved no more from the addition, or there were no similarity measures remaining in RD or RT . Therefore, we anticipate that a heterogeneous network model formed with the integrated similarity measures obtained by the FSI algorithm will show the best performance in predicting DTIs. However, the optimal subsets of similarity measures obtained by the FSI (OD and OT) may depend on the performance measure used to evaluate the heterogeneous network models and the integration method applied. In this study, we considered several performance measures used in the FSI and several integration functions to create an integrated similarity matrix from multiple similarity measures.

Algorithm 1: Forward Similarity Integration (FSI)

Input: A set of all drug–drug similarity measures (AD) and target–target similarity measures (AT)

Output: An optimal subset of drug–drug similarity measures (OD) and target–target similarity measures (OT), which are orderly integrated to obtain the most optimally combined drug–drug and target–target similarity measures

1. **Initialize** $k = 0$, OD_0 , $OT_0 = \emptyset$, $RD_0 = AD$, $RT_0 = AT$, $PRF_0 = 0$.
2. **repeat**
3. $k = k + 1$
4. $x^*, y^* = \text{argmax}_{x \in RD_{k-1}, y \in RT_{k-1}} [\text{EstimatePRF}(\text{IntegrateSim}(OD_{k-1} \cup \{x\}), \text{IntegrateSim}(OT_{k-1} \cup \{y\}))]$ // Add both drug and target similarity
5. **Denote** the best performance as PRF_{both} , $X^* = \{x^*\}$, and $Y^* = \{y^*\}$
6. **if** $k == 1$ **then**
7. $PRF_k = PRF_{both}$
8. **else** // Also consider adding only drug or target similarity
9. $x_d^* = \text{argmax}_{x \in RD_{k-1}} [\text{EstimatePRF}(\text{IntegrateSim}(OD_{k-1} \cup \{x\}), \text{IntegrateSim}(OT_{k-1}))]$ **and denote** the best performance as PRF_{drug}
10. $y_t^* = \text{argmax}_{y \in RT_{k-1}} [\text{EstimatePRF}(\text{IntegrateSim}(OD_{k-1}), \text{IntegrateSim}(OT_{k-1} \cup \{y\}))]$ **and denote** the best performance as PRF_{target}
11. $PRF_k = \max (PRF_{both}, PRF_{drug}, PRF_{target})$
12. **if** $PRF_k == PRF_{drug}$ **then**
13. $X^* = \{x_d^*\}$ and $Y^* = \emptyset$
14. **else if** $PRF_k == PRF_{target}$ **then**
15. $X^* = \emptyset$ and $Y^* = \{y_t^*\}$
16. **end if**
17. **end if**
18. **if** $PRF_k > PRF_{k-1}$ **then**
19. **Update** $OD_k = OD_{k-1} \cup X^*$, $OT_k = OT_{k-1} \cup Y^*$, $RD_k = RD_{k-1} - X^*$, $RT_k = RT_{k-1} - Y^*$
20. **end if**
21. **until** ($PRF_k \leq PRF_{k-1}$ **or** $|RD_k| == 0$ **or** $|RT_k| == 0$)
22. return the latest OD and OT

Based on a subset of similarity measures, we can create an integrated similarity matrix to form a heterogeneous network model following Algorithm 2. Let $SM = \{s_1, s_2, \dots, s_p\}$ be a

set of given similarity measures that are arranged according to the order of integration, Xs_i be a matrix containing the similarity scores based on similarity measure s_i , and f represents a function to combine two similarity matrices. The algorithm used to create an integrated similarity matrix (*IntSimMat*) based on a set of given similarity measures (SM) and an integration function f , is demonstrated in Algorithm 2, where $\|\cdot\|$ is the number of elements in a set.

Algorithm 2: Integrating multiple similarity matrices

Input: A set of given similarity measures (SM) and a function of an integration method (f)

Output: An integrated similarity matrix (*IntSimMat*)

1. **for** $i = 1$ **to** $i = |SM| - 1$
2. **if** $i == 1$ **then**
3. $IntSimMat = f(Xs_i, Xs_{i+1})$
4. **else**
5. $IntSimMat = f(IntSimMat, Xs_{i+1})$
6. **end if**
7. **end for**

According to Algorithm 2, Xs_i can be either a drug–drug similarity matrix (*i.e.*, $Xs_i \in \mathbb{R}^{m \times m}$) or a target–target similarity matrix (*i.e.*, $Xs_i \in \mathbb{R}^{n \times n}$). In this study, we included both linear and non-linear functions for integrating multiple similarity matrices: functions average (AVG), maximum (MAX), minimum (MIN), and SNF (Wang et al., 2014). AVG, MAX, and MIN are defined in Eqs. (9) to (11), where $A = [a_{kl}]$ and $B = [b_{kl}]$ are two different matrices of similarity scores.

$$AVG(A, B) = \left[\frac{a_{kl} + b_{kl}}{2} \right] \quad (9)$$

$$MIN(A, B) = [min(a_{kl}, b_{kl})] \quad (10)$$

$$MAX(A, B) = [max(a_{kl}, b_{kl})]. \quad (11)$$

SNF (Wang et al., 2014) is a method that calculates and fuses similarity networks into a single similarity network. The SNF utilizes an iterative non-linear approach and updates the global similarity network of each layer using a local k -nearest neighbors (k NN) approach. A similarity value between nodes is propagated to the k NN.

In formal, $W^{(m)}$ is denoted as the similarity matrix for the m th data type. Initially, a transition probability matrix between all nodes is defined in Eq. (12), and a transition probability matrix between nearest neighbors is defined in Eq. (13), where N_i is a set of node i 's k NN in $W^{(m)}$ matrices.

$$P_0^{(m)}(i, j) = \begin{cases} \frac{W^{(m)}(i, j)}{2 \sum_{k \neq i} W^{(m)}(i, k)}, & j \neq i; \\ \frac{1}{2}, & j = i. \end{cases} \quad (12)$$

$$S^{(m)}(i, j) = \begin{cases} \frac{W^{(m)}(i, j)}{\sum_{k \in N_i} W^{(m)}(i, k)}, j \in N_i; \\ 0, \text{Otherwise.} \end{cases} \quad (13)$$

P is iteratively updated using the values passing between the nearest neighbors following Eq. (14), where $P_q^{(m)}$ is the matrix for m th data type at iteration q and M is the number of data types. Finally, the overall status matrix at iteration t^{th} is calculated, as shown in Eq. (15).

$$P_{q+1}^{(m)} = S^{(m)} \frac{\sum_{k \neq m} P_q^{(k)}}{M-1} S^{(m)q} \quad (14)$$

$$P^{(t)} = \frac{\sum_{m \in M} P_q^{(m)}}{M}. \quad (15)$$

Experimental environments

Our experiments can be divided into four parts. The first part is an identification of the optimal decay factor to reduce the model variable. The second part is a selection of the heterogeneous network model to find the optimal model by the FSI. The third part is the performance comparison between the optimal model and other models, including full similarity integration, random similarity integrations, and the original heterogeneous network (OHN). The final part is an identification of new DTIs to predict new links between drugs and target proteins. All experiments were conducted on a 4.90-GHz core i7 laptop with 20 GB of ram under the environments of Python version 3.9. To estimate performance of each model, we used several packages comprising NumPy version 1.22.4, Pandas version 1.4.2, Scikit-learn version 1.1.2, Matplotlib version 3.5.2 and SNFPy version 0.2.2. To fairly compare the performance of different models, such as models with different values of the decay factors used or with different similarity integrations, training and testing data sets were controlled as the same.

Performance evaluation

To evaluate the performance of all heterogeneous network models, we performed 10-fold cross-validation. All DTIs were randomly divided into 10 equal parts, where each part of positive and negative interactions was treated as test data in turn, and the remaining nine parts were used as training data. To completely evaluate the performance of all heterogeneous network models, we used general metrics, including the area under a precision–recall curve (AUPR), the area under a receiver operating characteristic curve (AUC), the F1-measure (F1), the precision (PRE), the recall (REC), the accuracy (ACC) and the Matthews correlation coefficient (MCC). These measures can be calculated as shown in Eqs. (16) to (20). TP, FP, FN, and TN denote the true positive, false positive, false negative, and true negative, respectively, which compares the predicted and actual classes.

$$PRE = \frac{TP}{TP + FP} \quad (16)$$

$$REC = \frac{TP}{TP + FN} \quad (17)$$

$$F_1 \text{ measure} = \frac{2 \cdot PRE \cdot REC}{PRE + REC} \quad (18)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad (20)$$

RESULTS

Data summary

According to all the collected datasets of drugs and target proteins, we constructed a heterogeneous network with 862 drug nodes and 1,517 target nodes. All drug data (*i.e.*, chemical structures, DDIs, DDAs, and SEs) and all target data (*i.e.*, protein sequences, GO functions, PPI-based information, and pathway information) were collected. The network comprised 3,583 known DTIs linking drug and target nodes. This number of links is only 0.2740% of all possible edges linking drugs and targets. Thus, it is possible that many potential DTIs are still undiscovered in this dataset.

Figure S1 shows the degree distributions of the drugs and target proteins in the drug–target bipartite network, which can explain the overall picture of how many target proteins are associated with a drug and how many drugs are associated with a target protein according to the known DTIs. Notably, most drugs and target proteins are associated with a few target proteins and drugs, respectively. We found that 19.48% of the drugs and 38.85% of the target proteins had node degrees less than 5. The maximal degrees of the drug and target nodes were 61 and 40, respectively. A drug can bind to approximately four target proteins on average, and a target protein can interact with approximately two drugs on average. Additionally, the majority of nodes are rarely connected to one another while only a few nodes have dense links. This suggests that there would be many undiscovered links between drugs and target proteins in the drug–target bipartite network.

Correlation analysis of similarity measures

In this section, we performed the Pearson correlation analysis between different drug–drug similarity measures and between different target–target similarity measures to preliminarily validate the orthogonality of the defined similarity measures. A Pearson correlation coefficient (ρ) is a measure of linear correlation between two data sets and assigns a value between -1 and 1 , where 0 is no correlation, 1 is a total positive correlation, and -1 is a total negative correlation. The Pearson correlation coefficients that we calculate for all pairs of drug–drug and target–target similarity measures are shown in Fig. 3.

Usually, drug–drug similarity matrices and target–target similarity matrices are slightly positively correlated with one another. This suggests that drug and target similarity

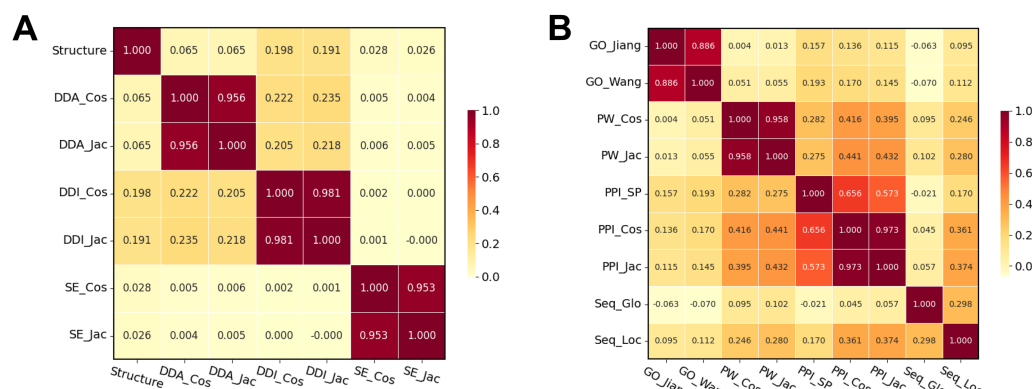


Figure 3 Heatmaps of the Pearson correlation coefficients of the drug and target similarity measures. (A) Drug similarity measures. (B) Target similarity measures.

Full-size [DOI: 10.7717/peerjcs.1124/fig-3](https://doi.org/10.7717/peerjcs.1124/fig-3)

measures provide information that complements that of one another, which is useful when integrating them for predicting DTIs. However, the similarity measures based on the same datasets, such as DDI_Cos and DDI_Jac; PPI_ISP, PPI_Jac, and PPI_Cos, have their own highly positive correlation values. Thus, we avoid integrating drug or target similarity measures based on the same dataset in the FSI algorithm despite the different methods used to compute the similarity scores.

Interestingly, the drug similarity measures based on DDIs, chemical structures, and DDAs are negligibly positively correlated together ($\rho \leq 0.3$). This suggests that the structural information of drugs is partly signified in DDIs and DDAs. Moreover, the target similarity measures based on PPI, protein sequences, and pathway information are slightly to moderately positively correlated together ($\rho \leq 0.5$).

Identification of the optimal decay factor

In the process of heterogeneous network propagation, one parameter required to be optimally tuned is the decay factor (α). This parameter indicates how much network propagation affects the weight updates of the edges linking the drug and the target nodes. Most studies (Luo et al., 2016; Yang et al., 2019) set the value of the decay factor at 0.4, which was suggested by Wang, Yang & Li (2013a). To discover the best model with an optimal similarity integration, we built numerous heterogeneous network models with different datasets combined and different integration methods used. With distinct data combinations, these models may have suitable values for the decay factor differently. To reduce the model variables, we preliminarily specified the value of the decay factor for all heterogeneous network models.

Since there are seven drug–drug similarity measures and nine target–target similarity measures, we have 5,671 possible combinations of drug and target similarity measures to construct heterogeneous network models with different similarity integration. In this experiment, we randomly selected 10% of all possible combinations, resulting in 567 combinations of drug and target similarity measures. For each combination with multiple drugs and target similarity measures, we integrated those drug and target similarity

Table 1 The FSI algorithm with different similarity integration methods and the performance measures used to select drug and target similarities.

Similarity integration method	Performance measure for selecting drug/target similarities		
	AUC	AUPR	F1
AVG	Model 1 {DDA_Jac} & {PPI_Jac, Seq_Loc}	Model 2 {DDA_Jac} & {Seq_Loc}	
MAX	Model 3 {DDA_Jac} & {PPI_Jac}		
MIN	Model 4 {DDA_Jac, Structures, DDA_Cos} & {PPI_Jac, Seq_Loc}	Model 5 {DDA_Jac, SE_Jac, Structures, DDI_Jac} & {Seq_Loc}	
SNF	Model 6 {DDA_Jac, DDI_Jac, PPI_Jac} & {Seq_Loc, PW_Jac}	Model 7 {DDA_Jac, DDI_Cos, Structures} & {Seq_Loc}	Model 8 {DDA_Jac, DDI_Jac, Structures} & {Seq_Loc}

measures by using all integration functions (*i.e.*, AVG, MAX, MIN, and SNF) to construct heterogeneous networks. In total, we have 567 different heterogeneous network models. In the network propagation of each model, we varied the values of the decay factor from 0.1 to 1, with a step of 0.1 and conducted 10-fold cross-validation to evaluate the model performance when using each value of the decay factor. For each model, we investigated the AUC, AUPR, and F1 values and used each of these metrics to select the optimal value of the decay factor. To compare all values of the decay factor, we counted the numbers of the models resulting in high performances in each decay factor value. The performances that we calculated were AUC, AUPR, and F1. The results are shown in Fig. 4.

Figure 4A shows that 85.23% of the models reach the maximum AUC values when setting the decay factor as 0.1. In Fig. 4B, 42.02% and 34.35% of the models yield maximum AUPR values when using decay factors of 0.9 and 0.1, respectively. Moreover, 38.67% and 26.37% of the models yields the maximum F1 scores when exploiting the decay factor as 0.9 and 0.1, respectively, as shown in Fig. 4C. To select the value of the decay factor that yields the best AUC, AUPR, and F1 values, we find the total number of models with the maximum AUC, AUPR, and F1 values as shown in Fig. 4D. As a result, it clearly shows that a decay factor of 0.1 yields the best coverage of the models with maximum AUC, AUPR, and F1 values. Therefore, we specified the value of the decay factor as 0.1 in the network propagation of all heterogeneous network models.

Selection of the heterogeneous network model

Based on a given method of similarity integration, we introduce the FSI algorithm for discovering the best sets of drug and target similarity measures, which are used to combine and form a heterogeneous network model. Herein, we investigated three comprehensive evaluation metrics—AUC, AUPR, and F1—to serve as the criteria to select drug and target similarities to integrate into a heterogeneous network model. With the different similarity integration methods and evaluation metrics used in the FSI, we obtained eight different models as shown in Table 1.

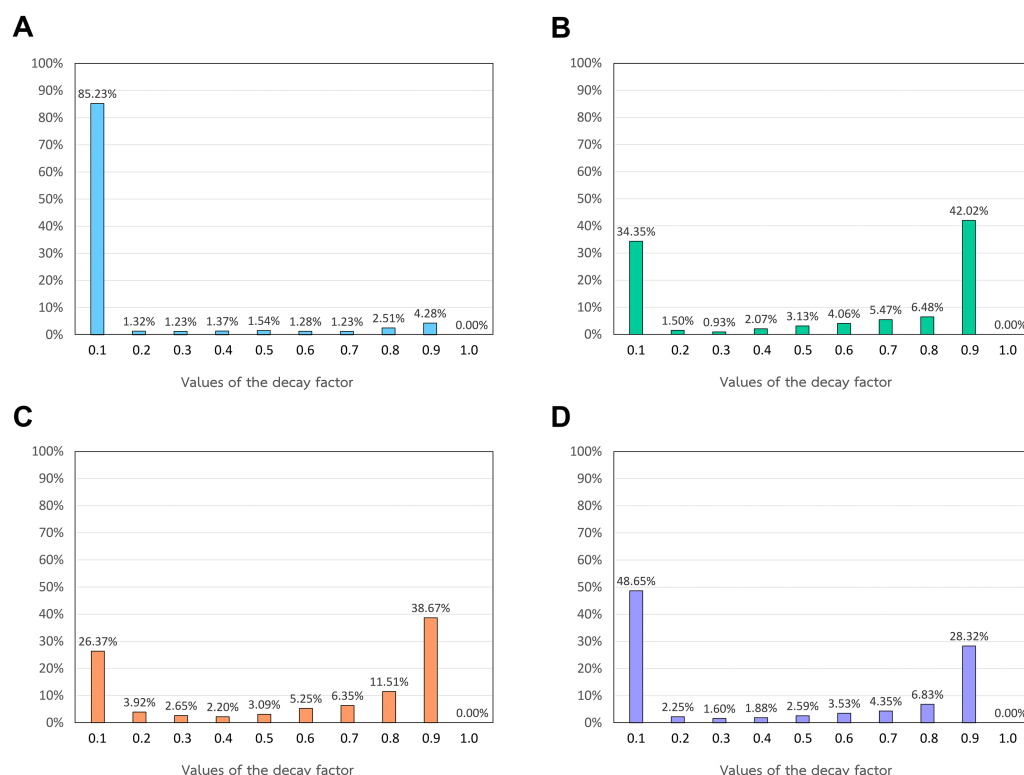


Figure 4 The percentage of models yielding high performances in different values of the decay factor. (A) The percentage of the models yielding the maximum AUC values. (B) The percentage of the models yielding the maximum AUPR values. (C) The percentage of the models yielding the maximum F1 values. (D) The percentage of the models yielding the maximum AUC/AUPR/F1 values (%).

[Full-size DOI: 10.7717/peerjcs.1124/fig-4](https://doi.org/10.7717/peerjcs.1124/fig-4)

As shown in Table 1, the FSI algorithm provides various output models generated by different integrations of drug and target similarities (either average (AVG), maximum (MAX), minimum (MIN), or similarity network fusion (SNF)) and different performance measures (AUC, AUPR, and F1 score). Each row in the table shows the best combination of similarities and integrations with the highest performance measures in column. For example, with the use of average (AVG) as a similarity integration method and the use of AUC as a performance evaluation, Model 1 using DDA_Jac, PPI_Jac for drug similarities, and Seq_Loc for target similarities yield the best AUC performance comparing to the other combinations of the similarity measures. The FSI presents the same or similar models when using AUPR or F1 as a performance evaluation for selecting drug and target similarities with either AVG, MAX, or MIN integration methods. We obtained the same model, formed by using DDA_Jac and Seq_Loc (Model 2) and the integration method AVG or MAX. FSI also returns Model 5 with using DDA_Jac, SE_Jac, Structures, and DDI_Jac for drug similarity and using Seq_Loc as a target similarity. Noticeably, DDA_Jac and Seq_Loc were always used in all obtained models because the models were preliminarily identified as the best by the FSI algorithm. Since this is an initial model where the FSI algorithm additionally

Table 2 Performance of eight FSI models and results of *t*-tests.

Model no.	AUPR	AUC	PRE	REC	F1	ACC	MCC
1	0.227	0.951	0.342	0.333	0.333	0.996	0.334
2	0.267 ^{a,b}	0.935	0.389 ^{a,b}	0.373 ^{a,b}	0.379 ^{a,b}	0.997 ^{a,b}	0.378 ^{a,b}
3	0.140	0.947	0.192	0.284	0.221	0.994	0.226
4	0.233	0.953	0.310	0.400	0.348	0.996	0.349
5	0.335 ^c	0.926	0.426 ^c	0.443 ^c	0.434 ^c	0.997 ^c	0.433 ^c
6	0.294	0.958	0.334	0.422	0.367	0.996	0.370
7	0.481^{d,e}	0.933 ^e	0.578^d	0.508 ^d	0.539^{d,e}	0.998^d	0.540^{d,e}
8	0.481	0.933	0.564	0.515	0.538	0.998	0.537

Notes.

The maximum value is shown in bold numbers.

^aSignificantly greater than the mean value of Model 1 at a significance level of 0.05.

^bSignificantly greater than the mean value of Model 3 at a significance level of 0.05.

^cSignificantly greater than the mean value of Model 4 at a significance level of 0.05.

^dSignificantly greater than the mean value of Model 6 at a significance level of 0.05.

^eSignificantly greater than the mean value of Model 8 at a significance level of 0.05.

selected other drug and target similarities into the heterogeneous model to increase the performance for DTIs predictions.

From these eight models in Table 1, we estimated the performance of each model by performing 10-fold cross-validation. The mean values of all evaluation metrics for each model are shown in Table 2. When we compared Model 1 and Model 2 by conducting *t*-tests, we found that Model 2 yielded significantly higher performances than Model 1 at a significance level of 0.05. Similar results were shown when we compared Model 1 and Model 3 or compared between Model 4 and Model 5. Moreover, the overall performance of Model 7 was significantly greater than that of Model 6 and 8 in most evaluation metrics, such as AUPR, F1, and MCC, at a significance level of 0.05. If we compared the models using the same integration methods, we found that using AUPR as a performance measure to select and integrate drug and target similarities to into a heterogeneous network model would produce a model with better performance than those using AUC or F1. Interestingly, Model 3 choose DDA_Jac and PPI_Jac as the optimal features when selecting the maximum as an integration method. Both similarities for drugs and targets are based on Jaccard indices. The maximum values of the overlap among the neighboring drugs and targets in the interaction network were much more important to yield better AUC. This might be because of the functional relation among drugs and their neighbors, as well as among targets and their similar neighboring targets. Model 4 provided DDA_Jac, Structures and DDA_Cos for drug similarities, and PPI_Jac and Seq_Loc for target similarities. Based on DDA_Jac and PPI_Jac in Model 4 with taking the maximum value, only one similarity was selected while in Model 3 with taking the minimum values, two or three similarities were selected. Therefore, taking the minimum value among the selected similarities could narrow the possibility of predicting an unrelated drug-target links with high score. In the better way, we should consider Structure and DDA_Cos also in the similarity integration for drugs and consider Seq_Loc more in the similarity integration for targets.

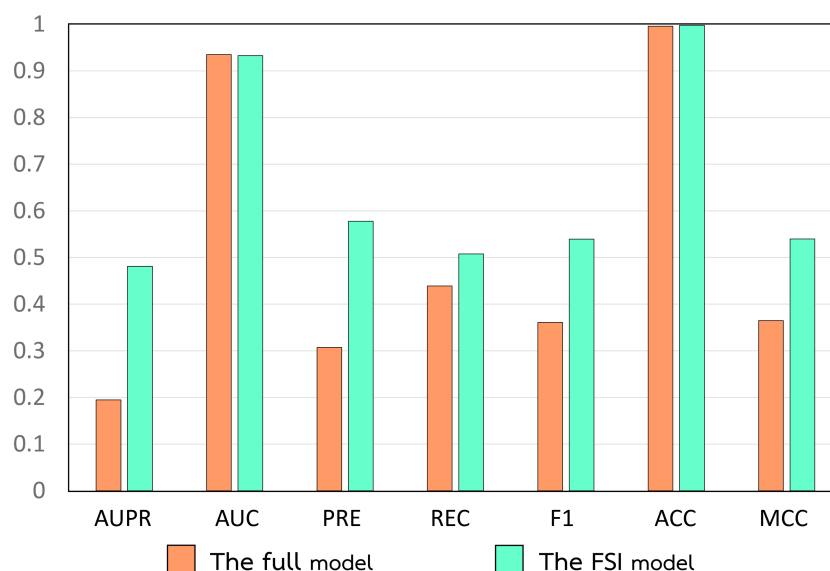


Figure 5 Performance comparison of the full model and the FSI.

Full-size [DOI: 10.7717/peerjcs.1124/fig-5](https://doi.org/10.7717/peerjcs.1124/fig-5)

According to the results in Table 2, it is noticeable that the maximum performance values are mostly of the models using SNF as a similarity integration method (Model 6, Model 7, and Model 8). Thus, we compared the performances of these models to select the best one. By performing t -tests, the mean values of almost all evaluation metrics, except that of AUC, of Model 7 are significantly greater than those of Model 6 at a significance level of 0.05. If we compare Model 7 and Model 8, the mean values of AUPR, AUC, F1, and MCC of Model 7 are significantly higher than those of Model 8 at a significance level of 0.05. Therefore, Model 7, which integrates DDA_Jac, DDI_Cos, and Structures into a drug similarity network by SNF and uses Seq_Loc as a target similarity, was selected as the best FSI model.

Performance evaluation of an optimal model

In this section, we demonstrate the superior performance of the heterogeneous network model with the drug and target similarity integration selected by the FSI, termed the FSI model. This model combines DDA_Jac, DDI_Cos, and Structures as an integrated drug similarity using SNF and employs Seq_Loc as a target similarity. To verify that the model selected by FSI is the best, we compared its performance to those of the models with full similarity integration, random similarity integrations, and the original heterogeneous network (OHN) model, which uses only Structures and Seq_Loc, proposed by Liu et al. (2022), Liu et al. (2016), Peng et al. (2022) and Wang, Yang & Li (2013a). To demonstrate the efficiency of FSI, we first compare the performance of the FSI model with that of the full model which fuses all drug similarities and target similarities, as shown in Fig. 5.

As shown in Fig. 5, the FSI model performs much better than the full model. By t -tests, the mean values of all evaluation metrics of the FSI model, except AUC, were significantly greater than those of the full model at a significance level of 0.01, where p -values when

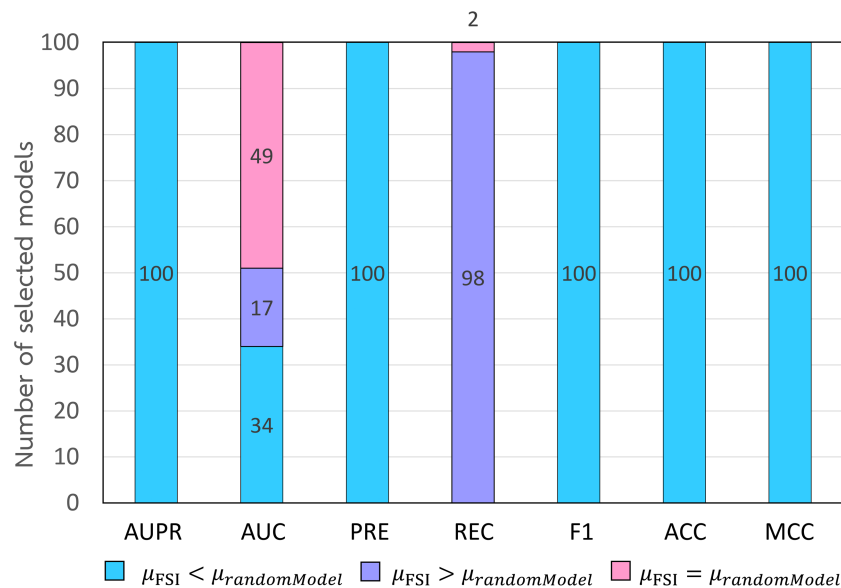


Figure 6 Performance comparison between 100 random integrating models and the FSI model.

Full-size [DOI: 10.7717/peerjcs.1124/fig-6](https://doi.org/10.7717/peerjcs.1124/fig-6)

considering AUPR, AUC, PRE, REC, F1, ACC, and MCC were 2.83E-13, 0.172, 3.10E-09, 2.79E-06, 1.99E-11, 1.81E-09, and 1.98E-11, respectively. This means that a model formed by selecting only some advantageous drug and target similarities by FSI is more efficient than a model that integrates all existing drug and target similarities without cautious consideration.

In addition, we compared the performance of the FSI model with that of random similarity integrations. We randomly selected drugs and target similarities 100 times to integrate and construct 100 different models. Then, we compare the performance of the FSI model with that of 100 random integrating models. The results of the comparison are shown in Fig. 6.

Based on each evaluation metric, Fig. 7 shows the results of the t -tests that compare the mean value of each evaluation metric of the FSI model with that of each random integrating model. In Fig. 6, the labeled numbers are the numbers of t -tests, resulting in the mean value of an evaluation metric of the FSI model being greater than (light blue bar) or lower than (purple bar) that of a random integrating model at a significance level of 0.05. The labeled numbers in the pink bar are the numbers of t -tests, resulting in the mean value of an evaluation metric of the FSI model not being different when compared to that of a random integrating model at a significance level of 0.05. The results showed that the mean values of all evaluation metrics of the FSI model, except AUC and REC, were significantly greater than those of the 100 random integrating models at a significance level of 0.05. This demonstrates that the model formed by systemically selecting drug and target similarities

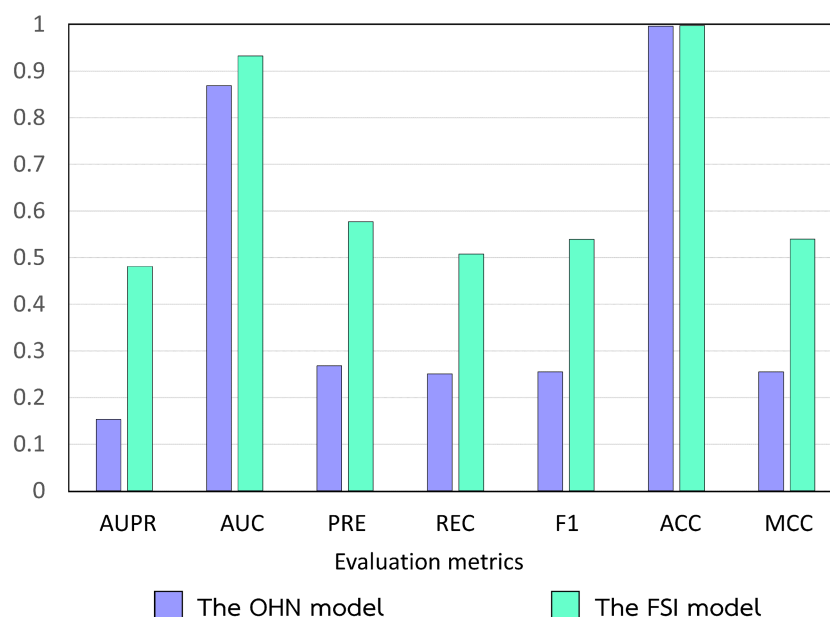


Figure 7 Performance comparison of the conventional model and the FSI model.

Full-size DOI: [10.7717/peerjcs.1124/fig-7](https://doi.org/10.7717/peerjcs.1124/fig-7)

by FSI is more efficient than a model that randomly selects drug and target similarities without cautious consideration.

To predict DTIs, the network models utilize the drug–drug similarity derived only from drug chemical structures and the target–target similarity based only on local sequence alignments of target proteins. To show the superior performance of the model optimally integrating multiple drug and target similarities by FSI, we compared the performances of the OHN model (Liu et al., 2022; Liu et al., 2016; Peng et al., 2022; Wang, Yang & Li, 2013a) with those of the FSI model (Fig. 7). Using *t*-tests, we found that the FSI model performed significantly better than the OHN model at a significance level of 0.01, no matter what evaluation metrics were compared. The *t*-test *p*-values when considering AUPR, AUC, PRE, REC, F1, ACC, and MCC were 2.09E-11, 7.23E-08, 2.25E-09, 7.69E-11, 5.73E-11, 1.23E-06, and 5.53E-11, respectively. This means that integrating DDA_Jac and DDI_Cos into a drug–drug similarity derived from only drug chemical structures could greatly improve the performance of the OHN model in predicting DTIs.

Significance of the drug similarity integration

In the FSI model, there were not only Structures and Seq_Loc used, but also two more similarities named DDA_Jac and DDI_Cos included. To verify the importance of these additional drug similarity measures for DTI predictions, we compared the performance of the FSI model with that of the removed DDA_Jac, namely the reduced DDA_Jac model, and removed DDI_Cos, namely the reduced DDI_Cos model. Moreover, we compared the performance of the FSI model with that of the models that permuted the DDA matrix, namely the permuted DDA model and the DDI matrix, namely the permuted DDI model.

Table 3 Performance comparison of the OHN model and the reduced model.

Model	AUC		AUPR		F1	
	average	p-value	average	p-value	average	p-value
The FSI model	9.33E-01	—	4.81E-01	—	5.39E-01	—
The reduced DDA_Jac Model	9.22E-01 ^a	6.46E-08	4.72E-01 ^a	3.65E-06	5.31E-01 ^a	3.91E-05
The reduced DDI_Cos Model	9.24E-01 ^a	2.47E-07	4.71E-01 ^a	1.43E-06	5.28E-01 ^a	1.36E-06
Permuted DDA Model	9.32E-01 ^a	9.9E-03	4.81E-01 ^b	2.82E-02	5.38E-01 ^{ns}	7.84E-02
Permuted DDI Model	9.14E-01 ^a	5.19E-09	4.69E-01 ^a	1.13E-06	5.24E-01 ^a	1.03E-05
Permuted DDA and DDI Model	9.14E-01 ^a	5.20E-09	4.69E-01 ^a	1.11E-06	5.24E-01 ^a	8.76E-06

Notes.
^aMean value of the FSI model is Significantly greater than those models at a significance level of 0.01.

^bMean value of the FSI model is Significantly greater than those models at a significance level of 0.05.

^{ns}Not significance

The mean AUPR, AUC, and F1 values of each model and the results of the *t*-tests are shown in Table 3.

Using *t*-tests, the mean values of AUC (0.4811), AUPR (0.4724), and F1 (0.4711) of the FSI models were greater than those of all the reduced models at a significance level of 0.01. This result showed that both DDA_Jac and DDI_Cos were important to DTI predictions, and we could not remove any of them from the heterogeneous network model. To verify the significance of the existing relationships in the DDA and DDI data, we generated the permuted models by randomly shuffling the existing edges in the DDA and DDI matrices, and then reconstructed heterogeneous network models with the permuted DDA_Jac and permuted DDI_Cos. Consequently, the mean AUC, AUPR, and F1 values of the FSI model were greater than those of the permuted DDI model and the model of both permuted DDAs and DDIs at a significance level of 0.01. Nevertheless, it cannot be concluded that the FSI model performs better than the permuted DDA model, especially when considering the mean F1 values. This may be due to the large sparsity of the DDA matrix, resulting in the fact that permuting this matrix does not change the DDA_Jac from the original one.

Comparison with the random forest

To show the performance of our final model derived from the FSI algorithm to other methods, we employed a random forest with the features based on only the similarities used in this study. The results show that all accuracies of the heterogeneous network model were better than those of the random forest using the same sets of similarities as the features for training and testing, as shown in Table 4. Using the original similarities (which are Structure and Seq_Loc similarities), the random forest yielded better in term of AUC. Taking all similarities, the random forest and the heterogeneous network model showed similar AUC but different levels of accuracy where the heterogeneous network is much better. Furthermore, the heterogeneous network model with the obtained optimal similarities (from our FSI method) provided better performance than the random forest in terms of both accuracy and AUC. Obviously, the FSI algorithm provided an optimal set of similarities which improves the performance of the heterogeneous network model.

Table 4 Performance of heterogeneous network model and random forest.

	Heterogeneous network model		Random Forest	
	Accuracy	AUC	Accuracy	AUC
Original similarities (OHN model)	0.9650	0.8850	0.8322	0.9032
All similarities (Full model)	0.9780	0.9330	0.8737	0.9382
Optimal similarities (FSI model)	0.9980	0.9330	0.8533	0.9215

Identification of new DTIs

In this section, the FSI model was used to predict new links between drugs and target proteins. We demonstrated the validation of the discovered DTIs through two types of case studies which were the case of target proteins with one known drug and the case of drugs with one known target protein. In the first case, we could obtain new drugs for a target protein with only one known drug. In the second case, we could predict new target proteins for a drug with only one known target. To find the list of new drugs for a certain target having only one known drug involved, we gathered the high-score predicted drug-target links. Three common targets which have about five proposed drugs were selected for more further verification and analysis (in Table 5). In the same manner, to find the list of new targets for a drug of interest having only one known target involved, in the high-score predicted drug-target links, common drugs having at least five proposed targets were selected as shown in Table 5. Three selected target proteins were tubulin beta-3 chain (Q13509), nicotinic acetylcholine receptor alpha-1 (P02708), and calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1C (Q14123), and two selected drugs were nicorandil (DB09220) and plerixafor (DB06809).

As shown in Table 5, the first row was the protein tubulin beta-3 chain (Q13509). Lxabepilone (DB04845) can bind to this protein and cause a reduction of improper cell division caused by several types of cancer, such as breast cancer, lung cancer, and lymphoma. From our predictions, the tubulin beta-3 chain could be associated with some drugs used in chemotherapy for cancer, including vincristine (DB00541), vinblastine (DB00570), and vinflunine (DB11641). Many studies have reported that these drugs interact with tubulins, such as for vincristine (Arora & Menezes, 2021; Lobert, Vulevic & Correia, 1996), for vinblastine (Arora & Menezes, 2021; Lobert et al., 1998), and for vinflunine (Aparicio, Pulido & Gallego, 2012; Arora & Menezes, 2021). Furthermore, beta-3 tubulin is predicted to be involved with albendazole (DB00518) and mebendazole (DB00643), which are drugs for helminth infections. It has been reported that both drugs interact with the alpha-1A and beta-4B tubulins (Berman et al., 2000; Chu et al., 2009; Ramírez et al., 2001; Solana et al., 2002). Thus, both albendazole and mebendazole could bind to a highly similar beta-3 tubulin protein.

Next, the nicotinic acetylcholine receptor alpha-1 or nAChR $\alpha 1$ (P02708) is a protein that has been reported as a target protein for lamotrigine (DB00555). Lamotrigine is an antiepileptic drug approved for the treatment of epilepsy and bipolar disorder (Nishimura et al., 2010; Vallés, Garbus & Barrantes, 2007). When lamotrigine binds to nAChRs, voltage-dependent sodium channels on this protein are blocked and prevent the release of excitatory

Table 5 Target proteins and drugs from high-score predicted links with their possible drugs and targets, respectively.

Targets	Known drug	Predicted drug (DrugBank ID)
Tubulin beta-3 chain (Q13509)	Ixabepilone (DB04845)	DB00541* , DB00570* , DB11641* , DB00518 , DB00643 ,
Nicotinic acetylcholine receptor alpha-1 (P02708)	Lamotrigine (DB00555)	DB00184* , DB00657* , DB01273 , DB00514 , DB00333
Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1C (Q14123)	Caffeine (DB00201)	DB01023 , DB01244 , DB00622 , DB01656 , DB00651
Drugs	Known Target	Predicted Target (Uniprot ID)
Nicorandil (DB09220)	ATP-binding cassette sub-family C member 9 (O60706)	Q09428* , P13569 , O15440 , Q92887 , P33527
Plerixafor (DB06809)	C-X-C chemokine receptor type 4 (P61073)	P35348 , P08913 , P35368 , P18825 , P18089

Notes.

*The drugs or targets that are found on publications are shown in bold numbers.

neurotransmitters, thereby preventing seizures ([Prabhavalkar, Poovanpallil & Bhatt, 2015](#); [Vallés, Garbus & Barrantes, 2007](#)). According to the DTI predictions, nAChR $\alpha 1$ could be associated with drugs that play several roles in voltage-dependent ion channels of nAChR $\alpha 1$, including nicotine ([DB00184](#)) and mecamlamine ([DB00657](#)). Nicotine is a stimulant drug that acts as an agonist for nicotinic acetylcholine receptors ([Wittenberg et al., 2020](#); [Xiao et al., 2020](#)). By binding this protein, nicotine activates voltage-gated calcium channels, causing the channel to open and allowing conductance of sodium, calcium, and potassium ([Wishart et al., 2018](#)). Nicotine is often used to relieve nicotine withdrawal symptoms and to aid in smoking cessation. Mecamlamine is a nicotine antagonist used to treat hypertension and uncomplicated malignant hypertension. By binding this protein, mecamlamine can act as a nicotinic acetylcholine receptor (nAChR) antagonist, inhibiting all known nAChR subtypes ([Nickell et al., 2013](#)). Many studies have shown that both drugs interact with nAChR $\alpha 1$, such as the studies of [Cosford et al. \(1996\)](#) and [Holladay, Dart & Lynch \(1997\)](#) for nicotine and the studies of [Cosford et al. \(1996\)](#) and [Kaushal & Tadi \(2020\)](#) for mecamlamine. Furthermore, nAChR $\alpha 1$ is predicted to involve varenicline ([DB01273](#)), dextromethorphan ([DB00514](#)), and methadone ([DB00333](#)), drugs for the relief of pain ([Eyler, 2013](#); [Hooten & Warner, 2015](#); [Weinbroum et al., 2000](#)) and treatment of addiction ([Ali et al., 2017](#); [Koyuncuoğlu & Saydam, 1990](#); [Nocente et al., 2013](#)). It has been revealed that these drugs interact with nAChR $\alpha 4$, nAChR $\alpha 7$, nAChR $\alpha 4$, and nACh $\beta 2$ ([Damaj et al., 2005](#); [Lee et al., 2006](#); [Mihalak, Carroll & Luetje, 2006](#); [Stensland et al., 2007](#); [Talka, Salminen & Tuominen, 2015](#); [Talka, Tuominen & Salminen, 2015](#)). Thus, varenicline, dextromethorphan, and methadone could bind to the highly similar protein nAChR $\alpha 1$.

Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1C or PDE1C ([Q14123](#)) is another protein involving high-score predicted drug-target links. Caffeine ([DB00201](#)), a stimulant present in tea, coffee, and analgesic drugs, is the only approved known drug that interacts to target PDE1C. Caffeine can increase alertness in a short period of time. Caffeine can bind PDE1C and cause vasodilation ([Echeverri et al., 2010](#)).

According to the DTI predictions, PDE1C could be associated with felodipine (DB01023), bepridil (DB01244), nicardipine (DB00622), and roflumilast (DB01656), which are drugs used in chemotherapy for cancer. Despite no clear evidence about these interactions, some publications support that these four drugs are involved in vasodilation, angina, and ischemic heart disease ([Bansal & Khandelwal, 2019](#); [DeWald et al., 2018](#); [Ekins, Puhl & Davidow, 2020](#); [Herych & Yatsyshyn, 2014](#)). In addition, PDE1C is predicted to interact with dyphylline (DB00651), a drug approved for asthma, bronchospasm, and chronic obstructive pulmonary disease (COPD) ([Ding et al., 2012](#); [Wishart et al., 2018](#); [Zhao et al., 2019](#)). Interestingly, this drug and caffeine are in a class of methylxanthines, a purine-derived group of pharmacologic agents used for bronchodilation and stimulation ([Basnet et al., 2017](#); [Gottwalt & Tadi, 2021](#)).

In the case of drugs having only one known targets, we selected nicorandil (DB09220) and plerixafor (DB06809) based on the high-score predicted drug-target links by the FSI model. Nicorandil is a vasodilatory drug used for patients with angina ([Ahmed, 2019](#)). It was found that the adenosine triphosphate (ATP)-binding cassette sub-family C member 9 or ABCC9 ([O60706](#)) interacted with nicorandil. This drug can activate vasodilation of arterioles and large coronary arteries ([Rang et al., 2011](#); [Wishart et al., 2018](#)). According to the DTI predictions, nicorandil could be associated with some target proteins in the ABCC subfamily, including ATP-binding cassette subfamily C member 8 or ABCC8 ([Q09428](#)), cystic fibrosis transmembrane conductance regulator or ABCC7 ([P13569](#)), ATP-binding cassette subfamily C member 5 or ABCC5 ([O15440](#)), ATP-binding cassette subfamily C member 2 or ABCC2 ([Q92887](#)), and ATP-binding cassette sub-family C member 1 or ABCC1 ([P33527](#)). Interestingly, it was found that nicorandil can reduce an excess of insulin secretion in ABCC8-deficient insulin-producing cells ([Guo et al., 2017](#)). Moreover, some known ABCC9 drugs, such as adenosine triphosphate (DB00171) and glyburide (DB01016), have a link with predicted ABCCs proteins ([Wishart et al., 2018](#)). Because these proteins share the same known drugs as ABCC9, and they are also in the same subfamily, which could be rather structurally conserved, it is possible that other proteins in the ABCC subfamily may bind to nicorandil as well ([Radivojac et al., 2013](#)).

Plerixafor is an anti-HIV agent specifically active against T4-lymphotropic HIV strains ([De Clercq, 2019](#)). Plerixafor has been used to mobilize stem cells by blocking the interaction between C – X – C motif chemokine 12 or CXCL12 ([P48061](#)) and C – X – C chemokine receptor type 4 or CXCR4 ([P61073](#)) ([Ilmer et al., 2020](#); [Innis-Shelton & Costa, 2019](#); [Wishart et al., 2018](#)). CXCR4 is the only target protein that interacts with plerixafor. According to the DTI predictions, plerixafor could be associated with some target proteins in a group of adrenoceptors, including alpha-1A adrenergic receptor or ADRA1A ([P35348](#)), alpha-2A adrenergic receptor or ADRA2A ([P08913](#)), alpha-1B adrenergic receptor or ADRA1B ([P35368](#)), alpha-2C adrenergic receptor or ADRA2C ([P18825](#)), and alpha-2B adrenergic receptor or ADRA2B ([P18089](#)). Despite no clear evidence of the associations between plerixafor and these proteins, some proteins in the adrenoceptor group have been reported to induce mobilization of stem cells and progenitor cells ([Agha et al., 2018](#); [Asada et al., 2013](#); [Baker et al., 2020](#); [Dar et al., 2011](#)).

DISCUSSION

To predict DTIs, the existing similarity-based methods (e.g., [Bleakley & Yamanishi, 2009](#); [Campillos et al., 2008](#)) mostly utilized the drug similarity only based on chemical structures and the target similarity only based on protein sequences. However, drug and target similarity measures solely based on a single feature of drugs and targets could introduce unreliable similarity scores and lead to inaccurate predictions of DTIs ([Chen et al., 2021](#)). In this study, we introduce a network propagation method with the FSI framework for systematically constructing an optimal model with suitable drug and target similarity integration. Through the FSI algorithm, we can select suitable drug and target similarity measures for similarity integration, the best integration method, and the best performance measure serving as the criteria for conducting the similarity integration.

As a result, the FSI model combines DDA_Jac, DDI_Cos, and Structures by using SNF as the integrated drug similarity and employs Sequences as the target–target similarity. By supplementing Structures with DDA_Jac and DDI_Cos, the FSI model perform significantly better than the OHN model, which utilizes only Structures and Sequences. This supports FSI efficiency in selecting and integrating beneficial information (i.e., DDA_Jac and DDI_Cos) for predicting DTIs. Because the therapeutic effects of drugs are associated with the ability to modulate drug targets at the molecular level ([Chen, Cheng & Li, 2020](#)), the similarity based on drug indications could suggest relationships between drugs and targets. DDAs occur when two drugs involve related molecular targets or processes, resulting in unintended effects ([Palleria et al., 2013](#)). The high similarity based on DDIs could convey highly similar targets or processes that are involved in drugs. Several recent methods include DDAs and DDIs for more effective predictions of DTIs, such as those by [Jiang et al. \(2022\)](#) and [Wang et al. \(2022\)](#).

In addition to the drug and target similarity measures that are integrated, the performance of a model in predicting DTIs also depends on the similarity integration methods applied in the FSI algorithm. We studied both linear (i.e., AVG, MAX, and MIN) and nonlinear (i.e., SNF) functions for similarity integration. With a particular performance measure used in the FSI algorithm, we usually obtain different selected drug and target similarities when using different similarity integration methods. However, the integration method with the best performance in predicting DTIs is the SNF method (i.e., Model 7 in [Table 2](#)). Linear integration is somewhat sensitive to outliers of some similarity scores. Rather than directly combining multiple similarity matrices, SNF is a network-based method that iteratively integrates multiple similarity networks into one composite network ([Wang et al., 2014](#)). Nowadays, SNF is an effective method widely exploited for aggregating multi-omics data in several biological applications, such as DTI prediction ([Shao et al., 2022](#); [Yan et al., 2017](#)) and DDA inference ([Jarada, Rokne & Alhajj, 2021a](#); [Jarada, Rokne & Alhajj, 2021b](#)).

In the FSI framework, the performance measure used for selecting drug and target similarity fusions is also important. We studied the most commonly used comprehensive metrics, including AUC, AUPR, and F1. According to the results ([Table 1](#)), using AUPR and F1 in FSI mostly obtains a similar selection of drug and target similarity measures.

In addition, the models obtained using AUPR and F1 performed significantly better than those obtained using AUC (Table 2). This is because the number of negative (unknown) DTIs greatly outnumbers the number of positive (known) DTIs in the dataset, resulting in the AUC values being deceptive and unsuitable for evaluating model performance in this situation (Saito & Rehmsmeier, 2015).

In this work, the heterogeneous network model was improved by plugging in the feature selection and integration with our proposed FSI algorithm. The algorithm searched for a suitable integration method of various similarity measures and resulted in a set of similarities for the network propagation that yields better performance than the original one. The random forest which was performed with the same set of the features obtained by the FSI algorithm also yield better performance. In principle, the direct comparison of the heterogeneous network model to the other state-of-the-art methods was not applicable since many current methods for drug-target predictions use various aspects of the features (not only similarities) to build up a prediction model and concern complicated tasks. Therefore, the prediction inference by the heterogeneous network model is simple and benefit for many applications and research studies. The development for improving this technique is still required and getting more interest as an essential tool now.

CONCLUSIONS

In this study, we propose a network propagation method with the FSI framework, which systemically integrates multiple similarity measures, for predicting new links between drugs and targets. Based on our datasets, seven drug–drug similarity matrices and nine target–target similarity matrices can be constructed. Four different similarity integration methods and three comprehensive metrics were tested. Consequently, FSI selects to combine DDA_Jac, DDI_Cos, and Structures as a drug similarity by using AUPR and SNF. Additionally, the FSI model exploits Seq_Loc as a target similarity. The FSI model performed significantly better than the OHN model and the models with full and random similarity integration. This demonstrates the efficiency of the FSI framework in predicting DTIs. Nevertheless, the FSI framework is not limited to DTI prediction with our drug and target similarity measures. FSI can be generally applied for drug–target identification with different existing data sets, integration methods, and performance measures or even other applications requiring integration of multiple similarity measures.

ACKNOWLEDGEMENTS

We acknowledge National e-Science Infrastructure Consortium (<http://www.e-science.in.th>) for providing computing resources that have contributed to the research results reported within this paper.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was funded by the National Science, Research and Innovation Fund (NSRF), and King Mongkut's University of Technology North Bangkok with Contract no. KMUTNB-FF-65-55. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

National Science, Research and Innovation Fund (NSRF).

King Mongkut's University of Technology North Bangkok: KMUTNB-FF-65-55.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Piyanut Tangmanussukum conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Thitipong Kawichai conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Apichat Suratanee conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Kitiporn Plaimas conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data and code are available at GitHub: https://github.com/piyanuttnk/FSI_framework.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.1124#supplemental-information>.

REFERENCES

- Agha NH, Baker FL, Kunz HE, Graff R, Azadan R, Dolan C, Laughlin MS, Hosing C, Markofski MM, Bond RA. 2018. Vigorous exercise mobilizes CD34+ hematopoietic stem cells to peripheral blood via the β 2-adrenergic receptor. *Brain, Behavior, and Immunity* 68:66–75 DOI 10.1016/j.bbi.2017.10.001.

- Ahmed LA. 2019.** Nicorandil: a drug with ongoing benefits and different mechanisms in various diseased conditions. *Indian Journal of Pharmacology* **51**(5):296–301 DOI [10.4103/ijp.IJP_298_19](https://doi.org/10.4103/ijp.IJP_298_19).
- Al-Karmalawy AA, Dahab MA, Metwaly AM, Elhady SS, Elkaeed EB, Eissa IH, Darwish KM. 2021.** Molecular docking and dynamics simulation revealed the potential inhibitory activity of ACEIs against SARS-CoV-2 targeting the h ACE2 receptor. *Frontiers in Chemistry* **9**:661230 DOI [10.3389/fchem.2021.661230](https://doi.org/10.3389/fchem.2021.661230).
- Ali S, Tahir B, Jabeen S, Malik M. 2017.** Methadone treatment of opiate addiction: a systematic review of comparative studies. *Innovations in clinical Neuroscience* **14**:8–19.
- Alonso H, Bliznyuk AA, Greedy JE. 2006.** Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews* **26**:531–568 DOI [10.1002/med.20067](https://doi.org/10.1002/med.20067).
- An Q, Yu L. 2021.** A heterogeneous network embedding framework for predicting similarity-based drug-target interactions. *Briefings in Bioinformatics* **22**:bbab275 DOI [10.1093/bib/bbab275](https://doi.org/10.1093/bib/bbab275).
- Aparicio LMA, Pulido EG, Gallego GA. 2012.** Vinflunine: a new vision that may translate into antiangiogenic and antimetastatic activity. *Anti-Cancer Drugs* **23**:1–11 DOI [10.1097/CAD.0b013e32834d237b](https://doi.org/10.1097/CAD.0b013e32834d237b).
- Arora RD, Menezes RG. 2021.** Vinca alkaloid toxicity. Treasure Island: StatPearls. Available at <https://www.ncbi.nlm.nih.gov/books/NBK557842/>.
- Asada N, Katayama Y, Sato M, Minagawa K, Wakahashi K, Kawano H, Kawano Y, Sada A, Ikeda K, Matsui T. 2013.** Matrix-embedded osteocytes regulate mobilization of hematopoietic stem/progenitor cells. *Cell Stem Cell* **12**(6):737–747 DOI [10.1016/j.stem.2013.05.001](https://doi.org/10.1016/j.stem.2013.05.001).
- Baker FL, Bigley AB, Agha NH, Pedlar CR, O'Connor DP, Bond RA, Bollard CM, Katsanis E, Simpson RJ. 2020.** Systemic β -adrenergic receptor activation augments the ex vivo expansion and anti-tumor activity of V γ 9V δ 2 T-cells. *Frontiers in Immunology* **10**:3082 DOI [10.3389/fimmu.2019.03082](https://doi.org/10.3389/fimmu.2019.03082).
- Bansal AB, Khandelwal G. 2019.** Felodipine. Treasure Island: StatPearls. Available at <https://www.ncbi.nlm.nih.gov/books/NBK542163/>.
- Basnet RM, Zizioli D, Guarienti M, Finazzi D, Memo M. 2017.** Methylxanthines induce structural and functional alterations of the cardiac system in zebrafish embryos. *BMC Pharmacology and Toxicology* **18**(1):72 DOI [10.1186/s40360-017-0179-9](https://doi.org/10.1186/s40360-017-0179-9).
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000.** The protein data bank. *Nucleic Acids Research* **28**(1):235–242 DOI [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235).
- Bhargava H, Sharma A, Suravajhala P. 2021.** Chemogenomic approaches for revealing drug target interactions in drug discovery. *Current Genomics* **22**(5):328–338 DOI [10.2174/1389202922666210920125800](https://doi.org/10.2174/1389202922666210920125800).
- Bleakley K, Yamanishi Y. 2009.** Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics* **25**(18):2397–2403 DOI [10.1093/bioinformatics/btp433](https://doi.org/10.1093/bioinformatics/btp433).

- Butina D, Segall MD, Frankcombe K. 2002.** Predicting ADME properties in silico: methods and models. *Drug Discovery Today* 7(11):S83–S88
DOI 10.1016/s1359-6446(02)02288-2.
- Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P. 2008.** Drug target identification using side-effect similarity. *Science* 321(5886):263–266
DOI 10.1126/science.1158140.
- Chen C, Shi H, Jiang Z, Salhi A, Chen R, Cui X, Yu B. 2021.** DNN-DTIs: improved drug-target interactions prediction using XGBoost feature selection and deep neural network. *Computers in Biology and Medicine* 136:104676
DOI 10.1016/j.combiomed.2021.104676.
- Chen H, Cheng F, Li J. 2020.** iDrug: integration of drug repositioning and drug-target prediction via cross-network embedding. *PLOS computational biology* 16(7):e1008040 DOI 10.1371/journal.pcbi.1008040.
- Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, Salzberg AC, Huang ES. 2007.** Structure-based maximal affinity model predicts small-molecule druggability. *Nature Biotechnology* 25(1):71–75 DOI 10.1038/nbt1273.
- Cheng F, Li W, Wu Z, Wang X, Zhang C, Li J, Liu G, Tang Y. 2013.** Prediction of polypharmacological profiles of drugs by the integration of chemical, side effect, and therapeutic space. *Journal of Chemical Information and Modeling* 53(4):753–762
DOI 10.1021/ci400010.
- Chu SW, Badar S, Morris DL, Pourgholami MH. 2009.** Potent inhibition of tubulin polymerisation and proliferation of paclitaxel-resistant 1A9PTX22 human ovarian cancer cells by albendazole. *Anticancer Research* 29(10):3791–3796.
- Cosford ND, Bleicher L, Herbaut A, McCallum JS, Vernier J-M, Dawson H, Whitten JP, Adams P, Chavez-Noriega L, Correa LD. 1996.** (S)-(–)-5-Ethynyl-3-(1-methyl-2-pyrroli-dinyl) pyridine Maleate (SIB-1508Y): a novel anti-parkinsonian agent with selectivity for neuronal nicotinic acetylcholine receptors. *Journal of Medicinal Chemistry* 39(17):3235–3237 DOI 10.1021/jm960328w.
- Damaj MI, Flood P, Ho K, May EL, Martin BR. 2005.** Effect of dextrometorphan and dextrorphan on nicotine and neuronal nicotinic receptors: in vitro and in vivo selectivity. *Journal of Pharmacology and Experimental Therapeutics* 312(2):780–785
DOI 10.1124/jpet.104.075093.
- Dar A, Schajnovitz A, Lapid K, Kalinkovich A, Itkin T, Ludin A, Kao W-M, Battista M, Tesio M, Kollet O. 2011.** Rapid mobilization of hematopoietic progenitors by AMD3100 and catecholamines is mediated by CXCR4-dependent SDF-1 release from bone marrow stromal cells. *Leukemia* 25(8):1286–1296
DOI 10.1038/leu.2011.62.
- Davis AP, Grondin CJ, Johnson RJ, Sciaky D, Wieggers J, Wieggers TC, Mattingly CJ. 2021.** Comparative toxicogenomics database (CTD): update 2021. *Nucleic Acids Research* 49:D1138–D1143 DOI 10.1093/nar/gkaa891.
- De Clercq E. 2019.** Mozobil® (Plerixafor, AMD3100), 10 years after its approval by the US Food and Drug Administration. *Antiviral Chemistry and Chemotherapy* 27:2040206619829382 DOI 10.1177/2040206619829382.

- DeWald LE, Dyall J, Sword JM, Torzewski L, Zhou H, Postnikova E, Kollins E, Alexander I, Gross R, Cong Y. 2018. The calcium channel blocker bepridil demonstrates efficacy in the murine model of marburg virus disease. *The Journal of Infectious Diseases* 218(suppl 5):S588–S591 DOI 10.1093/infdis/jiy332.
- Ding R, Shi J, Pabon K, Scotto KW. 2012. Xanthines down-regulate the drug transporter ABCG2 and reverse multidrug resistance. *Molecular Pharmacology* 81(3):328–337 DOI 10.1124/mol.111.075556.
- Echeverri D, Montes FR, Cabrera M, Galán A, Prieto A. 2010. Caffeine’s vascular mechanisms of action. *International Journal of Vascular Medicine* 2010:834060 DOI 10.1155/2010/834060.
- Ekins S, Puhl AC, Davidow A. 2020. Repurposing the dihydropyridine calcium channel inhibitor nicardipine as a Nav1, 8 inhibitor in vivo for Pitt Hopkins syndrome. *Pharmaceutical Research* 37(7):127 DOI 10.1007/s11095-020-02853-5.
- Eswaran D, Günnemann S, Faloutsos C, Makhija D, Kumar M. 2017. Zoobp: belief propagation for heterogeneous networks. *Proceedings of the VLDB Endowment* 10(5):625–636 DOI 10.14778/3055540.3055554.
- Eyler EC. 2013. Chronic and acute pain and pain management for patients in methadone maintenance treatment. *The American Journal on Addictions* 22(1):75–83 DOI 10.1111/j.1521-0391.2013.00308.x.
- Fiscon G, Paci P. 2021. SAveRUNNER: an R-based tool for drug repurposing. *BMC Bioinformatics* 22(1):150 DOI 10.1186/s12859-021-04076-w.
- Gottlieb A, Stein GY, Ruppin E, Sharan R. 2011. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Molecular Systems Biology* 7(1):496 DOI 10.1038/msb.2011.26.
- Gottwalt B, Tadi P. 2021. *Methylxanthines*. Treasure Island, Florida, USA: StatPearls Publishing. Available at <https://www.ncbi.nlm.nih.gov/books/NBK559165/>.
- Guo D, Liu H, Ruzi A, Gao G, Nasir A, Liu Y, Yang F, Wu F, Xu G, Y-x Li. 2017. Modeling congenital hyperinsulinism with ABCC8-deficient human embryonic stem cells generated by CRISPR/Cas9. *Scientific Reports* 7(1):3156 DOI 10.1038/s41598-017-03349-w.
- He J, Yang X, Gong Z. 2020. Hybrid attentional memory network for computational drug repositioning. *BMC Bioinformatics* 21:566 DOI 10.1186/s12859-020-03898-4.
- Hengphasatporn K, Plaimas K, Suratane A, Wongsripisant P, Yang JM, Shigeta Y, Chavasiri W, Boonyasuppayakorn S, Rungrotmongkol T. 2020. Target identification using homopharma and network-based methods for predicting compounds against dengue virus-infected cells. *Molecules* 25(8):1883 DOI 10.3390/molecules25081883.
- Herych P, Yatsyshyn R. 2014. Optimizing treatment of patients with co-existing cardiorespiratory pathology by administration of anti-inflammatory roflumilast and cardioprotective agent quercetin. *Galician Medical Journal* 21:65–69 DOI 10.1038/npre.2009.3154.1.

- Holladay MW, Dart MJ, Lynch JK. 1997.** Neuronal nicotinic acetylcholine receptors as targets for drug discovery. *Journal of Medicinal Chemistry* **40**(26):4169–4194 DOI [10.1021/jm970377o](https://doi.org/10.1021/jm970377o).
- Hooten WM, Warner DO. 2015.** Varenicline for opioid withdrawal in patients with chronic pain: a randomized, single-blinded, placebo controlled pilot trial. *Addictive Behaviors* **42**:69–72 DOI [10.1016/j.addbeh.2014.11.007](https://doi.org/10.1016/j.addbeh.2014.11.007).
- Huntley R, Dimmer E, Barrell D, Binns D, Apweiler R. 2009.** The gene ontology annotation (goa) database. In: *Nature Precedings*. 1–1 DOI [10.1038/npre.2009.3154.1](https://doi.org/10.1038/npre.2009.3154.1).
- Imer M, Weniger M, Niess H, Wu Y, Zhang C, Westphalen CB, Kruger S, Angele MK, Werner J, D’Haese JG. 2020.** Stories of drug repurposing for pancreatic cancer treatment—past, present, and future. In: *Drug repurposing in cancer therapy*. Cambridge, Massachusetts: Academic Press, 231–272 DOI [10.1016/b978-0-12-819668-7.00009-9](https://doi.org/10.1016/b978-0-12-819668-7.00009-9).
- Innis-Shelton R, Costa LJ. 2019.** Sources of cells for hematopoietic cell transplantation: practical aspects of hematopoietic cell collection. In: *Hematopoietic cell transplantation for malignant conditions*. Amsterdam: Elsevier, 73–84.
- Jaccard P. 1912.** The distribution of the flora in the alpine zone. 1. *New Phytologist* **11**(2):37–50 DOI [10.1111/j.1469-8137.1912.tb05611.x](https://doi.org/10.1111/j.1469-8137.1912.tb05611.x).
- Janyasupab P, Suratane A, Plaimas K. 2021.** Network diffusion with centrality measures to identify disease-related genes. *Mathematical Biosciences and Engineering* **18**(3):2909–2929 DOI [10.3934/mbe.2021147](https://doi.org/10.3934/mbe.2021147).
- Jarada TN, Rokne JG, Alhajj R. 2021a.** SNF-NN: computational method to predict drug-disease interactions using similarity network fusion and neural networks. *BMC Bioinformatics* **22**(1):28 DOI [10.1186/s12859-020-03950-3](https://doi.org/10.1186/s12859-020-03950-3).
- Jarada TN, Rokne JG, Alhajj R. 2021b.** SNF-CVAE: computational method to predict drug-disease interactions using similarity network fusion and collective variational autoencoder. *Knowledge-based Systems* **212**:106585 DOI [10.1016/j.knosys.2020.106585](https://doi.org/10.1016/j.knosys.2020.106585).
- Jiang L, Sun J, Wang Y, Ning Q, Luo N, Yin M. 2022.** Identifying drug-target interactions via heterogeneous graph attention networks combined with cross-modal similarities. *Briefings in Bioinformatics* **23**(2):bbac016 DOI [10.1093/bib/bbac016](https://doi.org/10.1093/bib/bbac016).
- Kanehisa M, Goto S. 2000.** KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**(1):27–30 DOI [10.1093/nar/28.1.27](https://doi.org/10.1093/nar/28.1.27).
- Kaushal S, Tadi P. 2020.** *Nicotinic ganglionic blocker*. Treasure Island: StatPearls. Available at <https://www.ncbi.nlm.nih.gov/books/NBK556073/>.
- Kawichai T, Suratane A, Plaimas K. 2021.** Meta-path based gene ontology profiles for predicting drug-disease associations. *IEEE Access* **9**:41809–41820 DOI [10.1109/ACCESS.2021.3065280](https://doi.org/10.1109/ACCESS.2021.3065280).
- Kim E, A-s Choi, Nam H. 2019.** Drug repositioning of herbal compounds via a machine-learning approach. *BMC Bioinformatics* **20**(Supp 10):247 DOI [10.1186/s12859-019-2811-8](https://doi.org/10.1186/s12859-019-2811-8).

- Koyuncuoğlu H, Saydam B. 1990. The treatment of heroin addicts with dextromethorphan: a double-blind comparison of dextromethorphan with chlorpromazine. *International Journal of Clinical Pharmacology, Therapy, and Toxicology* **28**(4):147–152.
- Kuhn M, Letunic I, Jensen LJ, Bork P. 2016. The SIDER database of drugs and side effects. *Nucleic Acids Research* **44**(D1):D1075–D1079 DOI [10.1093/nar/gkv1075](https://doi.org/10.1093/nar/gkv1075).
- Lee J-H, Shin E-J, Jeong SM, Kim J-H, Lee B-H, Yoon I-S, Lee J-H, Choi S-H, Lee S-M, Lee PH. 2006. Effects of dextrorotatory morphinans on $\alpha 3\beta 4$ nicotinic acetylcholine receptors expressed in *Xenopus* oocytes. *European Journal of Pharmacology* **536**(1-2):85–92 DOI [10.1016/j.ejphar.2006.02.034](https://doi.org/10.1016/j.ejphar.2006.02.034).
- Liu B, Pliakos K, Vens C, Tsoumakas G. 2022. Drug-target interaction prediction via an ensemble of weighted nearest neighbors with interaction recovery. *Applied Intelligence* **52**:3705–3727 DOI [10.1007/s10489-021-02495-z](https://doi.org/10.1007/s10489-021-02495-z).
- Liu Y, Wu M, Miao C, Zhao P, Li X-L. 2016. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. *PLOS Computational Biology* **12**(2):e1004760 DOI [10.1371/journal.pcbi.1004760](https://doi.org/10.1371/journal.pcbi.1004760).
- Lobert S, Ingram JW, Hill BT, Correia JJ. 1998. A comparison of thermodynamic parameters for vinorelbine- and vinflunine-induced tubulin self-association by sedimentation velocity. *Molecular Pharmacology* **53**(5):908–915.
- Lobert S, Vulevic B, Correia JJ. 1996. Interaction of vinca alkaloids with tubulin: a comparison of vinblastine, vincristine, and vinorelbine. *Biochemistry* **35**(21):6806–6814 DOI [10.1021/bi953037i](https://doi.org/10.1021/bi953037i).
- Luo H, Wang J, Li M, Luo J, Peng X, Wu F-X, Pan Y. 2016. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics* **32**(17):2664–2671 DOI [10.1093/bioinformatics/btw228](https://doi.org/10.1093/bioinformatics/btw228).
- Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, Peng J, Chen L, Zeng J. 2017. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature Communications* **8**(1):573 DOI [10.1038/s41467-017-00680-8](https://doi.org/10.1038/s41467-017-00680-8).
- Mihalak KB, Carroll FI, Luetje CW. 2006. Varenicline is a partial agonist at $\alpha 4\beta 2$ and a full agonist at $\alpha 7$ neuronal nicotinic receptors. *Molecular Pharmacology* **70**(3):801–805 DOI [10.1124/mol.106.025130](https://doi.org/10.1124/mol.106.025130).
- Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**(3):443–453 DOI [10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4).
- Nickell JR, Grinevich VP, Siripurapu KB, Smith AM, Dwoskin LP. 2013. Potential therapeutic uses of mecamylamine and its stereoisomers. *Pharmacology Biochemistry and Behavior* **108**:28–43 DOI [10.1016/j.pbb.2013.04.005](https://doi.org/10.1016/j.pbb.2013.04.005).
- Nishimura N, Reien Y, Matsumoto A, Ogura T, Miyata Y, Suzuki K, Nakazato Y, Daida H, Nakaya H. 2010. Effects of nicorandil on the cAMP-dependent Cl[−] current in guinea-pig ventricular cells. *Journal of Pharmacological Sciences* **112**(4):415–423 DOI [10.1254/jphs.09237fp](https://doi.org/10.1254/jphs.09237fp).
- Nocente R, Vitali M, Balducci G, Enea D, Kranzler HR, Ceccanti M. 2013. Varenicline and neuronal nicotinic acetylcholine receptors: a new approach to the treatment of

- co-occurring alcohol and nicotine addiction? *The American Journal on Addictions* 22(5):453–459 DOI 10.1111/j.1521-0391.2013.12037.x.
- Öztürk H, Ozkirimli E, Özgür A. 2016. A comparative study of SMILES-based compound similarity functions for drug-target interaction prediction. *BMC Bioinformatics* 17:128 DOI 10.1186/s12859-016-0977-x.
- Palleria C, Di Paolo A, Giofrè C, Caglioti C, Leuzzi G, Siniscalchi A, De Sarro G, Gallelli L. 2013. Pharmacokinetic drug-drug interaction and their implication in clinical management. *Journal of Research in Medical Sciences: the Official Journal of Isfahan University of Medical Sciences* 18(7):601–610.
- Peng J, Li J, Shang X. 2020. A learning-based method for drug-target interaction prediction based on feature representation learning and deep neural network. *BMC Bioinformatics* 21(Supp 13):394 DOI 10.1186/s12859-020-03677-1.
- Peng Y, Wang J, Wu Z, Zheng L, Wang B, Liu G, Li W, Tang Y. 2022. MPSM-DTI: prediction of drug–target interaction via machine learning based on the chemical structure and protein sequence. *Digital Discovery* 1(2):115–126 DOI 10.1039/D1DD00011J.
- Perlman L, Gottlieb A, Atias N, Ruppín E, Sharan R. 2011. Combining drug and gene similarity measures for drug-target elucidation. *Journal of Computational Biology* 18(2):133–145 DOI 10.1089/cmb.2010.0213.
- Pinzi L, Rastelli G. 2019. Molecular docking: shifting paradigms in drug discovery. *International Journal of Molecular Sciences* 20(18):4331 DOI 10.3390/ijms20184331.
- Prabhavalkar KS, Poovanpallil NB, Bhatt LK. 2015. Management of bipolar depression with lamotrigine: an antiepileptic mood stabilizer. *Frontiers in Pharmacology* 6:242 DOI 10.3389/fphar.2015.00242.
- Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, Graim K, Funk C, Verspoor K, Ben-Hur A. 2013. A large-scale evaluation of computational protein function prediction. *Nature Methods* 10(3):221–227 DOI 10.1038/nmeth.2340.
- Ramírez T, Benítez-Bribiesca L, Ostrosky-Wegman P, Herrera LA. 2001. In vitro effects of albendazole and its metabolites on the cell proliferation kinetics and micronuclei frequency of stimulated human lymphocytes. *Archives of Medical Research* 32(2):119–122 DOI 10.1016/s0188-4409(01)00259-4.
- Rang HP, Dale MM, Ritter JM, Flower RJ, Henderson G. 2011. *Rang & Dale's pharmacology*. 7th Edition. Amsterdam, Netherlands: Elsevier Health Sciences, 286.
- Rayhan F, Ahmed S, Mousavian Z, Farid DM, Shatabda S. 2020. FRnet-DTI: deep convolutional neural network for drug-target interaction prediction. *Heliyon* 6(3):e03444 DOI 10.1016/j.heliyon.2020.e03444.
- Rohani N, Eslahchi C. 2019. Drug-drug interaction predicting by neural network using integrated similarity. *Scientific Reports* 9(1):13645 DOI 10.1038/s41598-019-50121-3.
- Rudrapal M, Khairnar SJ, Jadhav AG. 2020. Drug repurposing (DR): an emerging approach in drug discovery. In: *Drug repurposing-hypothesis, molecular aspects and*

- therapeutic applications*. London, United Kingdom: IntechOpen [Online] Available at <https://www.intechopen.com/chapters/72744>.
- Saito T, Rehmsmeier M. 2015.** The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE* **10**(3):e0118432 DOI [10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432).
- Schenone M, Dančik V, Wagner BK, Clemons PA. 2013.** Target identification and mechanism of action in chemical biology and drug discovery. *Nature Chemical Biology* **9**:232–240 DOI [10.1038/nchembio.1199](https://doi.org/10.1038/nchembio.1199).
- Shaikh F, Tai HK, Desai N, Siu SW. 2021.** LigTMap: ligand and structure-based target identification and activity prediction for small molecular compounds. *Journal of Cheminformatics* **13**:44 DOI [10.1186/s13321-021-00523-1](https://doi.org/10.1186/s13321-021-00523-1).
- Shao K, Zhang Y, Wen Y, Zhang Z, He S, Bo X. 2022.** DTI-HETA: prediction of drug–target interactions based on GCN and GAT on heterogeneous graph. *Briefings in Bioinformatics* **23**(3):bbac109 DOI [10.1093/bib/bbac109](https://doi.org/10.1093/bib/bbac109).
- Singhal A. 2001.** Modern information retrieval: a brief overview. *IEEE Data Engineering Bulletin Issues* **24**(4):35–43.
- Smith TF, Waterman MS. 1981.** Identification of common molecular subsequences. *Journal of Molecular Biology* **147**(1):195–197 DOI [10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5).
- Solana H, Sallovitz J, Lanusse C, Rodriguez J. 2002.** Enantioselective binding of albendazole sulphoxide to cytosolic proteins from helminth parasites. *Methods and Findings in Experimental and Clinical Pharmacology* **24**(1):7–14 DOI [10.1358/mf.2002.24.1.677121](https://doi.org/10.1358/mf.2002.24.1.677121).
- Steensland P, Simms JA, Holgate J, Richards JK, Bartlett SE. 2007.** Varenicline, an $\alpha 4 \beta 2$ nicotinic acetylcholine receptor partial agonist, selectively decreases ethanol consumption and seeking. *Proceedings of the National Academy of Sciences of the United States of America* **104**(30):12518–12523 DOI [10.1073/pnas.0705368104](https://doi.org/10.1073/pnas.0705368104).
- Suratane A, Buaboocha T, Plaimas K. 2021.** Prediction of human-plasmodium vivax protein associations from heterogeneous network structures based on machine-learning approach. *Bioinformatics and Biology Insights* **15**:11779322211013350 DOI [10.1177/11779322211013350](https://doi.org/10.1177/11779322211013350).
- Suratane A, Chokrathok C, Chutimanukul P, Khrueasan N, Buaboocha T, Chadchawan S, Plaimas K. 2018.** Two-state co-expression network analysis to identify genes related to salt tolerance in Thai rice. *Genes* **9**(12):594 DOI [10.3390/genes9120594](https://doi.org/10.3390/genes9120594).
- Suratane A, Plaimas K. 2014.** Identification of inflammatory bowel disease-related proteins using a reverse k-nearest neighbor search. *Journal of Bioinformatics and Computational Biology* **12**(4):1450017 DOI [10.1142/S0219720014500176](https://doi.org/10.1142/S0219720014500176).
- Suratane A, Plaimas K. 2015.** DDA: a novel network-based scoring method to identify disease-disease associations. *Bioinformatics and Biology Insights* **9**:175–186 DOI [10.4137/BBI.S35237](https://doi.org/10.4137/BBI.S35237).
- Suratane A, Plaimas K. 2017.** Reverse nearest neighbor search on a protein-protein interaction network to infer protein-disease associations. *Bioinformatics and Biology Insights* **11**:1177932217720405 DOI [10.1177/1177932217720405](https://doi.org/10.1177/1177932217720405).

- Suratane A, Plaimas K. 2018. Network-based association analysis to infer new disease-gene relationships using large-scale protein interactions. *PLOS ONE* 13(6):e0199435 DOI 10.1371/journal.pone.0199435.
- Suratane A, Plaimas K. 2020. Heterogeneous network model to identify potential associations between plasmodium vivax and human proteins. *International Journal of Molecular Sciences* 21(4):1310 DOI 10.3390/ijms21041310.
- Suratane A, Plaimas K. 2021. Hybrid deep learning based on a heterogeneous network profile for functional annotations of plasmodium falciparum genes. *International Journal of Molecular Sciences* 22(18):10019 DOI 10.3390/ijms221810019.
- Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P. 2021. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* 49(D1):D605–D612 DOI 10.1093/nar/gkaa1074.
- Talka R, Salminen O, Tuominen RK. 2015. Methadone is a non-competitive antagonist at the $\alpha 4\beta 2$ and $\alpha 3^*$ nicotinic acetylcholine receptors and an agonist at the $\alpha 7$ nicotinic acetylcholine receptor. *Basic & Clinical Pharmacology & Toxicology* 116(4):321–328 DOI 10.1111/bcpt.12317.
- Talka R, Tuominen RK, Salminen O. 2015. Methadone’s effect on nAChRs—a link between methadone use and smoking? *Biochemical Pharmacology* 97(4):542–549 DOI 10.1016/j.bcp.2015.07.031.
- Tanimoto TT. 1958. Elementary mathematical theory of classification and prediction. Internal IBM Technical Report.
- Thafar MA, Olayan RS, Ashoor H, Albaradei S, Bajic VB, Gao X, Gojobori T, Essack M. 2020. DTiGEMS+: drug-target interaction prediction using graph embedding, graph mining, and similarity-based techniques. *Journal of Cheminformatics* 12(1):44 DOI 10.1186/s13321-020-00447-2.
- Vallés AS, Garbus I, Barrantes FJ. 2007. Lamotrigine is an open-channel blocker of the nicotinic acetylcholine receptor. *Neuroreport* 18(1):45–50 DOI 10.1097/01.wnr.0000246323.66438.94.
- Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A. 2014. Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods* 11:333–337 DOI 10.1038/nmeth.2810.
- Wang G, Wang H, Guo F, Du M, Cao C. 2022. A novel method for drug-target interaction prediction based on graph transformers model. Available at https://assets.researchsquare.com/files/rs-1529816/v1_covered.pdf?c=1649688072.
- Wang W, Yang S, Li J. 2013a. Drug target predictions based on heterogeneous graph inference. In: *Biocomputing 2013: world scientific*. 53–64.
- Wang W, Yang S, Li J. 2013b. Drug target predictions based on heterogeneous graph inference. In: *Pacific Symposium on Biocomputing*. 53–64.
- Weinbroum AA, Rudick V, Paret G, Ben-Abraham R. 2000. The role of dextromethorphan in pain control. *Canadian Journal of Anesthesia* 47(6):585–596 DOI 10.1007/BF03018952.

- Weininger D. 1988.** SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* **28**(1):31–36 DOI [10.1021/ci00057a005](https://doi.org/10.1021/ci00057a005).
- Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z. 2018.** DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research* **46**(D1):D1074–D1082 DOI [10.1093/nar/gkx1037Z](https://doi.org/10.1093/nar/gkx1037Z).
- Wittenberg RE, Wolfman SL, Biasi MDe, Dani JA. 2020.** Nicotinic acetylcholine receptors and nicotine addiction: a brief introduction. *Neuropharmacology* **177**:108256 DOI [10.1016/j.neuropharm.2020.108256](https://doi.org/10.1016/j.neuropharm.2020.108256).
- Xiao C, Zhou C-Y, Jiang J-H, Yin C. 2020.** Neural circuits and nicotinic acetylcholine receptors mediate the cholinergic regulation of midbrain dopaminergic neurons and nicotine dependence. *Acta Pharmacologica Sinica* **41**(1):1–9 DOI [10.1038/s41401-019-0299-4](https://doi.org/10.1038/s41401-019-0299-4).
- Yan C, Wang J, Lan W, Wu F-X, Pan Y. 2017.** Sdtrls: predicting drug-target interactions for complex diseases based on chemical substructures. *Complexity* **2017**:2713280 DOI [10.1155/2017/2713280](https://doi.org/10.1155/2017/2713280).
- Yang C, Zhang J, Han J. 2019.** Neural embedding propagation on heterogeneous networks. In: *2019 IEEE International Conference on Data Mining (ICDM)*. 698–707 DOI [10.1109/ICDM.2019.00080](https://doi.org/10.1109/ICDM.2019.00080).
- Yang K, Liu G, Wang N, Zhang R, Yu J, Chen J, Zhou X. 2018.** Heterogeneous network propagation for herb target identification. *BMC Medical Informatics and Decision Making* **18**(Supp 1):17 DOI [10.1186/s12911-018-0592-z](https://doi.org/10.1186/s12911-018-0592-z).
- Yang M, Luo H, Li Y, Wang J. 2019.** Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* **35**(14):i455–i463 DOI [10.1093/bioinformatics/btz331](https://doi.org/10.1093/bioinformatics/btz331).
- Yang Y, Zhu Z, Wang X, Zhang X, Mu K, Shi Y, Peng C, Xu Z, Zhu W. 2021.** Ligand-based approach for predicting drug targets and for virtual screening against COVID-19. *Briefings in Bioinformatics* **22**(2):1053–1064 DOI [10.1093/bib/bbaa422](https://doi.org/10.1093/bib/bbaa422).
- Yu G. 2020.** Gene ontology semantic similarity analysis using GOSemSim. *Stem Cell Transcriptional Networks: Springer* **2117**:207–215 DOI [10.1007/978-1-0716-0301-7_11](https://doi.org/10.1007/978-1-0716-0301-7_11).
- Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. 2010.** GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**(7):976–978 DOI [10.1093/bioinformatics/btq064](https://doi.org/10.1093/bioinformatics/btq064).
- Yu J, Finley Jr RL. 2009.** Combining multiple positive training sets to generate confidence scores for protein–protein interactions. *Bioinformatics* **25**(1):105–111 DOI [10.1093/bioinformatics/btn597](https://doi.org/10.1093/bioinformatics/btn597).
- Zhang X, Lin X, Hu J, Ding W. 2022.** Unsupervised prediction method for drug-target interactions based on structural similarity. *International Conference on Intelligent Computing. Springer, Cham* **13394**:517–532 DOI [10.1007/978-3-031-13829-4_45](https://doi.org/10.1007/978-3-031-13829-4_45).
- Zhao HY, Ren YH, Ren XB, Wang Y. 2019.** Diprophylline inhibits non-small cell lung cancer A549 cell proliferation and migration, and promotes apoptosis,

by downregulating PI3K signaling pathway. *Oncology Letters* **17**(1):857–862

[DOI 10.3892/ol.2018.9678](https://doi.org/10.3892/ol.2018.9678).

Zhao P, Lin L, Wu M, Wang L, Geng Q, Li L, Zhao N, Shi J, Lu C. 2022. Targets preliminary screening for the fresh natural drug molecule based on Cosine-correlation and similarity-comparison of local network. *Journal of Translational Medicine* **20**(1):67

[DOI 10.1186/s12967-022-03279-w](https://doi.org/10.1186/s12967-022-03279-w).