Review

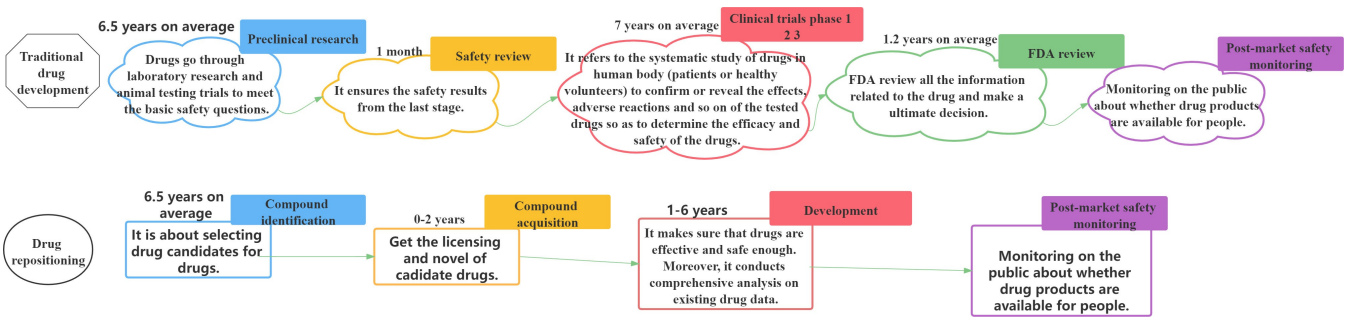
www.molinf.com

molecular informatics

doi.org/10.1002/minf.202100200

Network-Based Approaches for Drug Repositioning

Tao Song,[a, c] Gan Wang,[a] Mao Ding,*[b] Alfonso Rodriguez-Paton,[c] Xun Wang,*[a, d] and Shudong Wang[a]



Abstract: With deep learning creeping up into the ranks of big data, new models based on deep learning and massive data have made great leaps forward rapidly in the field of drug repositioning. However, there is no relevant review to summarize the transformations and development process of models and their data in the field of drug repositioning. Among all the computational methods, network-based methods play an extraordinary role. In view of these circumstances, understanding and comparing existing network-based computational methods applied in drug repositioning will help us recognize the cutting-edge technologies and offer valuable information for relevant researchers. Therefore, in this review, we present an interpretation of the series of important network-based methods applied in drug repositioning, together with their comparisons and development process.

Keywords: Network-based · Deep learning · Drug repositioning · Heterogeneous network

1. Introduction

The recent COVID-19 pandemic is a convincing case and prime example for faster and further exploration in drug repositioning. In recent years, with the exponentially increased modern technologies, drug repositioning has been creeping up into the ranks of big data quietly.[1,2] Equipped with abundant data relating to drug repositioning, such as genes, proteins, drugs, compounds, and interaction relationships among them,[3] the aforementioned data accelerate our understanding of the development process of drug repositioning which will not only promote the development of technology and algorithms from the perspective of data itself but also bring about a profound revolution to the field of life science.[4] Therefore, it is very important to compare all the current network-based methods used in the drug repositioning field and provide a clear context for researchers.

The rising era of deep learning has a great impact on drug repositioning.[5] While achieving promising results and even surpassing human accuracy in many challenging tasks, the adoption of deep learning in drug repositioning has been comparatively slow.[6,7] The principles of deep learning are closely interrelated to networks.[8] As an influential subfield of machine learning, its core is that it can automatically learn the high-level and abstract feature representation from the input data without specific domain knowledge.[9] More data hold the promise of a better understanding of the mechanisms underlying many complex biological data and structural models in the field of drug repositioning, which takes improved therapies for patients as the ultimate research purpose.[10]

Biological data are often represented in form of networks, such as drug-drug interaction (DDI) networks, gene regulatory networks, and other interaction networks including different biological entities.[6] Graph neural networks are a newly emerging type of deep neural network that can learn nonlinearity relationships of graph data[9] in which network is the carrier of connecting different biological entity data. Furthermore, different neural network classifiers also play a key role in prediction tasks in which network is the skeleton of the classifier.

Two of the most important sub-tasks in drug repositioning are drug-target interaction prediction and drug-drug interaction prediction. For the sake of efficiency, the former provides numerous possibilities for the new use of old drugs. At the same time, it can carry out rapid drug screening and save a lot of human and material resources instead of doing biological experiments one by one. Moreover, the latter can not only provide theoretical support for polypharmacy but also prevent some potential adverse drug-drug interactions in advance.[11] Consequently, in this review, we will explicitly elaborate on the development process of various emerging network-based methods in drug repositioning with the rising era of deep learning.

2. Commonly Used Datasets Summary in Drug Repositioning

Deep learning requires large amounts of training data, which have become widely available with the advent of the big data era. Data is not only the starting point to construct

[a] *T. Song, G. Wang, X. Wang, S. Wang*
College of Computer Science and Technology,
China University of Petroleum
Qingdao 266580, China
E-mail: wangsyun@upc.edu.cn
[b] *M. Ding*
Department of Neurology Medicine,
The Second Hospital,
Cheeloo College of Medicine,
Shandong University,
Ji Nan Shi, Jinan 250033, China
E-mail: 18264181312@163.com
[c] *T. Song, A. Rodriguez-Paton*
Department of Artificial Intelligence,
Faculty of Computer Science,
Polytechnical University of Madrid,
Campus de Montegancedo,
Boadilla del Monte 28660, Madrid, Spain
[d] *X. Wang*
China High Performance Computer Research Center
Institute of Computer Technology,
Chinese Academy of Science,
Beijing, 100190 Beijing, China

a model, but also the core to obtain the optimal model by comparing the advantages and disadvantages of different models.[12] As a matter of fact, there are two major challenges in the field of drug repositioning in terms of data preparation before building a model[13][14]. Firstly, the data is not always available or difficult to find from time to time, which makes many scholars unable to start hands-on practice. Second of all, different models of different papers use completely different datasets, making it difficult to objectively evaluate which model is actually better just by directly looking at the value of different evaluation indicators.

In addition, it is noted that many existing reviews do not have a comprehensive summary of useful datasets related to drug repositioning. As a result, for the sake of facilitating relevant scholars and saving their retrieval time,[15] we list some of the commonly used datasets in drug repositioning since 2003. The following Table 1 includes the characteristic, name, data content, and publication year of those datasets. For specific experiments, researchers could extract specific data from different datasets in Table 1 according to their tasks and objectives, then implement the same data on different models to do comparisons.

3. Network-based Approaches for Drug Repositioning

The development of drug repositioning models is mainly lying in two aspects: one is to enrich the information for prediction tasks with multi-source data in which many scholars construct heterogeneous networks as carriers to fuse multi-source information; the other is the improvement and optimization of algorithms. As the two most important sub-tasks of drug repositioning, drug-target interaction prediction and drug-drug interaction prediction play a crucial role in drug repositioning. As a consequence, we

Tao Song received his Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2013. He is currently a Full Professor in China University of Petroleum, Qingdao, China, and a Juan de la Cierva supported researcher in Polytechnic University of Madrid, Madrid, Spain. His current research interests include DNA computing, membrane computing, and bioinformatics.

Gan Wang is a graduate student in the Department of Computer Science and Engineering at China University of Petroleum, with a research interest in drug repositioning.

Mao Ding received her PH.D. degree from the Shandong University. She is an assistant researcher in the second hospital of Shandong university with a research interest in bioinformatics and biomedicine.

Alfonso Rodriguez-Patón received the B.S. degree in electronic and computational from the University of Santiago de Compostela, Santiago de Compostela, Spain, in 1992, and the Ph.D. degree in computer science from the Polytechnic University of Madrid, Madrid, Spain, in 1999. He is currently a Full Professor with the Faculty of Informatics, Polytechnic University of Madrid. His current research interests include interplay of computer science, biology, and engineering.

Xun Wang received her Ph.D. degree from the Tsukuba University, Tsukuba, Japan, in 2014. She is currently an Associated Professor in China University of Petroleum, Qingdao, China, and grand supported post-doc researcher in China High Performance Computer Research Center, Institute of Computer Technology, Chinese Academy of Science, Beijing, China. Her current research interests include high performance computing for bioinformatics.

Shudong Wang received her Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2004. She is currently a Full Professor in China University of Petroleum, Qingdao, China. Her current research interests include graph theory and bioinformatics.

Table 1. A summary of commonly used datasets applied in drug repositioning.

Characteristic	Highly-cited datasets	Data content	Year
Drug and target	Gold standard dataset (Enzymes, Ion Channels, GPCRs, Nuclear, Receptors)	DTIs	—
Drug-centric	BindingDB[17]	Molecular structure, protein sequence, activity with target	2007
	SuperTarget[18]	Molecular structure, molecular activity with target and pathway, ATC code and side effect	2008
	CHEMBL[19]	Molecular structure, molecular activity with target and therapeutic indications	2012
	OFFsides[20]	Drug side effect and DDIs	2012
	DrugBank[21]	Molecular structure, molecular activity with target	2014
	PubChem[22]	Molecular structure, ATC code, therapeutic indications and side effect	2016
	SIDER[23]	Therapeutic indications and side effect	2016
Protein-centric	STRING[24]	PPIs	2003
	Gene ontology[25]	Gene terms	2004
	BioGrid[26]	Gene terms and PPIs	2006
	HPRD[27]	Sequence and PPIs	2009
Disease-centric	Mesh[28]	Disease phenotypes and terms	2000
	OMIM[29]	Disease terms and associated genes	2005
	HPO[30]	Disease phenotypes and terms	2016

first introduce network-based DTI models in drug repositioning in which three stages of the development are interpreted in detail. Then we make a brief statement about the development of network-based DDI models as well as their similarities and differences with DTI models. Finally, we discuss the advantages of heterogeneous networks and deep learning algorithms.

The discovery of a new drug has always been a bumpy road full of laborious, time-consuming twists and turns.[4] De novo drug design costs a lot of money and faces many challenges. "The most fruitful basis for the discovery of a new drug is to start with an old drug" quoted Sir James Black, a Nobel Prize-winning pharmacologist.[31] However, the specific associations between drug candidates and the interactions with their potential counterparts, such as proteins, diseases and so on, are often difficult to foresee because the underlying mechanism associated with them is largely unknown,[4] complex, or dispersed and buried in masses of information. Hence, most of the successfully repositioned drugs are the result of "serendipity".[31]

In one classification, traditional drug repositioning methods were mainly grouped into three categories: 1 docking simulations, 2 ligand-based approaches and 3 literature text mining in 2012.[32] However, some limitations exist in the use of molecular docking in drug repositioning field. First of all, the three-dimensional (3D) structure of both chemical ligands and proteins severely hinders the application of docking simulation because the structures of many proteins are not entirely analyzed[33][34]. Secondly, molecular docking methods require huge computational resources that result in extended runtimes.[35] Moreover, the results of molecular docking have high false-positive rates resulting in the incomplete modeling of molecular interactions and errors in the determined protein structure.[33] For text mining approaches, shortcomings are usually based on keyword searching and so suffered from an inability to detect new biological findings and also the problem of redundancy in the compound/gene names in the literature.[36]

In another classification, which focuses on methodologies and model itself, has categorized drug repositioning methods into data-driven models and hypothesis-driven models in 2013.[34] One of the most frequently used data-driven approaches is network-based modeling. Networks are simple data structures from which a wide variety of biological data are represented in the form of networks.[34] And there has been a considerably growing interest in the investigation of the structure of such networks, the relationship between the networks, and the underlying biological properties in the network.[37] In comparison, one of the biggest obstacles for hypothesis-driven models is the necessity for quantitative details of interactions between biological entities.[34] On account of analysis and summary of the above two classification methods in drug repositioning, consequently, we focus on the combination of network-based models and ligand-based approaches in drug

repositioning. In the following sections, we introduce the methods related to networks in two important subtasks of the drug repositioning research field based on the timeline.

3.1. Development Process of Network-based Compound/Drug-protein/Target Interaction Prediction

The prediction of drug-target interaction has always been an important research field in the process of drug repositioning, which is of great importance for narrowing down the range of candidate medications.[38] However, due to the disadvantages of time-consuming labor and expensive traditional experiments, only a small proportion of interactions between drugs and targets were verified experimentally in the end. Thus, the computational prediction of drug-target interaction has become an essential way in the process of drug repositioning.

3.1.1. Chemical Molecular Structure and Protein Sequence in Network

Experimental determination of compound-protein interactions or potential drug–target interactions has remained very challenging since 2002.[39] Two of the most commonly used techniques for effective predictions on drug-target interaction were docking simulations and literature text mining around 2006 with some limitations in both methods.[36,40] In 2007, there was some new innovative progress, such as the considerations based on the structure of ligands and the analysis of the topology of nodes in the network.[41] However, none of aforementioned techniques comprehensively integrates protein sequence, chemical structure of drugs and known drug-target network information simultaneously. Under the same gold standard datasets, it seems that the main method to predict drug-target interactions from 2007 to 2010 was mainly based on the supervised network inference along with chemical structure of drugs and sequence of target proteins taken into consideration together[42–44] in 2008, 2009, 2010, respectively. Although the specific details of these[42–44] are slightly different from each other, the fundamental procedure of those models is to embed drugs and proteins in the carrier of interaction network based on the similarity of chemical structure and protein sequence in the first place then implement inference prediction in the network. The principle behind these methods is that when a drug interacts with many target proteins, other drugs with similar structure are more likely to share same targets with it, and the same is true for target proteins. Even if the pharmacological effects originated from chemical structure similarity of drugs[43] and protein sequence similarity were determinants for prediction results of drug-target interaction in 2010, the pharmacological effects were originated from chemical structure similarity of drugs as well. Therefore, their essence is invariant.

3.1.2. The advent of Heterogeneous Networks and Machine Learning Algorithms

With the further development of machine learning algorithms and complex network theory,[32,45,46] there were two main characteristics of drug-target interaction prediction from 2011 to 2016 along with some machine learning algorithms transformed into this subtask sphere. One is the development of heterogeneous networks with more data available in biological sphere, and the other is the more complex machine learning algorithms.[47,48] The simple structure of heterogeneous network is shown in Figure 1. A

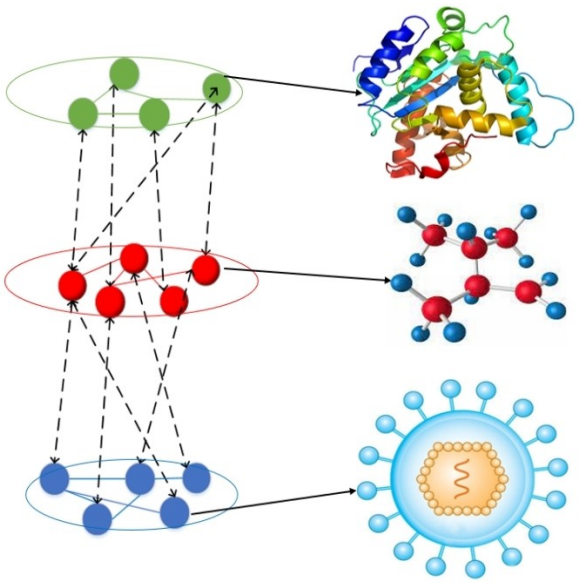


Figure 1. A simple heterogeneous network diagram. In this heterogeneous network diagram, green nodes represent drugs. Red nodes represent proteins, and blue nodes represent diseases. It can be seen as a three-layer architecture. Many machine learning algorithms and neural network algorithms could take this heterogeneous network as the carrier of data, such as random walk, graph neural network and so on. The heterogeneous network brings the drug repositioning task to another level, which can extract the features of the input entity more efficiently and fix it in a certain dimension. It is an end-to-end learning model which can complete the whole prediction task from original data to final outcomes.

simple heterogeneous network diagram. The green node represents protein, the red node represents drug, and the blue node represents disease. It can be regarded as a three-layer architecture. The network can have many types of nodes and multi-layer architecture. The solid line between nodes of the same type indicates that there is a certain relationship between them. Taking the drug layer as an

example, the relationship types can be as follows: 1 the structural similarity between drugs, 2 two drugs sharing one same target, 3 two drugs have synergy or side effects, etc. The double arrow line of the dotted line represents the interaction relationship between two different types of nodes. The same principle goes for other layers than only drug layer. And the black solid line arrow on the right refers to the general picture of three types of nodes.

We can see some traces of origin from.[49–52] Three of them do the prediction tasks based on heterogenous network from many sources of data. Even if almost each of them constructed a heterogeneous network in the first place except for J.-P. Mei,[50] they had different complex machine learning algorithms in the follow-up steps. Since 2010, a considerable number of drug repositioning models are closely related to the identity of heterogeneous networks, and almost half of them are constructed on the basis of heterogeneous networks. See Table 1 and Table 2 for details. Back then, J.-P. Mei[50] (2013) relatively tackled the fundamental problem that a new drug or protein cannot be predicted to found its potential pairing candidates without its known interaction information which is quite common based on the network missing-link inference methods. In 2012 and 2015 respectively, both X. Chen[49] and A. Seal[51] implemented random walk with restart algorithms on heterogeneous network which achieve more promising performances. In Y. Luo[52](2016), the processing steps on similarity data information after the construction of heterogeneous network were put in to place such as Compact feature learning and Matrix operation in which this model outperformed (detailed comparisons[52]) four excellent methods, including BLM-NII(2013),[50] NetLapRLS (2010),[53] HNM (2014)[54] and CMF(2013).[55] However, these methods from 2011 to 2016 partially improve some of shortcomings in previous models but they mainly did not break through the limitations based on the structural information of drugs or proteins. Because the multi-source information in heterogeneous networks along with machine learning algorithms is still based on similarity of the chemical structure of single drug and single protein sequence separately.

3.1.3. Changes Brought by Deep Learning

With deep learning quietly and quickly creeping up into the applications of drug repositioning since 2015, a new generation of technologies to improve the accuracy of drug-target interaction have become more complex and diverse. The common neural network models of deep learning applied in drug repositioning is shown in Figure 2 and Figure 3. Conv and Pool in figure 2 respectively represent convolution and pooling operations and FC represents the fully connected layer. Figure 3 consists of three parts: input layer, hidden layer and output layer. W^n and b^n represent weight parameter matrix and bias

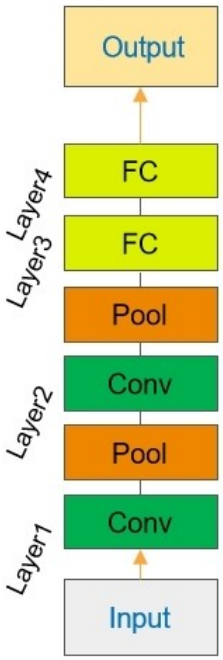


Figure 2. A general convolutional neural network flow chart. Convolutional neural network is mainly used to extract image features. In the field of drug repositioning, it can be used to extract features of both drug and protein. Generally, Conv layer is used to extract and compress features, while FC layer is used to predict the final outcomes.

parameter vector respectively. $X=A^0$ represents initial state. $\hat{Y}$ represents the final prediction result. In addition to this model, there are many other common neural networks, such as RNN and other more complex neural network models listed in table 2 and 3 for details. Last but not least, noticing one of the biggest characteristics of those models is the introduction of nonlinear factors which makes them superior to some traditional machine learning algorithms. Its principle basis can be expressed by mathematical activation function in the following:sigmoid function $f(x) = \frac{1}{1+e^{-x}}$, tanh function $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, and relu function $f(x) = \boldsymbol{max}(0, x)$ and so on.

The contributions of deep learning to drug-target interaction prediction are mainly lying in two aspects: feature learning and neural network classifiers. And there are several traits in the models that are network-based along with deep learning techniques. On the basis of heterogeneous networks based on similarity, the revolution of deep learning is mainly reflected in feature learning. Different techniques based on deep learning are put into place to better represent each drug and protein in the heterogeneous network. In 2018, Olayan R S[56] used nonelinear fusion methods to fuse different similarity networks with path-category-based features for both drugs and proteins in the network. Peng J[57](2020) implemented random walk with restart under heterogeneous network

Table 2. A summary of network-based methods for DTI.

Method	Dataset information	Techniques related to network	Prediction	Year
Bipartite graph learning[42]	gold standard dataset (Enzymes, Ion Channels, GPCRs, Nuclear Receptors) 932 drugs, 989 targets,5127 interactions	embed compounds and proteins on the interaction network into a unified space	drug-target interaction	2008
Bipartite local models[44]	gold standard dataset (Enzymes, Ion Channels, GPCRs, Nuclear Receptors)	use side information of vertices of the drug-target bipartite network to do prediction inference	drug-target interaction	2009
Integrated framework[43]	gold standard dataset (Enzymes, Ion Channels, GPCRs, Nuclear Receptors)	embed drugs and targets into a unified feature space on interaction network based on chemical, pharmacological and genomic data	drug-target interaction	2010
NRWRH[49]	445 210 223 54 drugs 664 204 95 26 targets 2926 1476 635 90 Interactions (KEGG, BRENDA, SuperTarget and DrugBank respectively)	a novel method of network-based Random Walk with Restart on the Heterogeneous network based on similarity	drug-target interaction	2012
BLM-NII[50]	gold standard dataset (Enzymes, Ion Channels, GPCRs, Nuclear Receptors)	predict interaction between a new drug and protein based on neighbor interaction profile in the network	drug-target interaction	2013
RWR[51]	727 compounds,3519 target proteins and 2557 connections among them (DrugBank)	1.construct the heterogeneous network 2.predict DTI based on random walk	drug-target interaction	2015
DTINet[52]	4 types of nodes and 6 types of edges 12015 nodes and 1895445 edges (DrugBank, HPRD, Sider and so on)	the heterogeneous network integrates 4 types of nodes (drugs, proteins, diseases and side-effects) and 6 types of edges	drug-target interaction	2016
SAE(Stacked Autoenc-oder Deep Neural Network)[60]	gold standard dataset (Enzymes, Ion Channels, GPCRs, Nuclear Receptors)	use stacked autoencoders(a special neural network)to learn features of input information and do the prediction task	drug-target interaction	2017
DDR[56]	gold standard dataset (Enzymes, Ion Channels, GPCRs, Nuclear Receptors)	use fusion method to fuse different similarity networks to construct heterogeneous networks	drug-target interaction	2018
DeepConv-DTI[61]	11950 compounds,3675 proteins and 32568 DTIs between them (DrugBank,KEGG and IUPHAR)	1.use deep learning network to yield representation of both proteins and drugs 2.DNN as the classifier	drug-target interaction	2019
DeepCPI[64]	1795801compounds (PubChem,DrugBank,-ChemBL) 464122 proteins (PubChem,DrugBank,ChemBL and so on)	use multimodal neural network to predict binding score between a new compound and protein	compound-protein interaction	2020
TransformerCPI[8]	7003 compounds and 585 proteins	1.a novel transformer neural network 2.the representation of each atom is learned by integrating the features of adjacent atoms in the network	compound-protein interaction	2020
LINE network represent-tation[58]	4601 nodes(134 drugs, 2062 diseases,613 proteins and so on) 94439 interactions(18416 drug-disease,19237 protein-protein and etc)	a novel network embedding method (LINE) to better embed into low-dimensional vector space	drug-target interaction	2020
DeepCDA[62]	442, 229, 81417 drugs (Davis, KIBA, BindingDB) and 68, 2111, 79536 proteins (Davis, KIBA. BindingDB) and interactions among them	use the combination of CNN and LSTM to encode compound and protein	compound-protein binding affinity	2020
DTICNN[57]	12015 nodes and 1895445 edges 4 types of nodes (drug, protein and etc)and 6 types of edges (drug-protein, drug-drug, drug-disease and so on)	1.restart random walk to extract features from the heterogeneous network 2.convolutional neural network as classifier to do the prediction task 3.use autoencoder to reduce feature dimension	drug-target interaction	2020
iDrug[65]	1321 drugs,3966 diseases and 111481 interactions between them (CTD database) 946 drugs,3610 targets and 10234 interactions between them (DrugBank database)	use Within-network factorization and Cross-network consistency methods to construct the final network for better embedding	drug-target interaction	2020
DeepACTION[66]	5877 drugs, 3348 targets and 12674 interactions (DrugBank)	the features of drugs and proteins are input to Convolutional neural network(CNN) for prediction	drug-target interaction	2020
AOPEDF[59]	732 drugs,1915 proteins,440 diseases and so on. 4978 DPIs,132768 DDIs and etc	1.integrate 15 biological networks to construct heterogeneous network 2.arbitrary-order proximity preserved network embedding	drug-target interaction	2020

Table 2. continued

Method	Dataset information	Techniques related to network	Prediction	Year
FRnetDTI[63]	gold standard dataset (Enzymes, Ion Channels, GPCRs, Nuclear Receptors)	the model includes two convolutional neural networks: FRnet-Encode for generating features and FRnet-Predict for identifying interaction probability	drug-target interaction	2020

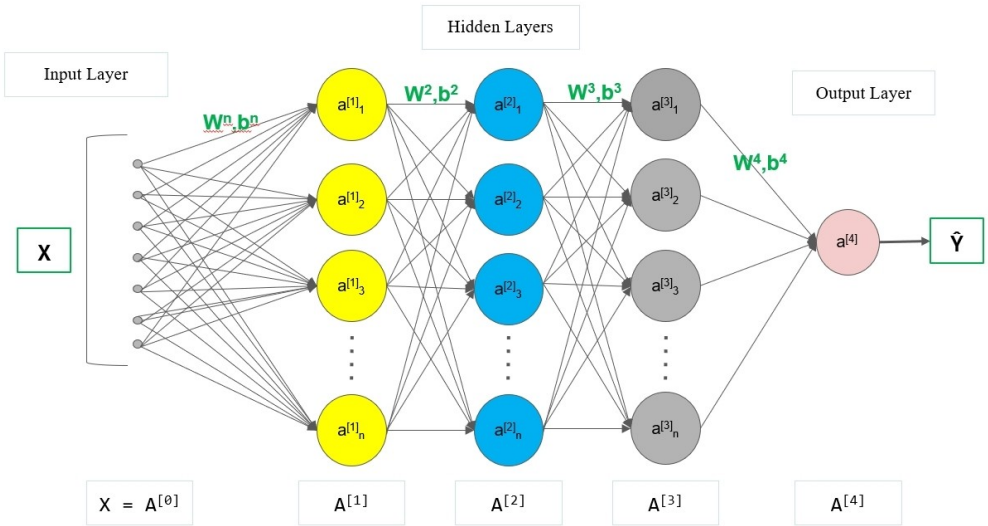


Figure 3. A general and simple deep neural network model diagram. The input layer is usually the input data. The hidden layer abstracts the input data to another dimension space to show its more abstract features, which can better classify the input data nonlinearly. Generally, there is only one neuron in the output layer, and sigmoid is usually used as the activation function. This kind of output layer is usually used to deal with binary classification tasks. Different scholars could choose different activation functions like softmax and DNN parameters according to different tasks.

based on similarity, then used a denoising autoencoder to learn the essential features along with CNN prediction classifier. In B.Y. Ji[58](2020), Ji B Y further extended the case beyond the feature generated by embedding nodes in heterogeneous network. They combined both structural and functional information of nodes based on Large-scale Information Network Embedding(LINE) in the heterogeneous network to form the final feature vector. The same idea was implemented in X. Zeng[59](2020), but arbitrary-order proximity preserved network embedding method.

Although these models mentioned just now have some improvements in the algorithms of network-based inference and machine learning, especially the functional embedding of nodes by deep learning techniques. However, it still cannot solve the problem mentioned above that a new drug or protein cannot be predicted to potentially pair candidates without pre-known interactions.

At the time, the L. Wang[60](2017) used stacked autoencoder to adequately extract the raw data information of protein sequence. Then final feature descriptors are constructed by combining with molecular fingerprint and fed into the rotation forest. In I. Lee[61](2019), convolutional neural network(CNN) was employed on raw protein sequences to extract its feature and fingerprints of drugs are extracted by fully connected layer as well along with DNN classifier, which outperformed DeepDTI(2017) and DeepDTA(2018). Karim A[62] used CNN-LSTM to better learn features of both drugs and proteins with DNN as prediction classifier. TransformerCPI[8] is an improved model of compound-protein interaction by sequence-based deep learning with self-attention mechanism and label inversion experiment. The authors of TransformerCPI[8] (2020) also make comparisons between their model and some other proposed models and baselines on C.elegans and BingdingDB datasets in which their model outperforms GraphDTA, GCN, CPI-GCN, KNN and etc. Rayhan F[63] employed two convolutional neural networks for both feature learning and prediction. These models substantively solve the prediction task given any new drugs or proteins.

3.2. Development Process of Network-based Drug-drug Interaction Prediction

As complex and co-existing diseases are commonly cured by drug combinations, polypharmacy of different drugs has

Table 3. A summary of network-based methods for DDI.

Method	Dataset information	Techniques/work related to network	Prediction	Year
An ensemble model of different network methods[70]	548 drugs and 48584 DDIs (TWOSIDES) and many other information like	combine neighbor recommender and random walk methods together to do the prediction task	drug-drug interaction	2017
Decagon[72]	645 drugs and 19085 proteins (715612 protein-protein 4651131 drug-drug and 18596drug-protein edges)	a new graph convolutional neural network for multi-relational link prediction	964 types of drug-drug side effects	2018
DeepDDI[73]	192284 DDIs(DrugBank gold standard)	use 8 layers deep neural network as prediction classifier	86 DDI types	2018
Conv-LSTM[71]	2898937 DDIs(DrugBank,KEGG Drug and so on) 260 DDI relation types	1.different embedding methods based on Network 2.Conv-LSTM for predicting DDI	drug-drug interaction	2019
Deep-forward network with autoencoder[74]	1597 drugs and 188258DDIs of 106 types	use a deep-forward network to predict a specific type of DDI	drug-drug interaction	2019
MHCADDI(multi-head co-attentive drug-drug interactions)[75]	645 drugs and 19085 proteins (715612 protein-protein 4651131 drug-drug and 18596drug-protein edges)	1.use multilayer perceptron neural network to project input to features 2.use GCN to get the neighborhood information between atoms for drug1 3.use transformer to get atom information of drug2 for drug1	binary classification or multi-label classification for all side effects	2019
TIP(Tri-graph information Propagation)[76]	P–P graph(19081 proteins and715612 P–P edges) P–D graph(3648 proteins,284 drugs and 18690 P–D edges) D–D graph(645 drugs and 63473 D–D edges)	1.TIP uses MPNN(message passing neural network) as encoder 2.TIP uses 2-layer neural network multi-label classifier as the decoder	multi types of drug side effects	2020
DeepDrug[77]	1599 drugs,559 proteins and 38716 DDIs,1268 DTIs	use graph convolutional networks to learn graphical representations of drugs and proteins	drug-drug interaction	2020
SkipGNN[78]	5018 drugs,2325 proteins and 15139 DTIs(BIOSNAP-DTI) and other datasets	a novel neural network for better embedding for each node	drug-drug interaction	2020
DPDDI[79]	three different datasets(DB1, DB2, DB3)	1.use GCN to extract features from DDI network 2.a deep neural network as a predictor	drug-drug interaction	2020
DDI-MDAE[80]	2367 drugs,2411 targets,285 enzymes and 314 pathways(DrugBank)	adopt deep neural network to manage network representation learning	drug-drug interaction	2020
GCN-BMP(Graph Convolutional Network with Bond-aware)[81]	548 drugs and 48548 drug-drug pairs for BinaryDDI 1704 molecules and 191400 interacting pairs for MultiDDI	1.use Siamese GCN as encoder to transfer molecular data into embedding vectors 2.use HOLE-style neural network as predictor for input drugs	binaryDDI and multiDDI	2020
SumGNN[69]	DrugBank(1709drugs, 136, 351 drug pairs, 86 types of relation)	a new graph neural network that summarizes diverse sources of drug information	multiple DDI relations	2021

drawn a great attention to fighting against sophisticated diseases. However, adverse drug events are also a serious problem in today's world. In the United States, thousands of dollars are forced to be spent every year on injuries and even deaths caused by the adverse effects of drugs. Therefore, the prediction of drug-drug interaction is very important. Synergy among drugs helps us enhance the effectiveness of drugs. At the same time, prediction on adverse effects help us avoid potential future harms in advance.

The concept that similar molecules have similar biological characteristics has been explored by pharmaceutical chemists since 2002.[67] The model based on molecular fingerprints has also been successfully applied to the recognition of structurally similar molecules.[68]

The idea can be extended and applied to the drug-drug interaction prediction task.[69] S. Vilar[69] calculated the similarity among different drugs based on their molecular fingerprint to predict the potential relations between them. They compared their model[69] with some classic and common network methods on two different datasets of DrugBank and TWOSIDES, such as MLP, Deepwalk, Node2-Vec and so on, which achieved the best performance in DDI prediction on two datasets.Zhang W[70](2017) briefly illus-

trated several drug-drug similarity methods based on known drug-drug interactions, such as Common neighbor similarity, Adamic-Adar similarity, Resource allocation similarity and so on. Karim M R[71](2019) compared six different network embedding methods such as PBG, SimplelE, KGloVe, TransE, CrossE, RDF2Vec in 2019. At the same time, they compared some classical machine learning methods with Conv-LSTM under same embedding methods. Experiments proved that their model was superior to some typical machine learning methods (such as SVM, RF, KNN and etc.) in different network embedding methods.

The same as drug- target interaction prediction, they are from the original cases based on structural similarity to the later heterogeneous network of multi-source information. Feature learning from the original raw data to extract information brought by the revolution of neural network improves accuracy of previous models. The difference is that the predictive value of drug-target interaction can be a number between 0 and 1, while DDI can be a multi-type predictive value, not just a binary classification problem. In view of that the essence of the development process is roughly the same between DDI and DTI, Table 3 only lists some network-based DDI models from 2017 to 2020. For detailed information, please refer to the specific literatures. See Table 3 for the specific network-based model method.

4. The Advantages of Heterogeneous Network and Deep Learning Algorithms in Drug Repositioning

Multi-source data and new algorithms are the two main contributions to improve the performance of the network-based models in drug repositioning. Deep learning is a new generation of evolutionary algorithms, which is characterized by the extraction of high-level features and strong classification ability.

The advantages of deep learning algorithms can be boiled down into the following three attributes. First of all, it can directly extract the features of high-dimensional data or reduce the dimension of data and then learn from data. Secondly, it can learn the nonlinear relationship between different complex data. Thirdly, it can efficiently complete the task of classification or regression without specific relevant domain knowledge. For heterogeneous networks, there are two main advantages. The first is that it can be constructed as the carrier of original data so that each node could generate its functional features according to their relative position with their neighbor nodes in the network without complex methods to generate different features of different data. Taking an instance of drug-target interaction data, it is not necessary to get the SMILES, Morgan fingerprint, and other information of drugs on which the same is true for proteins. Conversely, different algorithms can be implemented directly on heterogeneous networks to generate the functional features of original data nodes and complete prediction tasks. The second is that it is an end-to-end learning model, which can predict the potential interaction relationship directly from the original data without complex processes such as feature generation, data dimensionality reduction, feature learning, and data identification.

An increasing number of researchers have paid ever more attention to heterogeneous networks and deep learning algorithms in recent years. They construct a heterogeneous information network to extract features through graph convolution network and similarity method and finally use a three-layer MLP neural network to do the prediction task.[82] Kexin Huang[78] also constructs a heterogeneous network, which contains drug, protein, target, gene and disease nodes. Different from the traditional graph neural network, this paper considers extra information of the skip similarity between second-order interactions among different nodes in the network and achieves a relatively good model performance. High-order similarity and other revolutionary network embedding methods will also bring heterogeneous networks more promising performance. Hopefully, with the arduous efforts of network-based approaches together, drug repositioning will enter another new stage of development.

5. Contributions of Our Work

In this section, we boil down our innovative contributions to the scientific community into three points. To start with, we sort out the specific information (such as model names, references, model task and specific data information used and proposed year of the model) of different network-based models for predicting drug-target interaction prediction and drug-drug interaction prediction dedicating to make drug repositioning more efficient in chronological order. Second of all, we analyze and sum up the development process of drug-target interaction prediction and drug-drug interaction prediction in detail from 2002 to 2021 where three important stages are discussed. Thirdly, we list some important datasets in drug repositioning and elaborate advantages of deep learning methods and heterogeneous networks.

6. Summary

Key points:

1. There are three main manifestations of the network in network-based methods. The first is to use the network as the carrier of the original data to generate features. The second is to use neural networks as the classifier of the model. The third is the end-to-end network learning model, which includes the whole process of generating features, data dimensionality reduction, learning features, predicting the outcomes and so on.

2. Deep learning holds great promise for scientific advances in drug repositioning. There is also a general consensus that the lack of 'interpretability' represents a limitation to their deployment in actual practice. In the long term, interpretability will be a very important subtopic of deep learning in drug repositioning.
3. From the perspective of feature types, the current network-based methods in the field of drug repositioning can be roughly divided into two categories. One is based on the chemical structure of the entity, and the other is based on the functional importance of the entity in the interaction network. In the near future, if there is a better way to fuse or ensemble both structural information and functional information simultaneously, it is hopeful to further improve the performance of the prediction task.

Acknowledgements

This work was supported by Natural Science Foundation of China (Grant Nos. 61873280, 61972416, 61873281), Taishan Scholarship (tsqn201812029), Natural Science Foundation of Shandong Province (No. 2019GGX101067), Foundation of Science and Technology Development of Jinan (201907116), Fundamental Research Funds for the Central Universities (18CX02152A, 19CX05003A-6) and Grant from Juan de la Cierva and Talento-Comunidad de Madrid.

Conflict of Interest

None declared.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

[1] Y. Jing, Y. Bian, Z. Hu, L. Wang, X. Q. S. Xie, *AAPS J.* **2018**, *20*, DOI 10.1208/s12248-018-0210–0.
[2] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, T. Blaschke, *Drug Discovery Today* **2018**, *23*, 1241–1250.
[3] Y. Zhu, O. Elemento, J. Pathak, F. Wang, *Briefings Bioinf.* **2018**, *20*, 1308–1321.
[4] S. Schroedl, *Drug Discovery Today Technol.* **2019**, *32–33*, 9–17.
[5] S. Agatonovic-Kustrin, R. Beresford, *J. Pharm. Biomed. Anal.* **2000**, *22*, 717–727.
[6] P. Mamoshina, A. Vieira, E. Putin, A. Zhavoronkov, *Mol. Pharm.* **2016**, *13*, 1445–1454.
[7] S. Min, B. Lee, S. Yoon, *Briefings Bioinf.* **2017**, *18*, 851–869.
[8] L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, X. Luo, K. Chen, H. Jiang, M. Zheng, *Bioinformatics* **2020**, *36*, 4406–4414.
[9] D. Berrar, W. Dubitzky, *Briefings Bioinf.* **2021**, *22*, 1513–1514.
[10] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. Huang, A. Gitter, C. S. Greene, *bioRxiv* **2017**, DOI 10.1101/142760.
[11] T. Song, S. Pang, S. Hao, A. Rodríguez-Patón, P. Zheng, *Neural Process. Lett.* **2019**, *50*, 1485–1502.
[12] T. Song, X. Zeng, P. Zheng, M. Jiang, A. Rodriguez-Paton, *IEEE Trans. Nanobioscience* **2018**, *17*, 474–484.
[13] T. Song, P. Zheng, M. L. D. Wong, X. Wang, *Inf. Sci.* **2016**, *372*, 380–391.
[14] T. Song, A. Rodriguez-Paton, P. Zheng, X. Zeng, *IEEE Trans. Cogn. Dev. Syst.* **2018**, *10*, 1106–1115.
[15] T. Song, L. Pan, T. Wu, P. Zheng, M. L. D. Wong, A. Rodriguez-Paton, *IEEE Trans. Nanobioscience* **2019**, *18*, 176–190.
[16] K. Huang, C. Xiao, L. Glass, J. Sun, *arXiv* **2020**, DOI 10.1093/bioinformatics/btaa880.
[17] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, M. K. Gilson, *Nucleic Acids Res.* **2007**, *35*, 198–201.
[18] S. Günther, M. Kuhn, M. Dunkel, M. Campillos, C. Senger, E. Petsalaki, J. Ahmed, P. Bork, R. Preissner, *Nucleic Acids Res.* **2008**, *36*, 919–922.
[19] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, *Nucleic Acids Res.* **2012**, *40*, 1100–1107.
[20] N. P. Tatonetti, P. P. Ye, R. Daneshjou, R. B. Altman, *Sci. Transl. Med.* **2012**, *4*, DOI 10.1126/scitranslmed.3003377.
[21] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. B. Han, Y. Zhou, D. S. Wishart, *Nucleic Acids Res.* **2014**, *42*, 1091–1097.
[22] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, S. H. Bryant, *Nucleic Acids Res.* **2016**, *44*, D1202–D1213.
[23] M. Kuhn, I. Letunic, L. J. Jensen, P. Bork, *Nucleic Acids Res.* **2016**, *44*, D1075–D1079.
[24] C. von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, B. Snel, *Nucleic Acids Res.* **2003**, *31*, 258–261.
[25] M. A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, R. White, *Nucleic Acids Res.* **2004**, *32*, 258–261.
[26] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, *Nucleic Acids Res.* **2006**, *34*, 535–539.
[27] T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, S. Ramabadran, R. Chaerkady, A. Pandey, *Nucleic Acids Res.* **2009**, *37*, 767–772.
[28] N. Baumann, *Int. J. Clin. Pract.* **2016**, *70*, 171–174.
[29] A. Hamosh, A. F. Scott, J. Amberger, C. Bocchini, D. Valle, V. A. McKusick, *Nucleic Acids Res.* **2002**, *30*, 52–55.
[30] S. Köhler, N. A. Vasilevsky, M. Engelstad, E. Foster, J. McMurry, C. J. Mungall, M. Haendel, P. N. Robinson, *Nucleic Acids Res.* **2017**, *45*, D865–D876.
[31] J. K. Yella, S. Yaddanapudi, Y. Wang, A. G. Jegga, *Pharmaceuticals* **2018**, *11*, DOI 10.3390/ph11020057.
[32] M. Gönen, *Bioinformatics* **2012**, *28*, 2304–2310.
[33] J. T. Dudley, T. Deshpande, A. J. Butte, *Briefings Bioinf.* **2011**, *12*, 303–311.
[34] M. Lotf Shahreza, N. Ghadiri, S. R. Mousavi, J. Varshosaz, J. R. Green, *Briefings Bioinf.* **2018**, *19*, 878–892.
[35] H. Ding, I. Takigawa, H. Mamitsuka, S. Zhu, *Briefings Bioinf.* **2013**, *15*, 734–747.
[36] S. Zhu, Y. Okuno, G. Tsujimoto, H. Mamitsuka, *Bioinformatics* **2005**, *21*, 245–251.
[37] L. U. Sudha, J. Baskaran, S. Thamizharasan, *Int. J. Appl. Eng. Res.* **2015**, *10*, 1511–1518.

[38] C. Zheng, Z. Guo, C. Huang, Z. Wu, Y. Li, X. Chen, Y. Fu, J. Ru, P. Ali Shar, Y. Wang, Y. Wang, *Sci. Rep.* **2015**, *5*, 1–10.
[39] F. G. Kuruvilla, A. F. Shamji, S. M. Sternson, P. J. Hergenrother, S. L. Schreiber, *Nature* **2002**, *416*, 653–657.
[40] A. C. Cheng, R. G. Coleman, K. T. Smyth, Q. Cao, P. Soulard, D. R. Caffrey, A. C. Salzberg, E. S. Huang, *Nat. Biotechnol.* **2007**, *25*, 71–75.
[41] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, B. K. Shoichet, *Nat. Biotechnol.* **2007**, *25*, 197–206.
[42] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, *Bioinformatics* **2008**, *24*, i232–i240.
[43] Y. Yamanishi, M. Kotera, M. Kanehisa, S. Goto, *Bioinformatics* **2010**, *26*, i246–i254.
[44] K. Bleakley, Y. Yamanishi, *Bioinformatics* **2009**, *25*, 2397–2403.
[45] F. Cheng, C. Liu, J. Jiang, W. Lu, W. Li, G. Liu, W. Zhou, J. Huang, Y. Tang, *PLoS Comput. Biol.* **2012**, *8*, DOI 10.1371/journal.-pcbi.1002503.
[46] M. Sun, S. Zhao, C. Gilvary, O. Elemento, J. Zhou, F. Wang, *Briefings Bioinf.* **2020**, *21*, 919–935.
[47] X. Y. Yan, S. W. Zhang, C. R. He, *Comput. Biol. Chem.* **2019**, *78*, 460–467.
[48] H. Shi, S. Liu, J. Chen, X. Li, Q. Ma, B. Yu, *Genomics* **2019**, *111*, 1839–1852.
[49] X. Chen, M. X. Liu, G. Y. Yan, *Mol. BioSyst.* **2012**, *8*, 1970–1978.
[50] J.-P. Mei, C.-K. Kwoh, P. Yang, X.-L. Li, J. Zheng, *Bioinformatics* **2012**, *29*, 238–245.
[51] A. Seal, Y. Y. Ahn, D. J. Wild, *J. Cheminf.* **2015**, *7*, DOI 10.1186/s13321-015-0089-z.
[52] Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, W. Kuang, J. Peng, L. Chen, J. Zeng, *Nat. Commun.* **2017**, *8*, DOI 10.1038/s41467-017-00680–8.
[53] Z. Xia, L. Y. Wu, X. Zhou, S. T. C. Wong, *BMC Syst. Biol.* **2010**, *4*, 1–16.
[54] W. Wang, S. Yang, X. Zhang, J. Li, *Bioinformatics* **2014**, *30*, 2923–2930.
[55] G. Pandey, Association for Computing Machinery, ACM Digital Library., I. . International Conference on Knowledge Discovery & Data Mining (19th: 2013: *Chicago, USA, August 2013*, **n.d.**
[56] R. S. Olayan, H. Ashoor, V. B. Bajic, *Bioinformatics* **2018**, *34*, 1164–1173.
[57] J. Peng, J. Li, X. Shang, *BMC Bioinf.* **2020**, *21*, DOI 10.1186/s12859-020-03677–1.
[58] B. Y. Ji, Z. H. You, H. J. Jiang, Z. H. Guo, K. Zheng, *J. Transl. Med.* **2020**, *18*, DOI 10.1186/s12967-020-02490-x.
[59] X. Zeng, S. Zhu, Y. Hou, P. Zhang, L. Li, J. Li, L. F. Huang, S. J. Lewis, R. Nussinov, F. Cheng, *Bioinformatics* **2020**, *36*, 2805–2812.
[60] L. Wang, Z. H. You, X. Chen, S. X. Xia, F. Liu, X. Yan, Y. Zhou, K. J. Song, *J. Comput. Biol.* **2018**, 361–373.
[61] I. Lee, J. Keum, H. Nam, *PLoS Comput. Biol.* **2019**, *15*, DOI 10.1371/journal.pcbi.1007129.
[62] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J. B. Ghasemi, A. Masoudi-Nejad, *Bioinformatics* **2020**, *36*, 4633–4642.
[63] F. Rayhan, S. Ahmed, Z. Mousavian, D. M. Farid, S. Shatabda, *Heliyon* **2020**, *6*, DOI 10.1016/j.heliyon.2020.e03444.
[64] F. Wan, Y. Zhu, H. Hu, A. Dai, X. Cai, L. Chen, H. Gong, T. Xia, D. Yang, M. W. Wang, J. Zeng, *Genomics Proteomics Bioinf.* **2019**, *17*, 478–495.
[65] H. Chen, F. Cheng, J. Li, *PLoS Comput. Biol.* **2020**, *16*, DOI 10.1371/journal.pcbi.1008040.
[66] S. M. Hasan Mahmud, W. Chen, H. Jahan, B. Dai, S. U. Din, A. M. Dzisoo, *Anal. Biochem.* **2020**, *610*, DOI 10.1016/j.ab.2020.113978.
[67] Y. C. Martin, J. L. Kofron, L. M. Traphagen, *J. Med. Chem.* **2002**, *45*, 4350–4358.
[68] S. Costanzi, S. Vilar, D. Micozzi, F. M. Carpi, G. Ferino, A. Vita, S. Vincenzetti, *ChemMedChem* **2011**, *6*, 1452–1458.
[69] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, C. Friedman, *J. Am. Med. Informatics Assoc.* **2012**, *19*, 1066–1074.
[70] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, X. Li, *BMC Bioinf.* **2017**, *18*, DOI 10.1186/s12859-016-1415–9.
[71] M. Rezaul Karim, M. Cochez, J. B. Jares, M. Uddin, O. Beyan, S. Decker, in *ACM-BCB 2019 - Proc. 10th ACM Int. Conf. Bioinformatics, Comput. Biol. Heal. Informatics*, Association For Computing Machinery, Inc, **2019**, pp. 113–123.
[72] M. Zitnik, M. Agrawal, J. Leskovec, *Bioinformatics* **2018**, *34*, i457–i466.
[73] J. Y. Ryu, H. U. Kim, S. Y. Lee, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4304–E4311.
[74] G. Lee, C. Park, J. Ahn, *BMC Bioinf.* **2019**, *20*, 1–8.
[75] A. Deac, Y. H. Huang, P. Velickovic, P. Lio, J. Tang, *arXiv* **2019**.
[76] H. Xu, S. Sang, H. Lu, *arXiv* **2020**, 1–8.
[77] X. Cao, R. Fan, W. Zeng, *bioRxiv* **2020**, DOI 10.1101/2020.11.09.375626.
[78] K. Huang, C. Xiao, L. M. Glass, M. Zitnik, J. Sun, *Sci. Rep.* **2020**, *10*, DOI 10.1038/s41598-020-77766–9.
[79] Y. H. Feng, S. W. Zhang, J. Y. Shi, *BMC Bioinf.* **2020**, *21*, 419.
[80] Y. Zhang, Y. Qiu, Y. Cui, S. Liu, W. Zhang, *Methods* **2020**, *179*, 37–46.
[81] X. Chen, X. Liu, J. Wu, *Methods* **2020**, *179*, 47–54.
[82] Z. Wang, M. Zhou, C. Arnold, *Bioinformatics* **2021**, *36*, I525–I533.

Received: November 9, 2021
Accepted: December 5, 2021
Published online on December 30, 2021

