Biomedical data and deep learning computational models for predicting compound-protein relations

Qichang Zhao†, Mengyun Yang†, Zhongjian Cheng, Yaohang Li, and Jianxin Wang*

Abstract—The identification of compound-protein relations (CPRs), which includes compound-protein interactions (CPIs) and compound-protein affinities (CPAs), is critical to drug development. A common method for compound-protein relation identification is the use of *in vitro* screening experiments. However, the number of compounds and proteins is massive, and *in vitro* screening experiments are labor-intensive, expensive, and time-consuming with high failure rates. Researchers have developed a computational field called virtual screening (VS) to aid experimental drug development. These methods utilize experimentally validated biological interaction information to generate datasets and use the physicochemical and structural properties of compounds and target proteins as input information to train computational prediction models. At present, deep learning has been widely used in computer vision and natural language processing and has experienced epoch-making progress. At the same time, deep learning has also been used in the field of biomedicine widely, and the prediction of CPRs based on deep learning has developed rapidly and has achieved good results. The purpose of this study is to investigate and discuss the latest applications of deep learning techniques in CPR prediction. First, we describe the datasets and feature engineering (i.e., compound and protein representations and descriptors) commonly used in CPR prediction methods. Then, we review and classify recent deep learning approaches in CPR prediction. Next, a comprehensive comparison is performed to demonstrate the prediction performance of representative methods on classical datasets. Finally, we discuss the current state of the field, including the existing challenges and our proposed future directions. We believe that this investigation will provide sufficient references and insight for researchers to understand and develop new deep learning methods to enhance CPR predictions.

Index Terms—Virtual Screening, Compound-protein Relation Prediction, Deep Learning

1 Introduction

Drug development remains a key issue and challenge in improving the current field of biomedicine. The identification of relations between drugs/compounds and their targets is crucial for drug development [1], [2]. Drug targets refer to the biomolecules, mainly protein molecules, that are involved in the process of disease production and can act with related drugs to treat diseases [3]. A series of functions and effects of drugs are related to the specific recognition of target proteins. Newly discovered compound-protein relations (CPRs) are critical to drug repositioning [4] and the discovery of new drugs that may be more effective in treating certain diseases at low risk.

There are many small-molecule compounds in the database of compounds and bioactivity that have not yet been used as drugs. In fact, the interactions of most small-molecule compounds with proteins are still unknown. There are only approximately 2,110 small-molecule drugs approved by the FDA (Food and Drug Administration) and 4,964 experimental drugs [5]. Only 3,150 out of the approximately 20,000 human proteins are associated with these drugs according to the statistics of DrugBank [6]. The number of human drug targets is approximated at 4,500 [7]. Therefore, a large number of potential drug-protein relations need to be verified. Detecting the relation of these compounds with disease-related genes and target proteins is particularly useful for drug development [8], [9], [10]. Furthermore, the information can help narrow the range of drug candidates in drug discovery [11], [12], [13].

During the drug development process, as shown in Figure 1, high-throughput screening (HTS) experiments are performed to identify the lead compounds acting on target proteins. After the lead compounds are optimized, they enter clinical trials. Only compounds that pass clinical trials can become approved drugs. However, conducting HTS experiments is expensive and time-consuming. Besides, the number of the expressed proteins and existing compounds is massive, so it isn't a realistic way to verify them one by one through HTS experiments [14].

To solve the above challenges and problems, computational methods have been introduced and have attracted increasing attention. The computational identification of unknown compound-protein pairs using statistical models or machine learning methods is called virtual screening (VS). In the drug development process, VS approaches are

- *Q. Zhao, M. Yang, Z. Cheng and J.Wang are with the School of Computer Science and Engineering, Central South University, Changsha 410083,China and Hunan Provincial Key Lab on Bioinformatics, Central South University, Changsha, 410083, China. E-mail: jxwang@mail.csu.edu.cn.*
- *M. Yang is also with School of Science, Shaoyang University, Shaoyang 422000, China. E-mail: mengyunyang@csu.edu.cn.*
- *Y. Li is with the Department of Computer Science, Old Dominion University, Norfolk 23529, USA. E-mail: yaohang@cs.odu.edu.*
- * *Corresponding author.*
- † *Authors contributed equally.*

usually carried out before HTS, which significantly reduces the search space for drug targets and thus greatly saves time and reduces costs for wet-lab research [15].

At present, there are three types of computational methods for predicting CPRs. The first category encompasses ligand-based methods that are based on the principle of chemical similarity, which states that similar molecules tend to share similar biological activities and usually bind similar proteins [16], [17]. These methods rely on prior knowledge of biologically active ligands and protein structures [18], and they use the most chemically related structural similarities between protein ligands to predict interactions [19]. However, these methods are generally not suitable for situations where there are a few or no known ligands for a given protein. In these cases, models cannot be trained or prediction results may be unreliable [20].

The second method is the docking approach. It explores the main binding modes of compounds and proteins in three-dimensional (3D) space [21]. Because the 3D structures of some proteins are unknown and the determination of the 3D structure is still a challenging problem, the wide use of docking approaches at a large scale has been limited [14]. In addition, dealing with the flexibility of receptor proteins may be difficult because many variances must be considered [22].

The third category of computational methods is based on statistics or machine learning. It takes advantage of both compound (drug) and protein (target) information to make predictions. This is by far the most popular and cost-effective category and achieves more competitive results. These methods require compound-protein pairs to be explicitly represented as fixed-length feature vectors that reflect the physical, chemical, and molecular properties of the corresponding molecule. In addition, each compound-protein pair has a label to indicate whether it is a known interaction (binary relation) or affinity value (binding strength information). Using feature vectors and labels as input, various supervised machine learning methods can be developed to predict CPRs [23], [24], [25], [26], [27], [28].

However, due to the complexity and diversity of biological information data, traditional machine learning algorithms are limited in their ability to process these raw data. Moreover, the construction of a machine learning process must be based on the design of excellent feature engineering methods to transform the raw data into feature vectors. The quality of feature vectors directly affects the final performance of machine learning algorithms. Designing appropriate feature vectors requires rich professional knowledge and a precise engineering pipeline. The development of deep learning provides a new and promising direction for mining useful information in bioinformatics data. Deep learning is a multilevel representation learning method [29] which is a combination of multiple nonlinear modules. Each module transforms one level (starting with the raw input) representation to a higher and more abstract level representation. With enough of these transformations, complex functions and superior representations can be learned. The key advantage of deep learning is that the functionality of these layers is not designed by human engineers but is directly learned from data. Complex structures in high-dimensional data can be well discovered without human intervention. In the face of massive amounts of raw biological data, the automatic extraction capability of deep learning is particularly indispensable. Figure 1 shows the main stages of the drug development process and how deep learning is involved in this process. Although drug discovery based on deep learning is still in its infantry stage, it has shown great prospects. The research team from the University of Toronto tested the top compounds, predicted by AtomNet [30], that would bind to the glycoprotein and prevent Ebola virus entry into a cell in an *in vivo* assay and found a drug with a high binding affinity. The BenevolentAI developed Bavisant for excessive daytime sleepiness in patients with Parkinson's disease by their AI-based drug discovery process, which is in phase IIb trials. Bavisant was originally developed for attention-deficit hyperactivity disorder. BERG's leading product candidate, BPM31510 (ubidecarenone), works by correcting cancer cell metabolism, thereby reactivating apoptosis. BERG recently completed phase II trials for patients with advanced pancreatic cancer, which is extremely aggressive and difficult to treat. This is one of the most exciting results of artificial intelligence research related to drug discovery so far.

Compared with the previous reviews on this topic of CPR prediction [3], [31], [32], [33], our survey focuses on the application of deep learning in predicting CPRs. In addition, we describe a variety of data, including databases with storage properties and different kinds of relations, that might be used for CPR prediction tasks. Although deep learning models have also been covered in reviews [34], [35], [36], they are more likely to describe feature- or machine learning-based methods. Our survey is more comprehensive and up-to-date regarding deep-learning-based models, including the input of these models, feature extraction schemes, and different types of deep learning modules. We also perform an empirical comparison among various deep learning methods from the perspective of two different tasks, the compound-protein interactions (CPIs) task and the compound-protein affinities (CPAs) task, and discuss their advantages and limitations based on the results.

To sum up, the main contributions of this paper, which make it different from the other reviews, are as follows:

- Analyzing feature extraction modules corresponding to different types of inputs.
- Summarizing different types of CPR prediction models based on deep learning and highlighting their contributions.
- Comparing the performance of different methods on public datasets and providing an up-to-date benchmark dataset for CPI prediction.

The rest of this article is organized as follows. The second section describes various data sources required to perform the prediction of CPRs. The third section lists a variety of representative deep learning feature extraction methods for both compounds and proteins. Next, the fourth section outlines the classification and development process of deep learning-based methods for CPR prediction tasks. The fifth section explains performance metrics and presents a comparison of the results of several typical methods applied to benchmark data in the cases of CPIs and CPAs. Finally, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2021.3069040, IEEE/ACM Transactions on Computational Biology and Bioinformatics

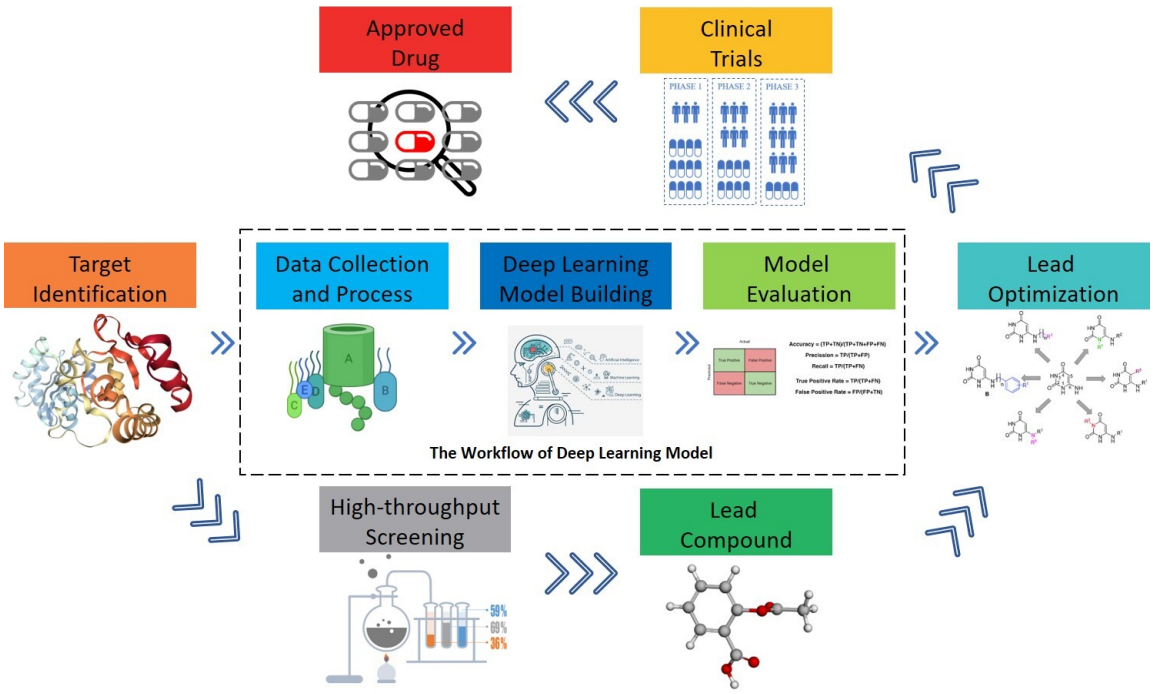


Fig. 1. A broad overview of drug development and the place of deep learning in this process.

last section summarizes the field and briefly discusses the future development direction.

2 Available Biomedical Data

With the development of high-throughput technology and virtual screening in biology, a large amount of biological activity data has been produced in recent years. It has opened a new era for computational biomedical research. To predict potential CPRs, researchers need to gain a comprehensive understanding of authoritative public databases, which can alleviate the difficulties in constructing their own datasets and speed up the process of scientific research. The purpose of this section is to provide a comprehensive overview of the available databases, which can be used to extract biomedical information, such as medicinal chemical structure, side effects, and anatomical therapeutic chemistry (ATC) codes. Popularly-used biological activity and compound data resources include DrugBank [6], SuperTarget [37], ChEBI [38], ZINC 15 [39], ChemSpider [40], ChEMBL [41], PubChem [42], KEGG [43], UniProtKB [44], Protein Data Banks (PDB) [45], HPRD [46], InterPro [47], STITCH [48], Binding MOAD [49] and BindingDB [50]. These popular public data sources can be classified into three categories, compound-centric data, protein-centric data, and compound-protein relation data, according to the biomedical entity they concern.

2.1 Compound-centric data

Compound-centric data contain several compound properties, such as the molecular structure, pharmacology and target information. In the past few decades, the pharmaceutical industry and academia have made great progress in drug discovery. With the release of large-scale compound data, research groups have collected the available information about compounds. In this sense, we list authoritative, freely available, and popularly used databases in Table 1. Although these compound-centric databases share common attributes, they also have different characteristics due to their different biomedical data sources and collection methods. For example, as a high-quality manually managed biomedical resource, DrugBank [6] covers almost all aspects of drugs, such as their identification (molecular formulas and SMILES strings [51]), pharmacology, clinical trials, pharmacoeconomic information, experimental properties, and comprehensive target protein information, which includes sequences, structures, and specific functions. PubChem [42] is also a large-scale compound and bioactivity database, and most of its data is derived from HTS experiments and updated daily. One of the most significant differences of ChEMBL [41] is that the data are mainly extracted and collated from the scientific literature, which makes ChEMBL [41] more reliable than the other databases. ChEBI [38] is a free database of molecular entities focusing on “small" compounds, such as any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, and conformer.

2.2 Protein-centric data

Proteins provide the basis of rational drug design. Available protein data that can be obtained include amino acid sequences, secondary structure information, 3D protein structures, Gene Ontology information, disease associations, and protein-protein interaction network information, which is useful for identifying potential target proteins. Table 2 shows the distribution of this information in several authoritative databases [44], [45], [46], [47]. Some databases have been cross-referenced and refined. For example, UniProtKB

TABLE 1
Public compound-centric databases

Compound	Molecular Structure		Line descriptor		Pharmacology			Targets		The number of compound
	2D	3D	SMILES	InChI	ATC code	Indication	Pathways	Proteins	Compound	
DrugBank [6]	√	√	√	√	√	√	√	√	√	13,874
SuperTarget [37]	√		√	√	√		√	√		195,770
ChEBI [38]	√	√	√	√						58,603
ZINC 15 [39]	√	√	√	√	√			√		980 million
ChemSpider [40]	√	√	√	√		√			√	67 million
ChEMBL [41]	√		√	√	√	√	√	√		1,961,462
PubChem [42]	√	√	√	√	√	√	√	√		108,824,004
KEGG [43]	√				√	√	√	√	√	18,749

[44] provides a comprehensive protein repository and is the primary source of protein information. The PDB database [45] provides protein structural information, and its 3D structural information is essential for drug discovery.

2.3 Compound-protein relation data

In addition to databases such as DrugBank [6], some databases specifically store compound-protein relations. Table 3 lists some free and prominent databases and provides the corresponding statistics. DrugBank [6] and STITCH [48] only provide compound-protein interactions, where existing interactions are denoted as 1s and otherwise are denoted as 0s. PubChem [42], ChEMBL [41], Binding MOAD [49], BindingDB [50] and SuperTarget [37] provide affinity data with quantitative measurements such as the IC50, EC50, K_i and K_d. These affinity data are obtained from various measurements, including enzyme inhibition activity, and enzyme kinetics, isothermal titration calorimetry (ITC), nuclear magnetic resonance (NMR) and radioactive ligand competition measurement.

Due to the heterogeneity among databases, researchers experience some difficulties in organizing available datasets from multiple data sources. The first problem is related to the mapping between databases. Although researchers have established the mapping relation between databases, some mapping information will inevitably be missing. Second, the different information stored in databases also presents a major challenge. For instance, for a given compound, there may only be SMILES strings [51] in some databases, with no stored graph structure representation. To this end, researchers have developed public packages, websites, and software to convert one representation into other representations; for example, through RDKit [52], the SMILES string can be converted into a graph structure. The connection and unification of databases will greatly increase the convenience of drug discovery and development.

Researchers usually collect and organize data from the above-mentioned databases and then publish some datasets, such as C.elegans [53] and KIBA [54], which can be directly used for research. For more details about these datasets, please refer to Section 1 in the supplementary material.

3 Feature Extraction Based on Deep Learning

As shown in Figure 2, compound, protein, and protein-ligand complex are three commonly used inputs for CPR prediction. A core challenge for CPR prediction is to effectively encode the inputs into task-related representations which reflect the intrinsic physical, chemical, and biological properties of the corresponding compounds or proteins. The main advantage of deep learning is that the feature extraction is performed in a fully automatic and data-driven manner without the elaborated feature engineering in advance. This section describes the deep learning-based feature extraction methods according to the different input formats, including manually crafted descriptors, sequence information, graph structure, and 3D structure.

3.1 Manually crafted descriptors

Descriptors are numerical vectors (eigenvectors) generated based on the structure and physicochemical properties of a molecule, where each position represents the presence or absence of certain attributes (bonds, substructures, functional groups, etc.). There are approximately 3,300 different types of compound descriptors recorded in the literature [55]. Several studies have explored the effect of fingerprint selection on difference tasks [56], [57], [58], [59]. Currently, the PubChem fingerprint [60] and Extended Connectivity Fingerprints (ECFP) [61] are considered as the most suitable fingerprints for CPR prediction. Moreover, a series of protein descriptors designed from different perspectives are listed in reference [35]. For example, Protein Sequence Composition (PSC) which has 8,420 descriptor values, is a widely used descriptor.

High-dimensional and sparse descriptor can easily trigger the curse of dimensionality, making models difficult to train. However, these problems can be solved by unsupervised learning methods. Table 4 shows examples of the dimensionality reduction of descriptors using unsupervised learning methods. Notably, Stacked Autoencoder (SAE) [62] is a typical unsupervised representation learning method, consisting of an encoder to learn the latent representation and a decoder to reconstruct the original input. Chan et al. [63] trained SAE by inputting multi-scale local descriptors of proteins [64] and PubChem fingerprints of compounds to get low-dimensional representations. Subsequently, Wang et al. [65] applied SAE to reconstruct the Position-specific Scoring Matrix (PSSM) [66] of protein. In addition to SAE, the restricted Boltzmann machine (RBM) [67] and deep belief network (DBN) [68] are introduced to solve the above problems. For example, Wang et al. [69] proposed a method based on RBM. A 199-dimensional protein binding site

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2021.3069040, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 2
Public protein-centric databases

Protein	Structure information			Targets		Gene Ontology	The number of protein/entry
	Amino acid sequences	Secondary structure	Tertiary structure	Disease	Protein		
UniProtKB [44]	√	√	√	√	√	√	Swiss-Prot:563,972; TrEMBL:209,157,139
PDB [45]	√	√	√				172,175
HPRD [46]	√			√	√		30,047
InterPro [47]	√	√				√	165 million

TABLE 3
The statistics of CPRs databases

Datasets	The number of compounds	The number of proteins	The number of relations	Relation type
DrugBank [6]	13,874	4,882	18,848	Binary
STITCH [48]	∼500,000	9,643,763	∼1.6 billion	Binary
PubChem [42]	108,824,004	99,361	273,300,136	Affinity
ChEMBL [41]	1,961,462	13,382	16,066,124	Affinity
Binding MOAD [49]	18,939	10,500	38,702	Affinity
BindingDB [50]	920,703	8,185	2,096,653	Affinity
SuperTarget [37]	195,770	6,219	332,828	Affinity

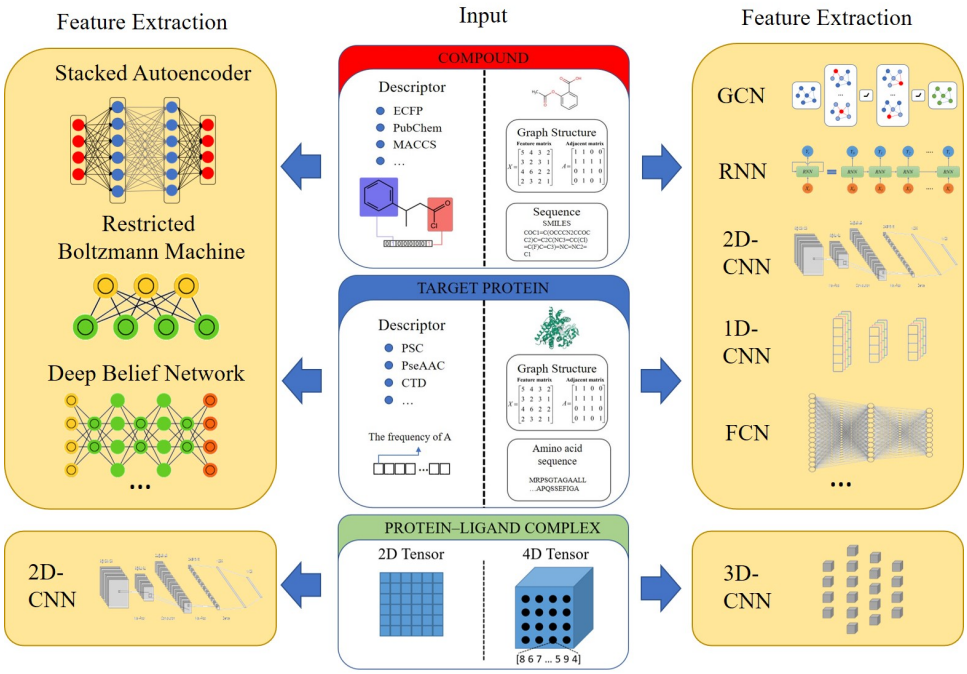


Fig. 2. Feature extraction based on deep learning.

vector and a 413-dimensional binary compound vector were used to train the RBM model. Then, only the encoded layer network was used as the prediction model. Furthermore, DeepDTI [70] utilizes DBN to reconstruct the feature vectors of compound-protein pairs. The model inputs are ECFP [61] of compounds and PSC of proteins. Additionally, although CGBVS-DNN [71] has similar architecture as DeepDTI, the inputs of the model are from PROFEAT [72] and Dragon [73]. Overall, these unsupervised models alleviate the dimensional disasters, and the reconstructed depth features can significantly improve the performance of subsequent classifiers.

3.2 Sequence information

Two types of sequence information are frequently-used in CPR prediction: SMILES string [51] and amino acid sequence. The first step in sequence feature extraction is encoding, including one-hot encoding and word embedding. The purpose of encoding is to give each ASCII character a unique semantic information. Some encoding methods including Mol2vec [74], BioVec [75], and SPVec [76] are inspired by Word2vec [77]. These methods try to learn

TABLE 4
Overview of papers using AutoEncoder to reduce descriptor dimensions in CPR prediction tasks

Citation	Methods	Input protein features (Dimension)	Input compound features (Dimension)	Out dimension
Chan et al. [63]	SAE	multiscale feature vectors(567)	The PubChem descriptor(881)	600
Wang et al. [65]	SAE	PSSM	The PubChem descriptor(881)	∼
Wang et al. [69]	RBM	Binding site information(199)	Structure information(413)	∼
DeepDTI [70]	DBN	PSC(8,420)	ECFP(6,144)	2,000
CGBVS-DNN [71]	DBN	Descriptor from PROFEAT(1,080)	Descriptor from Dragon(894)	2,000

the vector space, where the similar substructures (subsequences) are closely located.

1D-CNN followed by a global-max-pooling layer [78] can efficiently extract sequence semantic information [79], which is a widely used module in CPR prediction [80], [81], [82], [83], [84], [85]. In particular, when the convolution filter slides on the amino acid sequence, the different combination of amino acids in the sliding window are captured to simulate the local residue pattern [83], [84], [86], [87]. Some works also utilize 2D-CNN followed by max-pooling layer to extract features [88], [89]. The reasons why the structure of CNN followed by max-pooling is commonly used are that CNN has higher computational efficiency, and the max-pooling operation can ignore the length of the sequence greatly reducing the impact of different sequence lengths. Moreover, Recurrent neural network (RNN) is also a commonly used feature extractor [90]. To capture long-distance dependencies, Gao et al. [91] utilized an LSTM layer to directly process amino acid sequences. However, using a single feature extractor often has limitations. In order to capture the spatial correlation and timing information at the same time, researchers have proposed many improved modules, such as CNN–LSTM combination [92], [93], ConvLSTM [94], [95], [96], and Transformer [97]. It is worth mentioning that Transformer [97] is a state-of-the-art deep learning architecture. Unlike CNN and RNN, Transformer assumes that all elements in the input are connected. Transformer uses self-attention mechanism to dynamically calculate relationships between each element, which allows it to extract both local patterns and long-distance dependencies. Therefore, it is critical for sequence modeling, especially on proteins. Two amino acids that are far apart in the amino acid sequence may be located closely in actual 3D structure, thus, extraction of this spatial relationship is significantly important in CPR prediction.

3.3 Graph structure

Compared with SMILES string, the graph structure of compounds is a more natural representation, which takes atoms as nodes and bonds as edges. SMILES string can be converted into graph structure by using RDKit toolkit [52]. In general, graph representation of a protein is generated by considering amino acid residues as distinct nodes, and spatial proximities between residues are encoded into adjacency matrix. Particularly, Graph neural network (GNN) is a type of neural network that specializes in processing graph structures, and Graph convolutional network (GCN) [98] is a typical GNN model, which has been widely applied in computer-aided drug design. A GCN generally consists of two main operators: the aggregator and the update. The function of the aggregator operator is to gather information of the neighborhood atoms and provide the atom with its chemical environment. The update operator updates the atomic feature expression to improve the feature extraction capability of the graph convolution model. General GCN model treats all bonds equally, however, molecules are not just a simple collection of atoms. Same atoms often result in different properties depending on their local chemical environments, thus Graph attention network (GAT) introduces an attention mechanism [99] to obtain such structural information. Specifically, the relation between each pair of atoms is weighted by the attention weight. Moreover, Graph Isomorphism Network (GIN) [100] is an improved GCN model in which injective aggregation updates the parameters and performs the feature vector mapping to obtain maximum discriminative power. According to the results of [83], [85], [101], [102], GIN performs the best in the feature extraction of compounds, while GCN performs the best in the feature extraction of proteins. At present, there are many drug molecule graph convolution models [103], [104], [105], each of which has its own advantages. GCN [103] define a specific weight matrix for each node degree in each layer. Atoms are calculated with different weight matrices depending on its degree. To increase the efficiency of information transmission between distant atoms, Weave model [104] combines information from neighborhood atoms and their corresponding bonds. Additionally, Coley et al. [105] proposed a model in which the feature vectors of the neighborhood atoms concatenate with the features of their respective connecting bonds to form atom-bond feature vectors.

3.4 3D structure

The protein-ligand complex structure shows the most intuitive expression compound-protein combination in 3D space, which is a direct perspective for determining their functions and bioactivities. The 3D structure of the protein-ligand complex requires specific transformation and encoding before inputting into the deep learning model. In general, each complex is cropped into a cubic box, centered on the binding site. The default size of the box is usually set between 20 and 35 Å (1Å=10^{-10}m), and the default resolution is set to 1 Å [30], [106]. Each grid point stores the representation of a heavy atom at that site which holds basic structural features from atom types to more complex protein-compound descriptors. So, the complex can be represented as a 4D tensor. 3D-CNN is one of the mainstream feature extractors to explore the biochemical

interactions between atoms that occur locally in the protein-ligand complex.

The most representative model is AtomNet [30], which is the first deep learning system to process 3D structure of the protein-ligand complex. AtomNet utilizes four layers of 3D-CNNs to extract complex feature. However, the 3D grid representation has disadvantages such as sparseness and extremely high computational cost. Therefore, a series of improved methods have emerged [106], [107], [108]. Max-pooling is added after the 3D-CNN to reduce the amount of calculation [106], [107]. Li et al. [108] proposed a lightweight 3D-CNN framework, called DeepAtom, to automatically extract binding-related atomic interaction patterns. Compared with other CNN-based approaches, the design of lightweight model effectively improves the model representation capacity, even with limited available training data. Instead of directly processing 4D tensors, Zhang et al. [109] and Zheng et al. [110] mapped 3D interface spatial information into 2D space and applied stacked 2D-CNNs for feature extraction.

Inevitably, using deep learning methods to extract features has some disadvantages. First, the descriptors will lose their original biological meaning after deep learning feature reorganization, making them no longer biologically interpretable. Second, with the increasing complexity of structure and number of layers, the model tends to overfit. Hence, hyperparameter selection is particularly important. Judging from the experimental results of the methods [83], [85], [101], the feature combination of compound and protein is essential for determining the ability of the model. Simply improving the feature extraction of compound or protein cannot significantly improve the performance of the model. Therefore, researchers should be careful in the selection of inputs and corresponding feature extractors. In the next section, we will introduce some representative deep learning models in CPR prediction.

4 Computational methods based on deep learning

In recent years, deep learning algorithms have been widely used in CPR prediction. As shown in Figure 3, these models often consist of two modules. The first module is feature extraction to obtain features from input. The second module is a prediction based on these features. We classify these methods into six groups based on the feature extraction process and salient contribution: fully connected deep neural network-based models, convolution-based models, attention-based models, multi-modal-based methods, generative deep learning approaches, and other methods.

4.1 Fully connected deep neural network-based models

Fully connected deep neural network (FCDNN) is a type of artificial neural network in which all the nodes in one layer are connected to the neurons in the next layer while there is no connection between nodes on the same layer [29]. FCDNN-based models use descriptors as input and construct fully connected deep neural network to perform predictions.

Tian et al. [111] constructed a three-hidden-layer model with 2,000 hidden units in each layer. In this model, compounds are represented by the PubChem descriptors and proteins are encoded as a 5,523-dimensional vector by domain information. Du et al. [112] developed a neural network combining wide and deep models, giving the model some anti-overfitting capabilities. The inputs of the model are a series of physicochemical descriptors of the compound and protein computed by PyDPI [27]. Inspired by the intrinsic nonlinear patterns revealed by the LINCS project [113], Xie et al. [114] developed a four-layer model using the Z-score of genome-wide gene expression in level 4. They computed the Pearson correlation coefficient matrix for trails of a certain drug or a gene and applied the k-means algorithm to divide the drugs or genes into several clusters to build a 978-bit descriptor. Feng et al. [115] proposed a FCDNN-based framework with four layers, called PADME-ECFP, which combines the ECFP [61] of compounds and the PSC of proteins. Table 5 gives an overview of FCDNN-based methods. The "Structure" column in the table indicates the layer number of the FCDNN-based model and the neuron number in each layer. For example, "6,404-[2,000 $\times$ 3]-2" means that there are 6,404 neurons in the input layer, 3 hidden layers with 2,000 neurons in each layer, and 2 neurons in the output layer.

This type of model is the most basic one, so it is easy to train and implement. In the case of sufficient training data, the FCDNN-based model can make full use of descriptors to reveal the complex pattern. However, when there is not enough training data, prediction performance of this type of model would rely heavily on the descriptor selection and tend to overfit regardless of many techniques designed to prevent overfitting, such as dropout [116], batch normalization [117] and regularization.

4.2 Convolution-based models

With the continuous increasing amount of training data and the advancement of deep learning technology, researchers have developed a series of end-to-end methods to make more accurate CPR prediction. The most typical one is the convolution-based model. Convolution-based models receive two independent inputs, compound and protein. Each input has an independent feature extraction module, which belongs to the pairwise input neural networks (PINN).

Lee et al. [84] proposed a model, called DeepConv-DTI. DeepConv-DTI utilizes fully connected layer to process ECFP descriptor [61] of compounds and 1D-CNN block with global max-pooling to obtain protein features. The most significant contribution of this model is that multi-scale 1D-CNNs are applied to capture the local residual conformation patterns of different lengths of proteins. This structure is similar to Inception [118], which avoids the bias caused by a single-scale convolution kernel. After processing compounds and proteins, the feature vectors are concatenated and fed to fully connected layer to predict drug-target interactions.

DeepDTA [80] comprises two separate 1D-CNN blocks with with a global max-pooling layer. One block aims to learn representations from SMILES and the other aims to learn representations from protein sequences. For each 1D-CNN block, DeepDTA utilizes three consecutive 1D-CNN

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2021.3069040, IEEE/ACM Transactions on Computational Biology and Bioinformatics

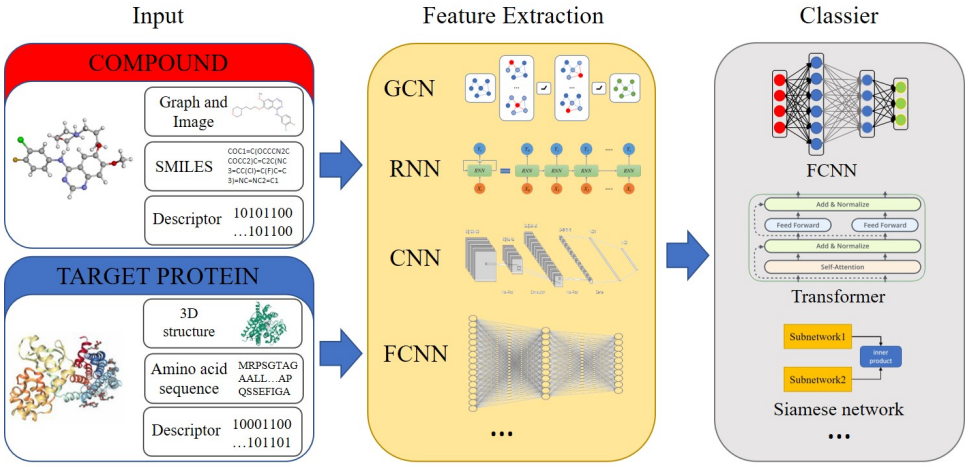


Fig. 3. Flowchart of a standard deep learning CPR prediction task.

TABLE 5
Comparison of different FCDNN-based methods

Citation	Input protein features(Dimension)	Input compound features(Dimension)	Structure	Task
Tian et al. [111]	Domain descriptor(5,523)	The PubChem descriptor(881)	6,404-[2,000×3]-2	CPI
Du et al. [112]	A series of descriptors(1819)	A series of descriptors(603)	2,422-[500×250]-20	CPI
Xie et al. [114]	Drug perturbations(978)	Gene knockout perturbations(978)	1,956-200-10-2	CPI
PADME-ECFP [115]	ECFP(1,024)	PSC(8,420)	9,444-[1,500×3]-1	CPA,CPI

layers, each containing 32, 64, and 96 filters respectively. This kind of stacked structure can extract wider semantic information, especially for proteins, which can cover relatively long amino acid chain. The last convolutional layer of each block is followed by a global-max-pooling layer, whose outputs are concatenated and fed into an FCDNN block to conduct prediction.

Based on DeepDTA, Nguyen et al. [85] proposed GraphDTA, in which GCNs are employed to capture the structural information of compounds. Then, outputs of GCN and protein vectors processed by 1D-CNN are concatenated together and fed into FCDNN. Similarly, Quan et al. [101] proposed GraphCPI that leverages GNNs to process compounds and 1D-CNN to learn protein vectors. It is worth mentioning that GraphCPI adapts a fixed-length N-gram splitting method to partition the sequence into meaningful "biological words" and introduces BioVec [75] for the embedding of protein sequences. Inspired by GraphDTA, Jiang et al. [102] proposed DGraphDTA, which encodes both compound and protein using GNNs. This method is the first attempt to construct the protein graph by utilizing the contact map generated by Pconsc4 [119]. After processed by GNNs, the molecular feature vectors of compounds and proteins are concatenated and fed into an FCDNN block. On the basis of GraphDTA, Wang et al. [83] introduced a new method, the dipeptide frequency of word frequency encoding, to overcome the lower expression ability of protein sequence based on one-hot encoding, improving the ability of 1D-CNN to extract protein features. Furthermore, Hu et al. [120] proposed a convolutional neural network model, named InterpretableCNN-CPR. First, the model applies the one-hot encoding to amino acid sequences and SMILES to form 2D input tensors. Then, the inputs are processed by two stacked 2D-CNNs with a pooling layer, of which the outputs are flattened, concatenated, and fed into FCDNN.

Gonczarek et al. [121] proposed a GCN-based model in which each compound or protein pocket is represented by a list of atom features and a list of atom connections. Given the input compound or protein data, the GCN and softmax operation are applied to get a fixed-size fingerprint which is further non-linearly transformed by FCDNN. The inner product of two fingerprints is then calculated to predict the interaction potential.

Torng et al. [122] developed a two-step graph-convolutional framework (Graph-CNN) for predicting CPIs. The first step is to build an unsupervised graph-autoencoder to learn features of protein pockets from a representative druggable pocket set. In the second step, they constructed two separate GCN modules for prediction. One GCN module extracted features of compounds, while the other GCN module initialized with learned weights from the first step learns protein features. Features learned from the GCN modules are then concatenated and fed into FCDNN to predict interactions.

To avoid the sparseness of the 3D grid representation, Lim et al. [123] introduced molecular graphs to efficiently express the exact distance between atoms of the complex. They embedded the 3D structure between the protein and compound atoms in two adjacency matrices. One of them represents purely covalent interactions and the other represents both covalent interactions and noncovalent intermolecular interactions. Then, two graph attention layers were employed to obtain two node feature matrices. One of them contains intermolecular information, while the other

does not. Using the matrix containing the intermolecular information to subtract the other matrix, the model learnes the difference between the binding structure and separated structure. This strategy allows the model to learn key features, rather than purely remembering certain patterns of ligand molecules. The feature vectors of ligand atoms are summed into one vector representing the protein-compound complex and fed into a FCDNN to make predictions.

Table 6 shows a comparison among the input types, feature extraction modules, and tasks of different convolution-based models. As presented in Table 6, CNN and GCN are mainstream feature extractors. CNN is known to capture important local patterns from the whole space; GCN regards the center nodes combined with neighborhood information as graph substructures. Therefore, convolution-based models contribute to present local patterns, which are similar to conceptions of functional groups in drugs and domains in proteins. Compared with RNN, the matrix operation mode of CNN or GCN can save a lot of training time, especially when dealing with relatively long protein sequences. The disadvantage of the convolution-based models is that they cannot capture long-distance dependencies. Although this problem can be solved by increasing the number of layers, it will also cause oversmoothness [124] (as nodes quickly become indistinguishable) in GCNs. Moreover, these methods make predictions based on the whole molecular structure of compounds and proteins, which would introduce noises and then affect the prediction performance. Also, the interpretability of convolution-based models has not drawn enough attention, since they do not present a straightforward path to indicate which substructures of compounds or subsequences of proteins contribute to the interactions.

4.3 Attention-based models

Although the above models have demonstrated good results, it is difficult to analyze them due to their black-box nature. Researchers have addressed this issue through the neural attention mechanism, determining which subsequences in a protein are binding sites and which atoms in a compound play a key role in binding with a protein. The attention mechanism also provides effective visualization by mapping regions with high weight values onto a known 3D protein-compound complex structure.

Tsubaki et al. [88] proposed a model called GNN-CPI, which applies GCN to learn descriptors from the graph of compounds. Additionally, it uses 2D-CNN to get the feature matrix of protein sequences, in which each row represents a subsequence. Moreover, GNN-CPI uses neural attention mechanisms [99] to assign high attention weights for protein fragments that are more inclined to bind to compounds. Then, the protein feature vector is computed as a weighted sum of the feature matrix according to the learned attention weights. This type of attention is called distinctive-attention [125]. Finally, the compound descriptor and protein feature vector are concatenated and fed into FCDNN for prediction.

Co-attention [125] is a more effective mechanism, which operates on two inputs at the same time and learns their attention weights jointly to capture interactions between each element of these inputs. InterpretableDTI [91] extracted feature matrices of protein and compound by LSTM and GCN [103], respectively. Through co-attention, InterpretableDTI generates an attention matrix, which presents the interaction between each amino acid and each atom. Then, feature vectors of compound and protein are generated according to this attention matrix. Finally, the attention-based vector representations are fed into a Siamese network [126] to perform prediction. In addition, Zhao et al. [82] proposed a model called AttentionDTA which applies co-attention on the basis of DeepDTA. Co-attention takes the outputs of CNN blocks as inputs and generates the attention matrix to measure the binding strength between each subsequence of SMILES strings and each fragment of amino acid sequences. Each row of outputs of CNN blocks multiplies by the attention coefficient before feeding into max-pooling. Abbasi et al. [92] proposed DeepCDA which combines CNN with LSTM to learn the occurrence patterns of local substructures through protein and compound sequences. The co-attention mechanism in DeepCDA is different from those in InterpretableDTI and AttentionDTA. This co-attention mechanism adjusts the binding affinities between substructures by the binding strength between the whole compound and the whole protein. Only when the two parts achieve high values, the strong binding relationships between the substructures will be presented, reflecting the idea of global control and local adjustment.

For overcoming the deficiency of CNN and RNN, Zheng et al. [127] used BiLSTM with a multi-head self-attention mechanism to process SMILES strings in order to gain some dependency between adjacent tokens. CNN module based on stacked residual blocks [128] and sequential self-attention blocks is used to code protein distance maps, which contributes to the binding site finding. In addition, Shin et al. [129] proposed Molecule Transformer DTI (MT-DTI). They used the self-attention mechanism to learn the high-dimensional structure of SMILES. Then they concatenated compound feature with protein feature extracted by CNNS and employed FCDNN for prediction.

Karimi et al. [130] designed a novel and interpretable deep learning framework called DeepAffinity. The first part of this framework is an AutoEncoder which is composed of bidirectional gated recurrent units (BiGRUs). It processes a large amount of unlabeled data to obtain representations, which captures long-term and nonlinear dependencies among residues/atoms in proteins/compounds. The second part is a novel deep learning model unifying BiGRUs and CNNs, which is trained from end to end using labeled data. The BiGRUs in the second part are initialized using the weights of BiGRUs trained in the first part. Furthermore, the method introduces both separate and joint attention mechanisms to interpret predictions by isolating the main contributors of compounds or proteins, which is exploited for predicting binding sites and providing interpretability.

Huang et al. [131] proposed a Molecular Interaction Transformer (MolTrans) to address the limitations of interaction modeling module. Given the input of SMILES or amino acid sequence, MolTrans extracts the sub-structure embedding of compound/protein sequence from Frequent Consecutive Sub-sequence Algorithm (FCS). Then, the sub-structure embeddings are fed into the compound/protein transformer encoder to obtain the augmented contextual representations. After that, an interaction map measuring

TABLE 6
Comparison of different convolution-based models

Citation	Inputs		Feature extractor		Task
	Protein	Compound	Protein	Compound	
DeepConv-DTI [84]	Amino acid sequence	ECFP	1D-CNN	FCDNN	CPI
DeepDTA [80]	Amino acid sequence	SMILES	1D-CNN	1D-CNN	CPI,CPA
GraphDTA [85]	Amino acid sequence	Graph structure	1D-CNN	GCN,GAT,GIN and GAT_GCN	CPA
DGraphDTA [102]	Graph structure	Graph structure	GNN	GNN	CPA
GraphDTI [101]	Amino acid sequence	Graph structure	1D-CNN	GCN,GAT and GIN	CPI
Wang et al. [83]	Amino acid sequence	Graph structure	1D-CNN	GAT_GCN	CPA
Gonczarek et al. [121]	Graph structure	Graph structure	GCN	GCN	CPI
InterpretableCNN-CPR [120]	Amino acid sequence	SMILES	2D-CNN	2D-CNN	CPI,CPA
Torng et al. [122]	Graph structure	Graph structure	GCN	GCN	CPI
Lim et al. [123]	Graph structure	Graph structure	GAT	GAT	CPI

interaction intensity among sub-structures is generated and fed into a 2D-CNN model to output the DTI probability.

Inspired by the great ability of capturing features between two sequences in Transformer, Chen et al. [132] found that the protein sequence representation and atom representation can be regarded as two sentences that fit the Transformer architecture. And the interactive features can be learned through the decoder of Transformer. Chen et al. modified the Transformer architecture by remaining the decoder of the Transformer and modifying its encoder and final linear layers to propose a novel transformer neural network, called TransformerCPI. The potential advantage of TransformerCPI is its inputs can be any length or size. TransformerCPI maintains the integrity of the input data, comparing to models that only receive fixed-length inputs.

Table 7 shows a comparison of different attention-based methods. Attention mechanism can significantly improve the capability of the model at the cost of adding fewer parameters. These models provide many visual examples to indicate that they can better predict binding site information, demonstrating the biological significance of the model. However, this type of model has an obvious limitation, which is it consumes a lot of memory especially when using a Transformer-like attention mechanism on proteins. Recently, a number of improved versions of Transformer have been proposed, many of which make improvements around computational and memory efficiency [133]. These models are expected to shine in CPR prediction tasks in near future.

4.4 Multi-modal-based methods

It is difficult to accomplish computer-aided CPR prediction by inputting single modal data. As described in Section 3, each original input corresponds to a unique feature extraction process. Different deep learning feature extractors focus on discovering specific patterns in the original input. CNNs can effectively capture invariant local patterns while RNNs can capture the long-term dependencies efficiently. Since the choice of descriptors and feature extractors markedly influence the model performance, researchers proposed multi-modal-based methods which take representations of multiple compounds and proteins as input to alleviate the influence of the feature extraction.

Öztürk et al. proposed WideDTA [81] based on DeepDTA. Besides inherently utilizing protein sequences and SMILES strings, WideDTA introduces the concept of "word" to improve the expressiveness of inputs. A "word" of a protein sequence corresponds to a three-residue subsequence, whereas a "word" of a ligand is equal to an 8-character subsequence. More importantly, two textual information sources were introduced to provide valuable clues about the interaction specificity, which are protein domains and motifs (PDM) of proteins from the database PROSITE [134] and ligand maximum common substructures (LMCS) [135] of compounds. The method applies a 1D-CNN block with a global-max-pooling layer for each text-based information module. Features extracted from these blocks are concatenated and fed into FCDNN to obtain the affinity values.

Methods which only utilize SMILES strings loss the molecular structure information, so, Mohammed et al. [95] proposed DeepH-DTA, which uses both molecular graphic structure and SMILES strings of the compounds as inputs. In DeepH-DTA, protein string is fed into DenseNet-like CNNs [136] with SE operation [137]. Simultaneously, a heterogeneous graph network [138] learns the topological representation of compounds, and a BiConvLSTM architecture learns the features of SMILES strings. After that, three extracted representations are concatenated and fed into FCDNN. Similarly, Li et al. [139] proposed DeepGS, which utilizes deep neural networks to extract local chemical context from protein sequences and SMILES sequences, as well as the topological structure information of compounds.

Furthermore, Lee et al. [140] proposed Multi-channel PINN (MCPINN). The compound inputs include SMILES strings, ECFPs [61], and vectors embedded by Mol2vec [74], and the protein inputs are amino acid sequences and vectors embedded by BioVec [75]. Each input channel layer extracts different levels of features, and then these features are concatenated and fed into FCDNN to predict the CPIs. In addition, Brighter et al. [141] proposed Integrated Views Predictive Generative Adversarial Network (IVPGAN), which forms a joint representation by concatenating the predefined molecular descriptor and the outputs of the parameterized descriptor learning function.

Different from models that directly concatenate features from different sources, Brighter et al. [142] proposed Joint View self-Attention (JoVA) approach to learn rich repre-

TABLE 7
Comparison of different Attention-based methods

Citation	Protein		Compound		Attention type	Task
	Input	Extractor	Input	Extractor		
GNN-CPI [88]	Amino acid sequence	CNN	Graph structure	GCN	Distinctive-attention	CPI
InterpretableDTI [91]	Amino acid sequence	LSTM	Graph structure	GCN	Co-attention	CPI
VQA [127]	2D pairwise distance map	2D-CNN	SMILES	LSTM	Self-attention	CPI
DeepAffinity [130]	Amino acid sequence	GRU,AE	SMILES	GRU,AE	Separate and joint attention	CPA
AttentionDTA [82]	Amino acid sequence	1D-CNN	SMILES	1D-CNN	Co-attention	CPA
MolTrans [131]	Amino acid sequence	Transformer	SMILES	Transformer	Self-attention	CPI
MT-DTI [129]	Amino acid sequence	CNN	SMILES	Transformer	Self-attention	CPA
TransformerCPI [132]	Amino acid sequence	CNN	Graph structure	GCN	Self-attention	CPI
DeepCDA [92]	Amino acid sequence	CNN-LSTM	SMILES	CNN-LSTM	Two-sided attention	CPA

sentations from multiple unimodal representations of compound and target. JoVA has the capacity to leverage multiple existing end-to-end and predefined descriptors by introducing the multi-view self-attention-based architecture. Intuitively, individual representation of the compounds and proteins are merged to form a fully connected representation by self-attention mechanism. Each sub-representation updates itself by considering all sub-representations.

In some cases, predetermined descriptors and end-to-end feature learning complement each other. Multi-modal-based methods fully exploit the advantages of different features and reduce the risk and cost of feature selection. However, when comparing to the single-channel models, multi-channel models do not guarantee better performance. The performance of multi-channel models depends on the combination of compound and protein features [140]. Therefore, the representation needs to be carefully selected for better prediction accuracy.

4.5 Generative deep learning approaches

Generative deep learning approaches present a fresh approach to *de novo* drug design, as they help researchers with the process of narrowing down the chemical search space and designing lead compounds binding to target proteins directly.

Gómez-Bombarelli et al. [143] proposed a model to generate SMILES strings by Variational Autoencoder (VAE). One trick in this paper is using Gaussian process in the latent space (which is a continuous space) to generate compounds with desired properties. In addition, Skalicet al. [144] proposed a generative adversarial network based on 3D-CNN to generate diverse 3D ligand shapes complementary to target proteins. Subsequently, the generated molecule shapes can be decoded into SMILES strings using a shape-captioning network. However, the problem is that there is no one-one correspondence between SMILES strings and molecules, so the generated SMILES strings normally fail to correspond to the molecules we need. Therefore, Kusner et al. [145] proposed Grammar VAE which converts the SMILES strings to the parse tree to generate more syntactically valid molecules.

However, to generate more valid molecules, SMILES-based models are required to learn SMILES grammar and atom ordering, which increases the burden of the model. In order to alleviate the burden of the model, molecular graph is a more preferable representation compared with the SMILES string. Li et al. [146] proposed a graph-based generator that is more suitable for molecules, which generates graphs by iteratively refining its intermediate structure. In addition, the conditional generative model is used to generate molecules with specific requirements. This model has advantages in generating drug candidates with good synthetic accessibility and targeting specific proteins.

The greatest potential of generative models lies in the ability to generate ligands for novel special targets, such as the SARS-CoV-2 viral proteins in the recent COVID-19 pandemic. Chenthamarakshan et al. [147] proposed an end-to-end *de novo* drug design framework, called Controlled Generation of Molecules (CogMol). CogMol has the ability to generate small drug-like compounds targeting novel viral proteins with high affinity and off-target selectivity. In CogMol, a VAE is trained under unsupervised learning to obtain the disentangled latent space of the molecules with SMILES as input. Moreover, to estimate the target specificity and the off-target selectivity, a CPA regressor is trained on the outputs of VAEs. Under the guidance of a set of attributes (affinity and selectivity) predictors, an efficient sampling scheme is introduced to generate molecules with desired attributes from the model of the VAE latent space.

De novo design of drug-like moleculars using generative models is a promising tool to accelerate drug discovery process when human facing the challenge of emerging diseases. Although some successes have been achieved, this research direction is still full of opportunities and challenges. The synthetic feasibility of the generated structures turns out to be a huge obstacle in connecting generative models with actual production. At the current stage, researchers focus on the generation of SMILES strings and descriptors. However, it is more promising to develop new methods that directly generate the 3D structures. Furthermore, the macromolecule generation will also be a encouraging future direction.

4.6 Other methods

We will show some other methods in this section, which contains some unique insights in completing CPR prediction from different perspectives.

DTI-RCNN [93] integrates LSTM networks with CNN to further improve CPI prediction accuracy using drug perturbation and gene knockout perturbation data. First, the model extracts potential semantic information between gene data and drug data via an LSTM network and constructs a CNN to extract the local knowledge from the LSTM outputs. After that, FCDNN is employed to conduct final prediction.

To alleviate the problem of lacking labeled data, Zhao et al. [148] proposed a semi-supervised generative adversarial network, called GANsDTA. GANsDTA comprises two parts: GAN for feature extraction and FCDNN for prediction. The semi-supervised mechanism allows GANsDTA to effectively learn useful features from both labeled and unlabeled data.

Hussein et al. [149] proposed Rosetta Energy Network (RosENet), a 3D-CNN model based on ResNet architecture [150]. RosENet combines voxelized molecular mechanics energies and molecular descriptors. The molecular mechanics have proven to be a valuable method for understanding the function and dynamics in biomolecular systems and provide critical information for estimating the binding affinity. The residual networks in Resnet have the capacity to learn variations between the input and output. It contributes to learn the energy distribution of complexes instead of simply the absolute energies.

In the field of machine learning, classification models tend to be more biased while regression models tend to overfit the training data to obtain large variance. Weng et al. [89] explored to balance the bias and variance by a multi-task learning framework, named Multitask-CPR, that can both fit the affinity curve and decide the correct boundary between positive and negative interactions. this model learns protein and compound feature vectors through 2D-CNN, max-pooling, and co-attention mechanism. Based on the shared feature representations, they computed the cosine similarity between two feature vectors in the prediction layer to measure CPRs.

It is worth mentioning that Li et al. [151] developed a multi-task neural network, MONN, to learn both pairwise non-covalent interactions and binding affinities between compounds and proteins. MONN takes graph structures of compounds and amino acid sequences of proteins as inputs, which are processed by GCNs and CNNs, respectively. In particular, MONN integrates graph warp module [152] into traditional GCN [153] to learn both local features from neighbors for individual atoms and global features for the whole compound. MONN captures the non-covalent interactions between atoms of a compound and residues of a protein under the supervision of binding site labels extracted from available high-quality 3D compound-protein complex structures. In addition, the pairwise non-covalent interaction prediction results are further utilized to benefit the prediction performance of binding affinities. After multi-task training, MONN can be used to infer the interaction sites and CPIs at the same time.

Different from the above methods, Rifaioglu et al. [154] proposed a large-scale CPI prediction system, called DEEPScreen. DEEPScreen takes compound images as input and constructs a convolutional network classifier to predict small-molecule ligands for a unique target protein.

5 Experiment and Comparison

To follow the latest research and perform a fair comparison, we select several state-of-the-art and easy-to-use (e.g., providing the source codes) methods which have been proposed in the past three years and conduct a comparison analysis. Some methods [91], [112], [130], [149], [149], [151] were excluded from comparison, because their inputs are so special that they are not available in our benchmark datasets. Methods [143], [144], [145], [146], [147] based on generative model are also excluded, as they cannot directly explore the interaction between a compound and a protein. These selected models adopted the same hyperparameters as those in original papers. For more details about the structure of these models, please refer to Section 2 of the supplementary material. We conduct 5-fold cross-validation comparison for each task. In each fold, $4/5$ of the relations are used as the training set, while the remaining $1/5$ of the relations are treated as the test set. We randomly take $1/5$ from the training set as the validation set to perform early stopping. We assume that if the model's performance on the validation set does not improve within 20 epochs, the model is considered well trained.

5.1 CPI prediction task

We choose nine methods for CPI prediction tasks, which are PADME-ECFP [115], GraphCPI [101], DeepConv-DTI [84], DeepDTA [80], InterpretableCNN-CPR [120], Multitask-CPR [89], TransformerCPI [132], GNN-CPI [88], and MolTrans [131]. DeepDTA originally is an CPR prediction model that is now used for CPI prediction by setting a threshold [80], [131]. We change the last layer of DeepDTA to make it suitable for the binary classification problem. Overviews of these models are shown in Figure S.1 to S.9 of supplementary material.

5.1.1 Datasets and Evaluation metrics

DrugBank [6] is the most commonly used dataset for CPI prediction. We use the version 5.1.5 (released on 2020-01-03) as the dataset. After brushing, there were 6,655 drugs, 4,294 proteins, and 17,511 positive samples in the dataset. We randomly select negative samples from the unlabeled compound-protein pairs which has an equal number as positive samples. We use the area under the curve (AUC), precision, accuracy, and recall as metrics to measure model performances.

5.1.2 Experiment results

The comparison of selected methods is shown in Table 8. The best results are highlighted in bold font. Overall, DeepConv-DTI [84] yields the best performance. Compared with other models, the biggest advantage of DeepConv-DTI is that it utilizes the multi-scale-based convolution structure in CNN module. This helps the model to avoid the deviation caused by setting the size of a single convolution kernel, thus helping CNN to learn more protein residues patterns. Then, Moltrans [131] achieved slightly worse performance than DeepConv-CPI, because it truncates proteins, which makes it lose useful information. Setting a longer protein length may be one of the solutions. It is worth mentioning that as a CPA prediction model, DeepDTA [80] achieves

competitive results in CPI prediction, which may stimulate researcher to focus on CPA prediction. InterpretableCNN-CPR [120] is similar to DeepDTA [80], but it performs slightly worse than DeepDTA. The biggest difference between them is that in InterpretableCNN-CPR [120], each layer of CNN is followed by a max-pooling layer, while in DeepDTA [80], only the last layer of CNN is followed by a global-max-pooling layer. The previous module can reduce computational requirements, but it will lose a lot of sequence information, thereby affecting the performance of the model. In GraphCPI [101], GraphDTA/GIN achieved the overall best results. The most likely reason is that GIN controls the update of atomic features through a learnable parameter in the aggregator stage, which makes the model more flexible in extracting the chemical environment information around the atoms. From the comparison of GraphCPI [101], TransformerCPI [132] and GNN-CPI [88], it can be concluded that attention mechanism indeed contributes to improve the performance of the model in the case of similar feature extraction.

5.2 CPA prediction task

We choose ten methods for CPA prediction tasks, which are PADME-ECFP [115], DeepDTA [80], WideDTA [81], GraphDTA [85], AttentionDTA [82], DeepGS [139], InterpretableCNN-CPR [120], Multitask-CPR [89], DeepCDA [92], and DGraphDTA [102]. Since the lack of PDM and LMCS, we use the WideDTA without PDM and LMCS here. Overviews of these models are shown in Figure S.10 to S.19 of supplementary material.

5.2.1 Datasets and Evaluation metrics

We evaluate the above models on KIBA [54] datasets, which is previously used as a benchmark dataset for binding affinity prediction. We use the mean squared error (MSE) and concordance index (CI) [155] as our metrics.

5.2.2 Experiment results

The results on KIBA dataset are given in Table 9, from which we can see that DGraphDTA [102] achieves the best performance. Besides AttentionDTA [82], GraphDTA [85] also surpasses other sequence-based models. This indicates that, graph structure-based methods are more competitive in CPA prediction tasks compared with sequence-based models. After all, the grammatical rules of SMILES are different from natural language which makes it hard to extract using a simple CNN module. Thus, the development of GNN will become a hot spot in CPA prediction research in the future. In addition, attention-based models, AttentionDTA [82] and DeepCDA [92], achieved better performance in sequence-based models [80], [81], [82], [89], [92], [120]. The excellent performance of AttentionDTA [82] indicates that utilizing attention mechanism to dynamically adjust the expression level of substructures in compounds and proteins is beneficial. Without introducing attention mechanism, DeepCDA based on CNN-LSTM achieves better results than those CNN-based models (DeepDTA [80], WideDTA [81] and InterpretableCNN-CPR [120]), which indicates that combined feature extractor performs preferable to single feature extractor. By comparing WideDTA and DeepDTA, we found that the word-based method (WideDTA) obtains a better performance than the character-based method (DeepDTA). Therefore, through the introduction of some natural language processing techniques, 1D-CNN's ability can be effectively improved. It is worth noting that the PADME-ECFP [115] has achieved eye-catching performance in CPI prediction, but less in CPA prediction. This indicates that in the CPA prediction task, the end-to-end feature extraction is more critical than the manually crafted descriptors.

6 Discussion

In this article, we present a review of the deep learning applications in CPR prediction. Primarily, we take a detailed investigation into different feature extraction processes and prediction models and summarize the architectures concisely. Then, we select several state-of-the-art methods and empirically compare these methods to show their performances and specialties. The development of deep learning in CPR prediction has exhibited a certain trend, which is that instead of relying on feature engineering, researchers have begun processing raw data and developing end-to-end models. In addition, researchers have gradually paid attention to the interpretability of models while improving their performance. Finally, we give some challenges and possible directions for further improvement of CPR prediction.

6.1 Fixed input size

Among current deep learning-based CPR prediction methods, most methods only receive fixed-size/length compound or protein input. For example, FCDNN-based models only receive fixed-length molecular descriptors, otherwise the models cannot be tested. Although the convolution-based models in theory can receive input of any size through global-pooling or sum-pooling layer, researchers still fixed the input size in order to accelerate the training speed and escape from local minima. Researchers need to record the length of all samples in the dataset to ensure that the length we set can cover enough information. Sequences that are longer than the length we set are truncated, whereas shorter sequences are zero-padded. Lee et al. [84] explored the impact of the maximum length setting on the DeepConv-DTI performance. Their results show that when the setting length is larger than the average length of sequences in the dataset, the performance of DeepConv-DTI is robust to the setting length. This is also confirmed by the experimental result of DeepGS [139]. However, this conclusion is not reliable for attention-based models. Truncation would result in the loss of input information and affect the distribution of attention weight, which greatly affect the ability of attention-based methods to accurately obtain local interactions between compounds and proteins.

6.2 Absence of reliable negatives

A common problem in CPI prediction is the lack of reliable negative samples. Only the accurate positive samples are kept in databases, and the rest are unmarked relations, including undiscovered positive and true negative relations. A common approach is to obtain negative samples from datasets that record affinity values. In most computational

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2021.3069040, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 8
Comparison on DrugBank under 5-fold cross-validation on CPI prediction task

Methods	AUC(std)	Precision(std)	Accuracy(std)	Recall(std)
PADME-ECFP [115]	0.8537(0.0000)	0.7901(0.0001)	0.7903(0.0000)	0.7907(0.0001)
DeepConv-DTI [84]	**0.8624**(0.0001)	0.7864(0.0002)	0.7909(0.0001)	**0.7995**(0.0002)
DeepDTA [80]	0.8594(0.0002)	0.7825(0.0001)	**0.7971**(0.0001)	0.7811(0.0001)
GNN-CPI [88]	0.8389(0.0000)	0.7704(0.0000)	0.7665(0.0000)	0.7595(0.0001)
GraphCPI/GCN [101]	0.7806(0.0001)	0.7306(0.0001)	0.7420(0.0001)	0.7643(0.0002)
GraphCPI/GAT [101]	0.7789(0.0000)	0.7173(0.0000)	0.7390(0.0000)	0.7865(0.0001)
GraphCPI/GIN [101]	0.8200(0.0000)	0.7493(0.0001)	0.7604(0.0000)	0.7808(0.0000)
GraphCPI/GAT_GCN [101]	0.8056(0.0000)	0.7492(0.0000)	0.7615(0.0000)	0.7844(0.0000)
InterpretableCNN-CPR [120]	0.8541(0.0000)	0.7745(0.0002)	0.7784(0.0000)	0.7860(0.0002)
Multitask-CPR [89]	0.8243(0.0000)	**0.8286**(0.0001)	0.7569(0.0000)	0.6480(0.0001)
TransformerCPI [132]	0.8416(0.0001)	0.7665(0.0001)	0.7753(0.0002)	0.7762(0.0001)
MolTrans [131]	0.8602(0.0000)	0.8104(0.0003)	0.7860(0.0001)	0.7477(0.0013)

TABLE 9
Comparison on KIBA under 5-fold cross-validation on CPAs task

Methods	MSE(std)	CI(std)
PADME-ECFP [115]	0.2113(0.0000)	0.8393(0.0000)
DeepDTA [80]	0.1970(0.0055)	0.8629(0.0020)
WideDTA [81]	0.1835(0.0005)	0.8705(0.0002)
GraphDTA/GCN [85]	0.1855(0.0003)	0.8680(0.0030)
GraphDTA/GAT [85]	0.1800(0.0063)	0.8733(0.0032)
GraphDTA/GIN [85]	0.1704(0.0004)	0.8825(0.0032)
GraphDTA/GAT_GCN [85]	0.1735(0.0053)	0.8764(0.0054)
AttentionDTA [82]	0.1621(0.0025)	0.8858(0.0040)
DeepGS [139]	0.1968(0.0005)	0.8680(0.0005)
InterpretableCNN-CPR [120]	0.1913(0.0008)	0.8700(0.0002)
Multitask-CPR [89]	0.1943(0.0005)	0.8680(0.0008)
DeepCDA [92]	0.1804(0.0001)	0.8725(0.0002)
DeepCDA without attention [92]	0.1899(0.0001)	0.8672(0.0003)
DGraphDTA [102]	**0.1565**(0.0020)	**0.8905**(0.0010)

studies, IC50 values below 10 μM are considered to be active [156]. However, during drug development, most of the drug candidates through the lead discovery and optimization process have activity values below micromolar concentrations [35]. The reason for relaxing threshold to 10 μM is to obtain more training data. However, marking cases with non-activity as active brings noise to the training data, making the classifier less effective. Thus, optimizing appropriate thresholds is a challenge. The other approach is to rely on certain criteria to select reliable negative samples. Liu et al. [53] developed a screening framework based on hypotheses that compounds would interact with proteins which are similar to known target proteins. In addition, to identify more reliable negative samples, they performed statistical tests to determine whether the variance of similarity between targets (or ligands) for each compound (or protein) was greater than the variance of the whole negative samples or not.

6.3 Affinity task

Binary datasets often ignore many important aspects of compound-protein interactions, including their dose-dependent and quantitative affinities. Some recently published data contains binding affinity information, so the prediction of relation can be viewed as a regression problem of binding affinity rather than a classification problem. DREAM Challenge (http://dreamchallenges.org/) ran a challenge called IDG-DREAM Drug-Kinase Binding Prediction Challenge on October 1, 2018. This challenge seeks to evaluate the power of statistical and machine learning models as a systematic and cost-effective means for catalyzing CPA mapping efforts by prioritizing the most potent interactions for further experimental evaluation. Participants are requested to predict currently unpublished K_d values for a given set of compound-kinase pairs. The challenge makes use of an open-data web-platform, DrugTargetCommons (https://drugtargetcommons.fimm.fi), which consists of CPR information including 1,746,997 compounds and 13,023 targets. The competition attracted more than 300 registration applications and 45 teams in the first round. Overall, regression models for binding affinity prediction will play a key role in predicting compound-protein relations in the future.

6.4 Representation learning from unlabeled data

By comparing the number of drugs in the DrugBank with the number of compounds in the Pubchem, we can find that when training with database as large as DrugBank, we rarely exploit the currently available large-scale protein and compound data to predict CPRs. Improving the representations of compounds and proteins from a large amount of unlabeled data has gradually become an essential step to improve CPR prediction models. The representation learning in DeepCPI [157] obtains the implicit expressive low-dimensional features of compounds and proteins from a massive amount of unlabeled data. Zhang et al. [76] proposed an unsupervised representation learning method, SPVec, which is inspired by Word2vec, to automatically represents SMILES strings and protein sequences into continuous, information-rich, and lower-dimensional vectors. Moreover, methods proposed in [130], [131] also benefit from a large number of unlabeled samples.

6.5 End-to-end learning under supervision in the biochemical background

In current deep learning-based approaches, there exists end-to-end models and non-end-to-end models. End-to-end

models extract high-level features from raw data without human intervention, however, traditional feature engineering in a non-end-to-end model extract feature descriptors with certain chemical and biological backgrounds. The attempt at developing end-to-end learning under supervision with biochemical background is a promising research direction.

6.6 Neural attention mechanism

Since the binding of compounds and proteins occurs locally, using the overall feature of compound or protein for prediction often introduces noise. Recognizing binding sites is critical for the accurate prediction of CPRs. Neural attention mechanism [99] allows the model to consider which subsequences in proteins and compounds are important for predicting CPRs, which is proven by [82], [88], [91], [127], [130]. In addition, by using the obtained weights, the neural attention mechanism provides a good visualization that makes the model easier to analyze. Therefore, in the future, attention mechanism is an important tool to improve CPR prediction task and provide biological insights for researchers.

6.7 Interpretability

Due to the black-box nature of deep learning, the interpretation of deep learning models is hard to achieve. Researchers developed some indirect or direct methods to perform model analysis. In order to map filters to chemical functions, Wallach et al. [30] applied convolution kernels to input data and examine the location where they are maximally activated, which demonstrates the ability of the AtomNet to learn complex chemical features from simpler ones. Goh et al. [158] developed an explanation mask that find the most important characters used in the task to demonstrate that neural networks can learn the technically accurate chemical concept, making interpretable deep neural networks a useful tool for chemical industry. With the successful application of attention mechanism in the field of natural language processing and computer vision, attention-based CPI prediction methods have been extensively studied in the past two years. Attention mechanism assigns large weights to protein fragments or drug atoms that play an important role in the inference process. These attention weights can be intuitively compared with the binding site information to achieve the purpose of model interpretability. Although researches have demonstrated the ability of attention mechanism to capture the binding sites through some cases, the experiments by Li et al. [151] show that attention-based approaches have difficulty in automatically capturing the local non-covalent interactions between compounds and proteins accurately without additional supervision. On the issue of CPR prediction, how to define and evaluate the interpretability of the model is worth considering in depth. This will make outstanding contributions to deep learning-based CPR prediction and help to facilitate the process from the drug research and development stage to actual production applications.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grants (Nos. U1909208, 61772557, 61732009), 111 Project (No. B18059), Hunan Provincial Science and Technology Program (No. 2018wk4001).

REFERENCES

[1] M. Yang, L. Huang, Y. Xu, C. Lu, and J. Wang, "Heterogeneous graph inference with matrix completion for computational drug repositioning," *Bioinformatics*, 12 2020, btaa1024. [Online]. Available: https://doi.org/10.1093/bioinformatics/btaa1024

[2] H. Luo, M. Li, M. Yang, F.-X. Wu, Y. Li, and J. Wang, "Biomedical data and computational models for drug repositioning: a comprehensive review," *Briefings in Bioinformatics*, 02 2020, bbz176. [Online]. Available: https://doi.org/10.1093/bib/bbz176

[3] X. Chen, C. C. Yan, X. Zhang, X. Zhang, F. Dai, J. Yin, and Y. Zhang, "Drug–target interaction prediction: databases, web servers and computational models," *Briefings in bioinformatics*, vol. 17, no. 4, pp. 696–712, 2016.

[4] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nature reviews Drug discovery*, vol. 3, no. 8, pp. 673–683, 2004.

[5] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda *et al.*, "DrugBank 5.0: a major update to the DrugBank database for 2018," *Nucleic Acids Research*, vol. 46, 2018.

[6] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, "DrugBank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic acids research*, vol. 34, no. suppl_1, pp. D668–D672, 2006.

[7] C. Finan, A. Gaulton, F. A. Kruger, R. T. Lumbers, T. Shah, J. Engmann, L. Galver, R. Kelley, A. Karlsson, R. Santos *et al.*, "The druggable genome and support for target identification and validation in drug development," *Science Translational Medicine*, vol. 9, no. 383, 2017.

[8] G. Wu, M. Yang, Y. Li, and J. Wang, "De novo Prediction of Drug-Target Interaction via Laplacian Regularized Schatten-p Norm Minimization," in *International Symposium on Bioinformatics Research and Applications*. Springer, 2020, pp. 154–165.

[9] H. Luo, J. Wang, C. Yan, M. Li, F. Wu, and Y. Pan, "A novel drug repositioning approach based on collaborative metric learning," *IEEE/ACM transactions on computational biology and bioinformatics*, 2019.

[10] C. Yan, J. Wang, W. Lan, F.-X. Wu, and Y. Pan, "SDTRLS: Predicting drug-target interactions for complex diseases based on chemical substructures," *Complexity*, vol. 2017, 2017.

[11] L. Yao, J. A. Evans, and A. Rzhetsky, "Novel opportunities for computational biology and sociology in drug discovery: corrected paper," *Trends in biotechnology*, vol. 28, no. 4, pp. 161–170, 2010.

[12] M. Yang, G. Wu, Q. Zhao, Y. Li, and J. Wang, "Computational drug repositioning based on multi-similarities bilinear matrix factorization," *Briefings in Bioinformatics*, 11 2020, bbaa267. [Online]. Available: https://doi.org/10.1093/bib/bbaa267

[13] M. Yang, H. Luo, Y. Li, and J. Wang, "Drug repositioning based on bounded nuclear norm regularization," *Bioinformatics*, vol. 35, no. 14, pp. i455–i463, 07 2019. [Online]. Available: https://doi.org/10.1093/bioinformatics/btz331

[14] A. L. Hopkins, "Predicting promiscuity," *Nature*, vol. 462, no. 7270, pp. 167–168, 2009.

[15] A. C. Schierz, "Virtual screening of bioassay data," *Journal of cheminformatics*, vol. 1, no. 1, p. 21, 2009.

[16] P. Willett, J. M. Barnard, and G. M. Downs, "Chemical similarity searching," *Journal of chemical information and computer sciences*, vol. 38, no. 6, pp. 983–996, 1998.

[17] A. Bender and R. C. Glen, "Molecular similarity: a key technique in molecular informatics," *Organic & biomolecular chemistry*, vol. 2, no. 22, pp. 3204–3218, 2004.

[18] A. Koutsoukas, B. Simms, J. Kirchmair, P. J. Bond, A. V. Whitmore, S. Zimmer, M. P. Young, J. L. Jenkins, M. Glick, R. C. Glen *et al.*, "From in silico target prediction to multi-target drug design: current databases, methods and applications," *Journal of proteomics*, vol. 74, no. 12, pp. 2554–2574, 2011.

[19] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, and B. K. Shoichet, "Relating protein pharmacology by ligand chemistry," *Nature biotechnology*, vol. 25, no. 2, pp. 197–206, 2007.

[20] L. Jacob and J.-P. Vert, "Protein-ligand interaction prediction: an improved chemogenomics approach," *Bioinformatics*, vol. 24, no. 19, pp. 2149–2156, 2008.

[21] J. Li, A. Fu, and L. Zhang, "An overview of scoring functions used for protein–ligand interactions in molecular docking," *Interdisciplinary Sciences: Computational Life Sciences*, pp. 1–9, 2019.

[22] S.-Y. Huang, "Comprehensive assessment of flexible-ligand docking algorithms: current effectiveness and challenges," *Briefings in Bioinformatics*, vol. 19, no. 5, pp. 982–994, 2018.

[23] H. Yu, J. Chen, X. Xu, Y. Li, H. Zhao, Y. Fang, X. Li, W. Zhou, W. Wang, and Y. Wang, "A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data," *PloS one*, vol. 7, no. 5, 2012.

[24] P. Zhang, L. Tao, X. Zeng, C. Qin, S. Chen, F. Zhu, Z. Li, Y. Jiang, W. Chen, and Y.-Z. Chen, "A protein network descriptor server and its use in studying protein, disease, metabolic and drug targeted networks," *Briefings in bioinformatics*, vol. 18, no. 6, pp. 1057–1070, 2017.

[25] X. Xiao, J.-L. Min, P. Wang, and K.-C. Chou, "iGPCR-Drug: A web server for predicting interaction between GPCRs and drugs in cellular networking," *PloS one*, vol. 8, no. 8, 2013.

[26] A. Ezzat, M. Wu, X.-L. Li, and C.-K. Kwoh, "Drug-target interaction prediction via class imbalance-aware ensemble learning," *BMC bioinformatics*, vol. 17, no. 19, p. 509, 2016.

[27] D.-S. Cao, Y.-Z. Liang, J. Yan, G.-S. Tan, Q.-S. Xu, and S. Liu, "PyDPI: freely available python package for chemoinformatics, bioinformatics, and chemogenomics studies," 2013.

[28] A. Ezzat, M. Wu, X.-L. Li, and C.-K. Kwoh, "Drug-target interaction prediction using ensemble learning and dimensionality reduction," *Methods*, vol. 129, pp. 81–88, 2017.

[29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[30] I. Wallach, M. Dzamba, and A. Heifets, "AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery," *arXiv preprint arXiv:1510.02855*, 2015.

[31] H. Ding, I. Takigawa, H. Mamitsuka, and S. Zhu, "Similarity-based machine learning methods for predicting drug–target interactions: a brief review," *Briefings in bioinformatics*, vol. 15, no. 5, pp. 734–747, 2014.

[32] Z. Mousavian and A. Masoudi-Nejad, "Drug–target interaction prediction via chemogenomic space: learning-based methods," *Expert opinion on drug metabolism & toxicology*, vol. 10, no. 9, pp. 1273–1287, 2014.

[33] T. Pahikkala, A. Airola, S. Pietilä, S. Shakyawar, A. Szwajda, J. Tang, and T. Aittokallio, "Toward more realistic drug–target interaction predictions," *Briefings in bioinformatics*, vol. 16, no. 2, pp. 325–337, 2015.

[34] A. Ezzat, M. Wu, X.-L. Li, and C.-K. Kwoh, "Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey," *Briefings in bioinformatics*, vol. 20, no. 4, pp. 1337–1357, 2019.

[35] A. S. Rifaioglu, H. Atas, M. J. Martin, R. Cetin-Atalay, V. Atalay, and T. Doğan, "Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases," *Briefings in bioinformatics*, vol. 20, no. 5, pp. 1878–1912, 2019.

[36] M. Bagherian, E. Sabeti, K. Wang, M. A. Sartor, Z. Nikolovska-Coleska, and K. Najarian, "Machine learning approaches and databases for prediction of drug–target interaction: a survey paper," *Briefings in Bioinformatics*.

[37] N. Hecker, J. Ahmed, J. von Eichborn, M. Dunkel, K. Macha, A. Eckert, M. K. Gilson, P. E. Bourne, and R. Preissner, "SuperTarget goes quantitative: update on drug–target interactions," *Nucleic acids research*, vol. 40, no. D1, pp. D1113–D1117, 2012.

[38] J. Hastings, G. Owen, A. Dekker, M. Ennis, N. Kale, V. Muthukrishnan, S. Turner, N. Swainston, P. Mendes, and C. Steinbeck, "ChEBI in 2016: Improved services and an expanding collection of metabolites," *Nucleic acids research*, vol. 44, no. D1, pp. D1214–D1219, 2016.

[39] T. Sterling and J. J. Irwin, "ZINC 15–ligand discovery for everyone," *Journal of chemical information and modeling*, vol. 55, no. 11, pp. 2324–2337, 2015.

[40] A. Williams and V. Tkachenko, "The Royal Society of Chemistry and the delivery of chemistry data repositories for the community," *Journal of computer-aided molecular design*, vol. 28, no. 10, pp. 1023–1030, 2014.

[41] A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Krüger, Y. Light, L. Mak, S. McGlinchey *et al.*, "The ChEMBL bioactivity database: an update," *Nucleic acids research*, vol. 42, no. D1, pp. D1083–D1090, 2014.

[42] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker *et al.*, "PubChem substance and compound databases," *Nucleic acids research*, vol. 44, no. D1, pp. D1202–D1213, 2016.

[43] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic acids research*, vol. 44, no. D1, pp. D457–D462, 2016.

[44] C. UniProt, "UniProt: a hub for protein information," *Nucleic Acids Res*, vol. 43, no. D1, pp. D204–12, 2015.

[45] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[46] T. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal *et al.*, "Human protein reference database—2009 update," *Nucleic acids research*, vol. 37, no. suppl_1, pp. D767–D772, 2009.

[47] R. D. Finn, T. K. Attwood, P. C. Babbitt, A. Bateman, P. Bork, A. J. Bridge, H.-Y. Chang, Z. Dosztányi, S. El-Gebali, M. Fraser *et al.*, "Interpro in 2017—beyond protein family and domain annotations," *Nucleic acids research*, vol. 45, no. D1, pp. D190–D199, 2017.

[48] M. Kuhn, D. Szklarczyk, S. Pletscher-Frankild, T. H. Blicher, C. Von Mering, L. J. Jensen, and P. Bork, "STITCH 4: integration of protein–chemical interactions with user data," *Nucleic acids research*, vol. 42, no. D1, pp. D401–D407, 2014.

[49] A. Ahmed, R. D. Smith, J. J. Clark, J. B. Dunbar Jr, and H. A. Carlson, "Recent improvements to Binding MOAD: a resource for protein–ligand binding affinities and structures," *Nucleic acids research*, vol. 43, no. D1, pp. D465–D469, 2015.

[50] M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, and J. Chong, "BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology," *Nucleic acids research*, vol. 44, no. D1, pp. D1045–D1053, 2016.

[51] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," *Journal of chemical information and computer sciences*, vol. 28, no. 1, pp. 31–36, 1988.

[52] "Rdkit: Open-source cheminformatics software, 2020. http//www.rdkit.org/ (3 March 2020, date last accessed)."

[53] H. Liu, J. Sun, J. Guan, J. Zheng, and S. Zhou, "Improving compound–protein interaction prediction by building up highly credible negative samples," *Bioinformatics*, vol. 31, no. 12, pp. i221–i229, 2015.

[54] J. Tang, A. Szwajda, S. Shakyawar, T. Xu, P. Hintsanen, K. Wennerberg, and T. Aittokallio, "Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis," *Journal of Chemical Information and Modeling*, vol. 54, no. 3, pp. 735–743, 2014.

[55] R. Todeschini and V. Consonni, *Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references*. John Wiley & Sons, 2009, vol. 41.

[56] A. Cereto-Massagué, M. J. Ojeda, C. Valls, M. Mulero, S. Garcia-Vallvé, and G. Pujadas, "Molecular fingerprint similarity search in virtual screening," *Methods*, vol. 71, pp. 58–63, 2015.

[57] S. Riniker and G. A. Landrum, "Open-source platform to benchmark fingerprints for ligand-based virtual screening," *Journal of cheminformatics*, vol. 5, no. 1, p. 26, 2013.

[58] R. Sawada, M. Kotera, and Y. Yamanishi, "Benchmarking a Wide Range of Chemical Descriptors for Drug-Target Interaction Prediction Using a Chemogenomic Approach," *Molecular informatics*, vol. 33, no. 11-12, pp. 719–731, 2014.

[59] G. Cano, J. Garcia-Rodriguez, A. Garcia-Garcia, H. Perez-Sanchez, J. A. Benediktsson, A. Thapa, and A. Barr, "Automatic selection of molecular descriptors using random forest: Application to drug discovery," *Expert Systems with Applications*, vol. 72, pp. 151–159, 2017.

[60] J. Shen, F. Cheng, Y. Xu, W. Li, and Y. Tang, "Estimation of ADME properties with substructure pattern recognition," *Journal of chemical information and modeling*, vol. 50, no. 6, pp. 1034–1041, 2010.

[61] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *Journal of chemical information and modeling*, vol. 50, no. 5, pp. 742–754, 2010.

[62] Y. Bengio, Y. LeCun *et al.*, "Scaling learning algorithms towards AI," *Large-scale kernel machines*, vol. 34, no. 5, pp. 1–41, 2007.

[63] P. Hu, K. C. C. Chan, and Z. You, "Large-scale prediction of drug-target interactions from deep representations," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 1236–1243.

[64] Z.-H. You, K. C. Chan, and P. Hu, "Predicting protein-protein interactions from primary protein sequences using a novel multi-scale local feature representation scheme and the random forest," *PloS one*, vol. 10, no. 5, p. e0125811, 2015.

[65] L. Wang, Z.-H. You, X. Chen, S.-X. Xia, F. Liu, X. Yan, and Y. Zhou, "Computational methods for the prediction of drug-target interactions from drug fingerprints and protein sequences by stacked auto-encoder deep neural network," in *International Symposium on Bioinformatics Research and Applications*. Springer, 2017, pp. 46–58.

[66] M. Gribskov, A. D. McLachlan, and D. Eisenberg, "Profile analysis: detection of distantly related proteins," *Proceedings of the National Academy of Sciences*, vol. 84, no. 13, pp. 4355–4358, 1987.

[67] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[68] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[69] C. Wang, J. Liu, F. Luo, Y. Tan, Z. Deng, and Q.-N. Hu, "Pairwise input neural network for target-ligand interaction prediction," in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2014, pp. 67–70.

[70] M. Wen, Z. Zhang, S. Niu, H. Sha, R. Yang, Y. Yun, and H. Lu, "Deep-learning-based drug–target interaction prediction," *Journal of proteome research*, vol. 16, no. 4, pp. 1401–1409, 2017.

[71] M. Hamanaka, K. Taneishi, H. Iwata, J. Ye, J. Pei, J. Hou, and Y. Okuno, "CGBVS-DNN: Prediction of Compound-protein Interactions Based on Deep Learning," *Molecular informatics*, vol. 36, no. 1-2, p. 1600045, 2017.

[72] Z.-R. Li, H. H. Lin, L. Han, L. Jiang, X. Chen, and Y. Z. Chen, "PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence," *Nucleic Acids Research*, vol. 34, no. suppl_2, pp. W32–W37, 2006.

[73] A. Mauri, V. Consonni, M. Pavan, and R. Todeschini, "Dragon software: An easy approach to molecular descriptor calculations," *Match*, vol. 56, no. 2, pp. 237–248, 2006.

[74] S. Jaeger, S. Fulle, and S. Turk, "Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition," *Journal of Chemical Information and Modeling*, vol. 58, no. 1, pp. 27–35, 2018.

[75] E. Asgari and M. R. K. Mofrad, "Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics," *PLOS ONE*, vol. 10, no. 11, 2015.

[76] Y.-F. Zhang, X. Wang, A. C. Kaushik, Y. Chu, X. Shan, M.-Z. Zhao, Q. Xu, and D.-Q. Wei, "SPVec: A Word2vec-Inspired Feature Representation Method for Drug-Target Interaction Prediction," *Frontiers in Chemistry*, vol. 7, p. 895, 2020.

[77] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *ICLR*, 2013.

[78] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of machine learning research*, vol. 12, no. Aug, pp. 2493–2537, 2011.

[79] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[80] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: deep drug–target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, 2018.

[81] H. Öztürk, E. Ozkirimli, and A. Özgür, "WideDTA: prediction of drug-target binding affinity," *arXiv preprint arXiv:1902.04166*, 2019.

[82] Q. Zhao, F. Xiao, M. Yang, Y. Li, and J. Wang, "AttentionDTA: prediction of drug–target binding affinity using attention model," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2019, pp. 64–69.

[83] X. Wang, Y. Liu, F. Lu, H. Li, P. Gao, and D. Wei, "Dipeptide Frequency of Word Frequency and Graph Convolutional Networks for DTA Prediction," *Front Bioeng Biotechnol*, vol. 8, p. 267, 2020.

[84] I. Lee, J. Keum, and H. Nam, "DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences," *PLoS computational biology*, vol. 15, no. 6, p. e1007129, 2019.

[85] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh, "GraphDTA: Predicting drug–target binding affinity with graph neural networks," *Bioinformatics*, 10 2020, btaa921.

[86] W. Elhefnawy, M. Li, J. Wang, and Y. Li, "DeepFrag-k: a fragment-based deep learning approach for protein fold recognition," *BMC bioinformatics*, vol. 21, no. 6, pp. 1–12, 2020.

[87] M. Zeng, F. Zhang, F.-X. Wu, Y. Li, J. Wang, and M. Li, "Protein–protein interaction site prediction through combining local and global features with deep neural networks," *Bioinformatics*, vol. 36, no. 4, pp. 1114–1120, 09 2019. [Online]. Available: https://doi.org/10.1093/bioinformatics/btz699

[88] M. Tsubaki, K. Tomii, and J. Sese, "Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences," *Bioinformatics*, vol. 35, no. 2, pp. 309–318, 2019.

[89] Y. Weng, C. Lin, X. Zeng, and Y. Liang, "Drug Target Interaction Prediction using Multi-task Learning and Co-attention," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Conference Proceedings, pp. 528–533.

[90] Y.-B. Wang, Z.-H. You, S. Yang, H.-C. Yi, Z.-H. Chen, and K. Zheng, "A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network," *BMC Medical Informatics and Decision Making*, vol. 20, no. 2, p. 49, 2020.

[91] K. Y. Gao, A. Fokoue, H. Luo, A. Iyengar, S. Dey, and P. Zhang, "Interpretable Drug Target Prediction Using Deep Neural Representation," in *IJCAI*, 2018, pp. 3371–3377.

[92] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J. B. Ghasemi, and A. Masoudi-Nejad, "DeepCDA: deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks," *Bioinformatics*, vol. 36, no. 17, pp. 4633–4642, 05 2020.

[93] X. Zheng, S. He, X. Song, Z. Zhang, and X. Bo, "DTI-RCNN: New Efficient Hybrid Neural Network Model to Predict Drug–Target Interactions," in *International Conference on Artificial Neural Networks*. Springer, 2018, pp. 104–114.

[94] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, pp. 802–810, 2015.

[95] M. Abdel-Basset, H. Hawash, M. Elhoseny, R. K. Chakrabortty, and M. Ryan, "DeepH-DTA: Deep Learning for Predicting Drug-Target Interactions: A Case Study of COVID-19 Drug Repurposing," *IEEE Access*, vol. 8, pp. 170 433–170 451, 2020.

[96] C. Lu, M. Zeng, F.-X. Wu, M. Li, and J. Wang, "Improving circRNA-disease association prediction by sequence and ontology representations with convolutional and recurrent neural networks," *Bioinformatics*, 12 2020, btaa1077. [Online]. Available: https://doi.org/10.1093/bioinformatics/btaa1077

[97] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[98] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[99] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[100] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

[101] Z. Quan, Y. Guo, X. Lin, Z. Wang, and X. Zeng, "GraphCPI: Graph Neural Representation Learning for Compound-Protein Interaction," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Conference Proceedings, pp. 717–722.

[102] M. Jiang, Z. Li, S. Zhang, S. Wang, X. Wang, Q. Yuan, and Z. Wei, "Drug–target affinity prediction using graph neural network and contact maps," *RSC Advances*, vol. 10, no. 35, pp. 20 701–20 712, 2020.

[103] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," in *Advances in neural information processing systems*, 2015, pp. 2224–2232.

[104] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, "Molecular graph convolutions: moving beyond fingerprints," *Journal of computer-aided molecular design*, vol. 30, no. 8, pp. 595–608, 2016.

[105] C. W. Coley, R. Barzilay, W. H. Green, T. S. Jaakkola, and K. F. Jensen, "Convolutional Embedding of Attributed Molecular Graphs for Physical Property Prediction," *Journal of Chemical Information and Modeling*, vol. 57, no. 8, pp. 1757–1772, 2017, pMID: 28696688.

[106] M. M. Stepniewska-Dziubinska, P. Zielenkiewicz, and P. Siedlecki, "Development and evaluation of a deep learning model for protein–ligand binding affinity prediction," *Bioinformatics*, vol. 34, no. 21, pp. 3666–3674, 2018.

[107] M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri, and D. R. Koes, "Protein–ligand scoring with convolutional neural networks," *Journal of chemical information and modeling*, vol. 57, no. 4, pp. 942–957, 2017.

[108] Y. Li, M. A. Rezaei, C. Li, X. Li, and D. Wu, "DeepAtom: A Framework for Protein-Ligand Binding Affinity Prediction," *arXiv preprint arXiv:1912.00318*, 2019.

[109] H. Zhang, L. Liao, K. M. Saravanan, P. Yin, and Y. Wei, "DeepBindRG: a deep learning based method for estimating effective protein–ligand affinity," *PeerJ*, vol. 7, p. e7362, 2019.

[110] L. Zheng, J. Fan, and Y. Mu, "OnionNet: a Multiple-Layer Intermolecular-Contact-Based Convolutional Neural Network for Protein–Ligand Binding Affinity Prediction," *ACS Omega*, vol. 4, no. 14, pp. 15 956–15 965, 2019.

[111] K. Tian, M. Shao, Y. Wang, J. Guan, and S. Zhou, "Boosting compound-protein interaction prediction by deep learning," *Methods*, vol. 110, pp. 64–72, 2016.

[112] Y. Du, J. Wang, X. Wang, J. Chen, and H. Chang, "Predicting drug-target interaction via wide and deep learning," in *Proceedings of the 2018 6th International Conference on Bioinformatics and Computational Biology*, 2018, pp. 128–132.

[113] Q. Duan, C. Flynn, M. Niepel, M. Hafner, J. L. Muhlich, N. F. Fernandez, A. D. Rouillard, C. M. Tan, E. Y. Chen, T. R. Golub *et al.*, "LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures," *Nucleic acids research*, vol. 42, no. W1, pp. W449–W460, 2014.

[114] L. Xie, S. He, X. Song, X. Bo, and Z. Zhang, "Deep learning-based transcriptome data classification for drug-target interaction prediction," *BMC genomics*, vol. 19, no. 7, p. 667, 2018.

[115] Q. Feng, E. V. Dueva, A. Cherkasov, and M. Ester, "PADME: A Deep Learning-based Framework for Drug-Target Interaction Prediction," *ArXiv*, vol. abs/1807.09741, 2018.

[116] G. E. Hinton, A. Krizhevsky, I. Sutskever, and N. Srivastva, "System and method for addressing overfitting in a neural network," Aug. 2 2016, uS Patent 9,406,017.

[117] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[118] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper With Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[119] M. Michel, D. Menéndez Hurtado, and A. Elofsson, "PconsC4: fast, accurate and hassle-free contact predictions," *Bioinformatics*, vol. 35, no. 15, pp. 2677–2679, 2019.

[120] F. Hu, J. Jiang, and P. Yin, "Interpretable Prediction of Protein-Ligand Interaction by Convolutional Neural Network," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Conference Proceedings, pp. 656–659.

[121] A. Gonczarek, J. M. Tomczak, S. Zaręba, J. Kaczmar, P. Dąbrowski, and M. J. Walczak, "Interaction prediction in structure-based virtual screening using deep learning," *Computers in biology and medicine*, vol. 100, pp. 253–258, 2018.

[122] W. Torng and R. B. Altman, "Graph Convolutional Neural Networks for Predicting Drug-Target Interactions," *Journal of Chemical Information and Modeling*, vol. 59, no. 10, pp. 4131–4149, 2019.

[123] J. Lim, S. Ryu, K. Park, Y. J. Choe, J. Ham, and W. Y. Kim, "Predicting Drug–Target Interaction Using a Novel Graph Neural Network with 3D Structure-Embedded Graph Representation," *Journal of chemical information and modeling*, vol. 59, no. 9, pp. 3981–3988, 2019.

[124] H. NT and T. Maehara, "Revisiting graph neural networks: All we have is low-pass filters," *arXiv preprint arXiv:1905.09550*, 2019.

[125] S. Chaudhari, G. Polatkan, R. Ramanath, and V. Mithal, "An attentive survey of attention models," *arXiv preprint arXiv:1904.02874*, 2019.

[126] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a "siamese" time delay neural network," in *Advances in neural information processing systems*, 1994, pp. 737–744.

[127] S. Zheng, Y. Li, S. Chen, J. Xu, and Y. Yang, "Predicting drug protein interaction using quasi-visual question answering system," *bioRxiv*, p. 588178, 2019.

[128] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.

[129] B. Shin, S. Park, K. Kang, and J. C. Ho, "Self-attention based molecule representation for predicting drug-target interaction," *arXiv preprint arXiv:1908.06760*, 2019.

[130] M. Karimi, D. Wu, Z. Wang, and Y. Shen, "DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks," *Bioinformatics*, vol. 35, no. 18, pp. 3329–3338, 2019.

[131] K. Huang, C. Xiao, L. M. Glass, and J. Sun, "MolTrans: Molecular interaction transformer for drug target interaction prediction," *Bioinformatics*, 10 2020, btaa880.

[132] L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, X. Luo, K. Chen, H. Jiang, and M. Zheng, "TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments," *Bioinformatics*, 2020.

[133] Y. Tay, M. Dehghani, D. Bahri, and D. Metzler, "Efficient transformers: A survey," *arXiv preprint arXiv:2009.06732*, 2020.

[134] C. J. Sigrist, L. Cerutti, E. De Castro, P. S. Langendijk-Genevaux, V. Bulliard, A. Bairoch, and N. Hulo, "PROSITE, a protein domain database for functional characterization and annotation," *Nucleic acids research*, vol. 38, no. suppl_1, pp. D161–D166, 2010.

[135] M. Woźniak, A. Wołos, U. Modrzyk, R. L. Górski, J. Winkowski, M. Bajczyk, S. Szymkuć, B. A. Grzybowski, and M. Eder, "Linguistic measures of chemical diversity and the "keywords" of molecular collections," *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.

[136] G. Huang, Z. Liu, L. V. Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," pp. 2261–2269, 2017.

[137] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[138] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in *The World Wide Web Conference*, 2019, pp. 2022–2032.

[139] X. Lin, K. Zhao, T. Xiao, Z. Quan, Z.-J. Wang, and P. S. Yu, "DeepGS: Deep Representation Learning of Graphs and Sequences for Drug-Target Binding Affinity Prediction," in *24th European Conference on Artificial Intelligence (ECAI)*, 2020, pp. 1–8.

[140] M. Lee, H. Kim, H. Joe, and H. Kim, "Multi-channel PINN: investigating scalable and transferable neural networks for drug discovery," *Journal of Cheminformatics*, vol. 11, no. 1, pp. 1–16, 2019.

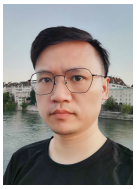
[141] B. Agyemang, W.-P. Wu, M. Y. Kpiebaareh, and E. Nanor, "Drug-Target Indication Prediction by Integrating End-to-End Learning and Fingerprints," in *2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing*. IEEE, 2019, pp. 266–272.

[142] B. Agyemang, W.-P. Wu, M. Y. Kpiebaareh, Z. Lei, E. Nanor, and L. Chen, "Multi-view self-attention for interpretable drug–target interaction prediction," *Journal of Biomedical Informatics*, vol. 110, p. 103547, 2020.

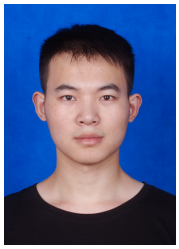
[143] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik, "Automatic chemical design using a data-driven continuous representation of molecules," *ACS central science*, vol. 4, no. 2, pp. 268–276, 2018.

[144] M. Skalic, D. Sabbadin, B. Sattarov, S. Sciabola, and G. De Fabritiis, "From target to drug: Generative modeling for the multimodal structure-based ligand design," *Molecular pharmaceutics*, vol. 16, no. 10, pp. 4282–4291, 2019.
[145] M. J. Kusner, B. Paige, and J. M. Hernández-Lobato, "Grammar variational autoencoder," *arXiv preprint arXiv:1703.01925*, 2017.
[146] Y. Li, L. Zhang, and Z. Liu, "Multi-objective de novo drug design with conditional graph generative model," *Journal of cheminformatics*, vol. 10, no. 1, p. 33, 2018.
[147] V. Chenthamarakshan, P. Das, S. Hoffman, H. Strobelt, I. Padhi, K. W. Lim, B. Hoover, M. Manica, J. Born, T. Laino *et al.*, "Cogmol: Target-specific and selective drug design for covid-19 using deep generative models," *Advances in Neural Information Processing Systems*, vol. 33, 2020.
[148] L. Zhao, J. Wang, L. Pang, Y. Liu, and J. Zhang, "GANsDTA: Predicting Drug-Target Binding Affinity Using GANs," *Frontiers in genetics*, vol. 10, pp. 1243–1243, 2020.
[149] H. Hassan-Harrirou, C. Zhang, and T. Lemmin, "RosENet: Improving Binding Affinity Prediction by Leveraging Molecular Mechanics Energies with an Ensemble of 3D Convolutional Neural Networks," *Journal of Chemical Information and Modeling*, vol. 60, no. 6, pp. 2791–2802, 2020.
[150] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[151] S. Li, F. Wan, H. Shu, T. Jiang, D. Zhao, and J. Zeng, "MONN: A Multi-objective Neural Network for Predicting Compound-Protein Interactions and Affinities," *Cell Systems*, vol. 10, no. 4, pp. 308–322.e11, 2020.
[152] K. Ishiguro, S.-i. Maeda, and M. Koyama, "Graph Warp Module: an Auxiliary Module for Boosting the Power of Graph Neural Networks in Molecular Graph Analysis," *arXiv preprint arXiv:1902.01020*, 2019.
[153] T. Lei, W. Jin, R. Barzilay, and T. Jaakkola, "Deriving neural architectures from sequence and graph kernels," *arXiv preprint arXiv:1705.09037*, 2017.
[154] A. S. Rifaioglu, E. Nalbat, V. Atalay, M. J. Martin, R. Cetin-Atalay, and T. Doğan, "DEEPScreen: high performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations," *Chem. Sci.*, vol. 11, pp. 2531–2557, 2020.
[155] M. Gönen and G. Heller, "Concordance probability and discriminatory power in proportional hazards regression," *Biometrika*, vol. 92, no. 4, pp. 965–970, 2005.
[156] E. D. Coelho, J. P. Arrais, and J. L. Oliveira, "Computational discovery of putative leads for drug repositioning through drug-target interaction prediction," *PLoS computational biology*, vol. 12, no. 11, 2016.
[157] F. Wan, Y. Zhu, H. Hu, A. Dai, X. Cai, L. Chen, H. Gong, T. Xia, D. Yang, M.-W. Wang, and J. Zeng, "DeepCPI: A Deep Learning-based Framework for Large-scale in silico Drug Screening," *Genomics, Proteomics & Bioinformatics*, vol. 17, no. 5, pp. 478–495, 2019.
[158] G. B. Goh, N. O. Hodas, C. Siegel, and A. Vishnu, "Smiles2vec: An interpretable general-purpose deep neural network for predicting chemical properties," *arXiv preprint arXiv:1712.02034*, 2017.

Qichang Zhao is a master student in the Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, School of Information Engineering at Xiangtan University, Hunan, China. His current research interests include machine learning, deep learning, and bioinformatics.



Mengyun Yang received the MS degree in computational mathematics from Hunan Normal University in 2012 and the PhD degree in computer science and technology from Central South University in 2020. He is an associate professor in the School of Science, Shaoyang University, Hunan, China. His current research interests include machine learning, deep learning, and bioinformatics.



Zhongjian Cheng is a master student in the Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering at Central South University, Hunan, China. His research interests include bioinformatics and machine learning.

Yaohang Li received the BS degree from South China University of Technology in 1997, and the MS and PhD degrees in computer science from Florida State University, Tallahassee, FL, USA, in 2000 and 2003, respectively. He is an associate professor in computer science at Old Dominion University, Norfolk, VA, USA. His research interests are in protein structure modeling, computational biology, bioinformatics, Monte Carlo methods, big data algorithms, and parallel and distributive computing.

Jianxin Wang received the BEng and MEng degrees in computer engineering from Central South University, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University, China, in 2001. He is the dean and a professor in the School of Computer Science and Engineering, Central South University, Changsha, Hunan, China. He is a senior member of the IEEE. His current research interests include algorithm analysis and optimization, parameraized algorithm, bioinformatics, and computer network.