International Journal of
Molecular Sciences

MDPI

Review

Comprehensive Survey of Recent Drug Discovery Using Deep Learning

Jintae Kim [1,†], Sera Park [1,†], Dongbo Min [2,*] and Wankyu Kim [1,3,*]

[1] KaiPharm Co. Ltd., Seoul 03759, Korea; contact@kaipharm.com (J.K.); srpark@kaipharm.com (S.P.)
[2] Computer Vision Lab, Department of Computer Science and Engineering, Ewha Womans University, Seoul 03760, Korea
[3] System Pharmacology Lab, Department of Life Sciences, Ewha Womans University, Seoul 03760, Korea
* Correspondence: dbmin@ewha.ac.kr (D.M.); wkim@ewha.ac.kr (W.K.)
† These authors contributed equally to this work.

Abstract: Drug discovery based on artificial intelligence has been in the spotlight recently as it significantly reduces the time and cost required for developing novel drugs. With the advancement of deep learning (DL) technology and the growth of drug-related data, numerous deep-learning-based methodologies are emerging at all steps of drug development processes. In particular, pharmaceutical chemists have faced significant issues with regard to selecting and designing potential drugs for a target of interest to enter preclinical testing. The two major challenges are prediction of interactions between drugs and druggable targets and generation of novel molecular structures suitable for a target of interest. Therefore, we reviewed recent deep-learning applications in drug–target interaction (DTI) prediction and de novo drug design. In addition, we introduce a comprehensive summary of a variety of drug and protein representations, DL models, and commonly used benchmark datasets or tools for model training and testing. Finally, we present the remaining challenges for the promising future of DL-based DTI prediction and de novo drug design.

Keywords: artificial intelligence-based drug discovery; deep learning; drug–target interaction; virtual screening; de novo drug design; molecular representation; benchmark tool

Citation: Kim, J.; Park, S.; Min, D.; Kim, W. Comprehensive Survey of Recent Drug Discovery Using Deep Learning. *Int. J. Mol. Sci.* **2021**, *22*, 9983. https://doi.org/10.3390/ijms22189983

Academic Editor: Michael J. Parnham

Received: 24 August 2021
Accepted: 10 September 2021
Published: 15 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The primary goal of drug discovery is to develop safe and effective medicines for human diseases. All the drug development processes—from target identification to step-by-step clinical trials—require significant amount of time and cost. As costs increase gradually with every step, it is essential to ensure that appropriate drug candidates are selected for the next phase at each milestone. In particular, the "hit-to-lead" process is a pivotal step in identifying promising lead compounds from hits and determining their potential as therapeutics. One of the reasons why clinical trials face side effects and lack in vivo efficacy is that single or multiple drugs often interact with multiple targets based on the concept of polypharmacology [1]. Ideally, full-scale in vivo tests for each disease model should be able to address this problem; however, that will require astronomical time and effort. Computer-aided drug discovery or design methods have played a major role in this hit-to-lead process by reducing the burden of consumptive validation experiments since the 1980s in modern pharmaceutical research and development (R & D) [2–4]. However, even this in silico approach has not prevented the decline in pharmaceutical industry R & D productivity since the mid-1990s.

Recently, much effort was invested in drug discovery through artificial intelligence (AI), which has enabled significant and cost-effective development strategies in academia and pharmaceutical industries. The vast amounts of chemical and biological data accumulated over decades, along with technological automation through the availability of

high-performance processors such as graphics processing unit computing, paved the way for AI in drug development [4–6]. Not only state-of-the-art AI technologies are adopted in the drug development process, but also diverse pipelines or frameworks for AI-driven drug development are being built [7–9]. Utilizing deep neural networks provides the advantage of understanding the very complex contexts of biological space. This is because nonlinear models can be constructed in hidden layers to extract complex patterns from multi-level representations. It also minimizes the work of manually preprocessing unformatted raw data and selecting all kinds of features. Consequently, advances made in the development of deep learning (DL)-based methods have led to successful outcomes for prediction of drug–target interactions (DTIs) and generation of novel molecules with desired properties [4,10,11]. However, since datasets for drug development exhibit types and distributions that are different from those used in traditional AI data, such as images and texts, further attempts are still required to analyze data from a different angle and apply the latest DL techniques.

In this review, we introduce essential data representations and DL models for DTI prediction and de novo drug design. In addition, we investigate recent advances and benchmark datasets in DL-based methods in the following sections: The "Data Representation" section introduces several data representations of the inputs that were used in DL-based drug discovery. The "Deep Learning Models" section explains DL methods for drug discovery via comparison of the strengths and weaknesses of models. In the two sections, "Deep Learning Methods for Drug–Target Interaction Prediction" and "Deep Learning Methods for De Novo Drug Design," we classify and describe the models for each of the DTI predictive models and de novo drug design models based on their utility. The "Benchmarking Datasets and Tools" section demonstrates the commonly used benchmark dataset and publicly available benchmarking tools. Finally, we discuss the advantages and limitations of the current methods as well as the remaining important challenges and future perspective for DL-based drug discovery in the "Limitation and Future Work" section.

In the "Deep Learning Methods for Drug–Target Interaction Prediction" section, the description of the following studies was minimized: (1) studies that predict compound properties not considering protein targets such as blood–brain barrier permeability, solubility, lipophilicity, and chemical-based adverse effect [10,11]; (2) target prediction studies that determine targets for the existing drugs, such as reverse docking simulation [12]; (3) studies that only use knowledge-based documents from which information is extracted by text mining techniques [13]; (4) studies that focus on only optimizing binding between the drug and target using molecular dynamics simulation [14–16].

2. Data Representation

The input data for DL-based drug discovery are molecules; drugs and protein targets are small molecules or macromolecules. To characterize these molecules, several types of molecular representations (often referred to as descriptors or features) have been used in many machine learning (ML) methods—from simple sequences of molecular entities to manually predefined molecular features [17,18] (Figure 1). However, because it is directly related to the knowledge of learning models, data representation has a significant impact on pre-training to improve performance of predictive models. There has been a surge of interest in research on representation of molecules, and these efforts can contribute to capturing unknown features of compounds and targets [19,20]. Learning expressive representation from molecular structures is one of the challenges of these studies [21]. Besides, many recent DL models tend to use three-dimensional (3D) representations based on protein–ligand complexes, including molecular graphs, atom-pair fingerprints, and voxels. The various molecular representations utilized in deep neural networks as drug representations and target representations are described separately in this section.

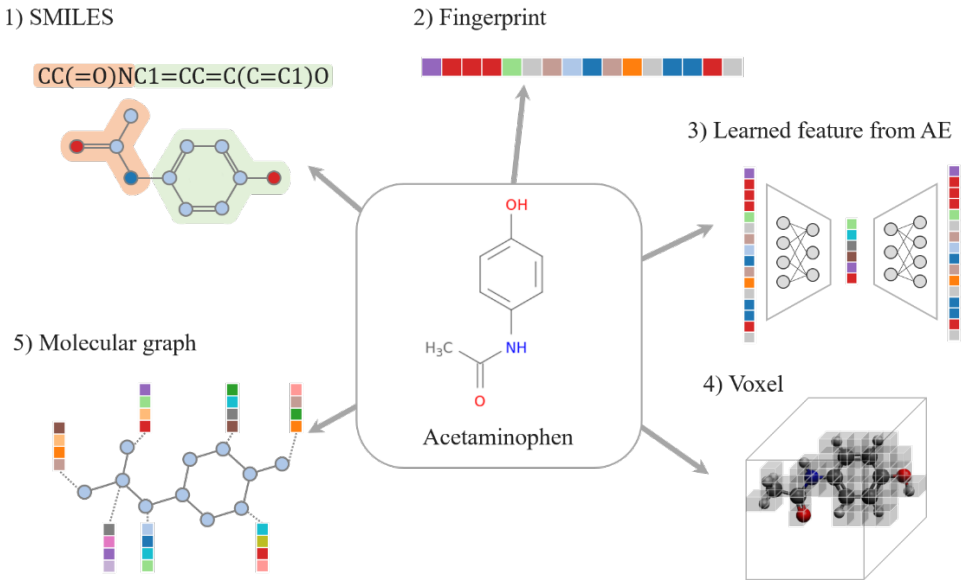

Figure 1. Different types of drug representations used in DL-based drug discovery. This figure shows the drug representations of acetaminophen, which is widely used to treat mild to moderate pain. (**1**) SMILES: a string that expresses structural features including phenol group and amide group. (**2**) Fingerprint: a 16-digit color-coded 64-bit MACCS key fingerprint. (**3**) Learned representations: In this case, it depicts the features learned from an autoencoder (AE). (**4**) Voxel: binary volume elements with atoms assigned to a cube with a fixed grid size. (**5**) Molecular graph: Each node encodes the network information of the molecular graph.

2.1. Drug Representations

2.1.1. SMILES

The most commonly used drug representation is a simplified molecular-input line-entry system (SMILES) string, which is a line notation that encodes structural, geometric, and topological properties of a molecule. The SMILES is simple and easy to obtain; therefore, it enables fast training. A number of molecular deep-learning models use the SMILES as the molecular representation [22–25]. As it is a sequence-based feature, a SMILES string can be directly used as a "sentence" to learn the representations. Sequence models using the SMILES can be successfully applied to predict chemical reactions. For example, many studies have shown promising results for synthetic prediction by directly converting to the predicted reactants in the SMILES format through seq-2-seq approaches to illustrate the reactions of compounds [26–28]. Furthermore, most molecules can be randomly generated with more than one SMILES string by starting in different atoms and changing the atomic orders, and this randomized SMILES model exhibits much better performance [29,30].

2.1.2. Fingerprint

Another chemical structure-implemented feature is molecular FP, which is a bit string encoding structural or pharmacological feature of a ligand. Many types of molecular FPs have been proposed for similarity comparisons for virtual screening (VS), including ligand-based similarity searching and quantitative structure-–activity relationship (QSAR) analysis. A number of deep-learning-based DTI prediction models also used FPs as input features [24,31–34]. We discuss three types of widely used molecular FPs: key-based FPs, hashed FPs, and pharmacophore FPs.

Key-based FPs include molecular ACCess system (MACCS) and PubChem FP. The MACCS keys are composed of predefined 166 substructures. PubChem FP [35] has 881 bits and each bit tests the presence of element count, type of ring, atom pairing, and nearest neighbors, etc. These structural keys are designed for substructure retrieval. Therefore,

although it is possible to quickly and accurately find substructures, there is a limit to classifying various characteristics.

Hashed FPs, such as Daylight FP, Morgan FP, extended-connectivity FP (ECFP), and functional-class FP (FCFP), are also used in the similarity analysis of compounds. Unlike key-based FPs, hashed FPs do not require predefined substructures and are instead created by a hash function to convert all possible fragments to numeric values. ECFP, a circular FP based on the Morgan algorithm [36], is often used in a wide range of applications, including DL models for the DTI prediction. This is because several DL methods using ECFP exhibited robustness in bioactivity prediction [8,37].

Pharmacophore FPs have pharmacophoric features such as aromatic, hydrophobic, charged, and hydrogen bond donor/acceptor. The pharmacophore FPs consider the overlapping of energy-minimized conformations of a set of known ligands and the extraction of recurrent pharmacophoric properties [38]. Many studies have used these pharmacophoric features to assess similarities between binding sites [39].

Finally, recent studies have attempted to add 3D structures to FPs to accurately predict binding affinity [8,40,41]. Gao et al. [33] reassessed the predictive power of 2D and 3D FPs and concluded that 2D FPs are still competitive in prediction of toxicity, physicochemical properties, and ligand-based binding affinity; however, 3D structure-based models outperformed 2D-based counterparts in the protein–ligand binding prediction. In other words, 2D FPs are still competitive; therefore, considering the structural properties at the 2D and 3D levels together will yield better results.

2.1.3. Learned Representations

"word2vec" is a very popular method in natural language processing for word embedding [42]. With word2vec, the meaning of the word is learned and reflected in the coordinates; therefore, an ML model can better characterize the words. The embedding method, which usually adds "(2)vec" to the end, is inspired by "word2vec" and treats a molecule or protein as a sentence or word of a natural language and converts it into a real vector. ProtVec [43] and Mol2vec [44] are representative representations, and there are other methods such as SPvec [40] and SMILES2vec [41]. When converted to the 2vec type, specific information such as atom (or amino acid) type or bonding relationship of the original data cannot be known without restoration. Therefore, it is not suitable for a de novo design and is mainly used for property prediction. It is also known that the 2vec type has better prediction accuracy than the SMILES or FP [32].

There are other learned representation methods that employ DL. Recently, many studies used deep representation learning to encode molecules. The most common learned representation method is AutoEncoder (AE) [45]. The AE extracts the potential characteristics that make the input data distinguishable and compress them into vectors of desired length, called as latent vectors. Currently, the transformer model is preferred over the AE [46]. However, to train representation learning, a large amount of data is required; however, a publicly available pre-trained transformer or AE model can also be used without making your own DL model [46]. X-Mol [46] or MolGNet [47] are well-known frameworks designed for this purpose. By fine-tuning these frameworks according to the purpose, they can be used in a variety of ways—from property prediction, DDI, DTI, de novo design, to molecule optimization.

Additionally, Denis et al. applied wave transform for efficient representation of sparse voxel data [48], and Ziyao et al. proposed HamNet [49] considering molecular conformation in their study.

2.1.4. Voxel

A voxel is a combination of "volume" and "pixel," which is a data representation that extends a 2D image into three dimensions. In the 3D space, a value is assigned to the geographic location where the atom exists, and the rest is filled with zeros. The value can be 1 to indicate the presence of only an atom, and it may be an encoded value corresponding

to the type of atom or a quantum chemical property such as hydropathy or electric charge [50]. In addition, as mentioned in Section 2.2, a voxel is used as the expression of the target protein and has the advantage that it can express only the pocket that reacts with the ligand instead of the entire protein [51]. Because the voxel has specific 3D information, it is a very suitable expression for binding prediction.

Resolution is important when using voxel. For example, if the size of a voxel is 20 × 20 × 20, the number of features in the input data is 8k, so the size is large; however, most of the data are filled with zeros. If the resolution is lowered to reduce the size, the accuracy will decrease, and if the resolution is increased, the data size and training speed will increase significantly. There are also 3D mesh [52] or point cloud [53] for geometric 3D representation; however, the voxel is most widely used in the DTI field.

2.1.5. Molecular Graph

A molecular graph is a mapping of atoms constituting a molecule to nodes and chemical bonds to edges. In molecular graphs, nodes are sometimes represented using symbols in the periodic table to indicate atom types, or using some kinds of functional groups or fragments [20]. The edge attributes can describe bond strength or bond resonances between two atoms, which is important training data that is expressed as the adjacent matrix in the graph convolution neural network model [11]. With the development of graph neural networks, recent DL-based works have adopted molecular graphs as drug or target representations for both DTI prediction models and novel molecular design models. Notably, molecular graphs can represent not only 2D structures but also 3D structures with the spatial information including atomic coordinates, bond angles, and chirality. However, since the arrangement of atoms in a three-dimensional space changes constantly, the space in a molecular graph is almost infinite. Some successful results using 3D graph representation were obtained for the DTI prediction by avoiding inefficient computations [54]; however, the 3D structure data of the protein–ligand complex is insufficient. Thus, the model can memorize the features of the training data extensively [55]. If you need a further explanation of molecular graphs, we recommend referring to Ref. [20], which provides a good review of molecular representations.

2.2. Target Representations

2.2.1. Sequence-Based Feature

The simple and primary feature of targets is a protein sequence composed of a linear composition of amino acid residues, which is easy to obtain and serves as the input of the recurrent neural network (RNN). Amino acid sequences including protein primary structures have been frequently used as a target representation in the predictive models. Information on protein primary structures can also generate a variety of target properties, such as monopeptide/dipeptide/tripeptide composition (also called protein sequence composition description; PSC), sequence motif, and functional domain. Lin Zhu et al. [56] listed the types of protein features derived from the amino acid composition: amino acid composition, sequence order, etc. Some studies [34,57] considered a position specific scoring matrix (PSSM) as the target feature. The PSSM is derived from an ordered set of sequences that are presumed to be functionally related and serves as an important feature that is widely used in the prediction of DNA or RNA binding sites (e.g., PSI-BLAST [58]) [34].

2.2.2. Structure-Based Feature

For known protein structures, as previously described in the drug presentation part, molecular descriptors such as atom-pair map, voxel, and molecular graph were often used for the target representation to determine structurally matching ligands or to design new ligands for the protein [17,40,59,60]. However, there are not many known protein structures. Genetically encoded amino acid sequences determine the remarkable diversity of the molecular functions performed by finely tuned 3D structures (i.e., tertiary structure)

through protein folding. Accurately predicting the folding structure of proteins in a real biological system is important in biomedicine and pharmacology. Scientists directly analyzed the stereoscopic structure of proteins using methods such as X-ray crystallography, nuclear magnetic resonance spectroscopy, or cryo-electron microscopy to decipher the structures of more than 170,000 protein species; however, it took a very long time to analyze a single protein. For this reason, the tertiary structure of proteins has been determined using computational methods. Since there are many variables in protein folding, this challenge has not been addressed over fifty years in computational biology [61]. Recently, DeepMind developed AlphaFold and succeeded in accurately predicting the 3D structure of proteins from amino acid sequences [62]. As these prediction results are open to the public, it is expected that all scientists will be able to gain new insights and spur their discovery in drug development.

2.2.3. Relationship-Based Feature

The key questions for druggable targets are how to extract important features of the drug-binding site and how to predict potential space. Many DL models have simply applied amino-sequence-based features; however, some studies have focused on a variety of aspects of target proteins. Some groups [59,63,64] utilize not only sequences but also pathway membership information such as gene ontology (GO) terms [60] and MSigDB pathways [65]. A number of studies [59,63,66–68] employed a protein–protein interaction (PPI) network, which is generated into network-based features by node2vec or AE. Other studies showed the transcriptome data, such as connectivity map (CMAP) [69] and Library of integrated network-based cellular signatures (LINCS)-L1000 database [70], can be utilized as target features for the DTI prediction models [63,71]. The transcriptional profile is an integrated result of many genetic processes; thus, the characteristic changes in the transcriptional profiles can denote the underlying mechanisms of diseases of interest [72].

3. Deep Learning Models

Basic DL models can be classified according to their purpose, loss function, learning method, and structure. When DL was initially applied for drug development, there were studies using only a single model; however, recently, there are very few cases where only a basic model is used. In most cases, two or more of the basic models introduced below are combined. There are many different types of DL models, but only very basic ones have been described in this section. We describe the strengths and weaknesses of each model and introduce the characteristics of the models from the point of view of drug discovery.

3.1. Multi-Layer Perceptron

Multi-layer perceptron (MLP) is the most common neural network structure and is also called the fully connected layer, linear layer, etc. MLP's strengths lie in classification and regression. Usually, it is trained by finding the optimal parameters that can minimize the error between the predicted value and the correct answer for the input. Since it is a standard model that has been studied extensively, various techniques have been established and almost all DL frameworks basically provide it; therefore, it is easy to apply, and stable performance can be expected. Because of its wide versatility, various data such as FP, transcriptome [71], bioassay [73], and molecular properties can be used along with the compound structure. Chen et al. used the MLP with four hidden layers for the DTI prediction, FP was used for compound, and various information such as PseAAC, PsePSSM, NMBroto, and structure feature were combined with the target protein information [57].

3.2. Convolutional Neural Network

Convolutional neural networks (CNN) extract local features by calculating several adjacent features through the same computational filter. By stacking the CNNs in several layers, global features including local features can be extracted. The CNN is generally used when all the input data are single-modal-like image recognition. The convolution filter is the same regardless of the amount of the input data. Even if the amount of the input data is large, the number of calculations increases; however, the number of parameters of the DL model does not increase much. Therefore, it is efficient for training. It is also relatively robust against noise from the input data. Because the CNN is well suited to atomistic geometry [74], it is often used in combination with the voxel or image-type data. DEEPScreen [19] used a 2D image of the molecule, and RoseNet [15], AK-score [75], and DeepDrug3D [51] predicted the DTI by converting the protein and ligand into a voxel. Although it is not optimized for sequential expression methods such as SMILES or amino acids, the CNNs are sometimes used instead of the RNNs [76]. Both DeepConv-DTI [31] and transformer-CNN [77] used the CNNs for sequential input data to build QSAR models.

3.3. Graph Neural Network

Most of the data used in ML is expressed in the form of a vector that matches the Euclidean space. In the case of single-vector-type data or sequential data, models such as MLP, RNN, or transformer can be used; however, it is not appropriate to apply these models to data expressed in relational graphs such as social networks. A graph neural network is a model designed to learn graph-type data in a DL method [78]. There are various graph types of data in drug discovery—compound structure, DTI relationship, PPI, patient–disease relationship, etc. There are several types of graph neural networks (GNN); however, the graph convolution network (GCN) [79,80] adopting the CNN method and the graph attention network(GAT) applying the attention mechanism are representative [47].

GNNs are widely used in many ways; however, they stand out primarily in three applications. The first is to predict the properties of compounds using representation learning. Yang et al. claimed that the GNN method that they employed has better property prediction performance than the existing methods [21]. According to this trend, recently, the GNN is widely applied for property prediction [81]. The second is to learn relationship information between different domains such as heterogeneous and bipartite [82,83]. For example, it is possible to learn the relationship between patients and diseases, and between genes and drugs, making it possible to utilize comprehensive meta data [83]. Lastly, in the field of de novo design, compounds are generated or optimized by the GNN [84].

3.4. Recurrent Neural Network

If sequential data are put into the recurrent neural network (RNN) one by one in order, it is influenced by the previous input value to derive the next output value. When the RNN was first introduced, AI performance in the natural language field, which was difficult to achieve previously, was greatly improved, and it became one of the famous models of DL. In addition, it can be used to embed structural information as a kind of representation learning by extracting the weight of the hidden layer and treating it as a feature with sequence information. However, the naive RNN has a simple structure, and there are performance limitations for application in various situations. The most important problem is the vanishing gradient problem, which exhibits poor performance for long-length data such as large proteins and large compounds because the length of the input sequence exponentially reduces the impact of items far from the currently entered item [85]. Moreover, since the same operation is executed repeatedly as the length of the input sequence, the length of the sequence increases the training time. Even when the

items in sequential data have complex intrinsic relationships, their characteristics are not well learned.

Long short-term memory (LSTM) was invented in the 1990s and began to be widely used in the late 2000s [86]. LSTM was introduced to address the fast-vanishing problem of naive RNNs. The LSTM can be used with good performance even on longer sequential data compared to the RNN. Since its introduction, various modifications of the LSTM have been proposed [87], and recently, gated recurrent units (GRU) with a simpler internal structure [88] has also been widely used. It can simply be used for the de novo drug design, which randomly generates short-length compounds, and can generate an appropriate candidate drug by inputting a target protein sequence [89]. The LSTM and GRU have exhibited significant improvements over the RNN and are widely used to replace the RNN in drug discovery [31,90,91]; however, the vanishing problem still persists, which makes it difficult to use very long sequence data.

3.5. Attention-Based Model

The self-attention technique is a method that was first proposed by the transformer model to introduce machine translation in the field of natural language processing. Self-attention is a technique that calculates the association between the elements included in a sequence and extracts features for each element based on the calculated result. Unlike the RNN, which uses a single hidden state in which all time step values are implied, the attention technique handles past data in parallel; therefore, the correlation with distant tokens can be used without reduction. Furthermore, bidirectional encoder representations from transformers (BERT), introduced by Devlin et al. [92] in 2018, has dramatically improved natural language presentation using DL and has been actively introduced in drug discovery.

In DTI, the transformer model was naturally absorbed into the traditional QSAR modeling using the RNN. Karpov et al. [77] used a model applying CNN to a transformer with SMILES as an input to predict drug activity. Shin et al. [23] proposed a molecular transformer DTI (MT-DTI) model that predicts drug-target binding affinity by embedding the protein sequence using the CNN and embedding the molecule structure using the BERT. Lennox et al. [93] also proposed a model with a concept similar to MT-DTI; however, it used BERT for both protein and chemical structures and was based on GCN. Lennox et al. evaluated that their model performed better in predicting binding affinity than MT-DTI.

3.6. Generative Adversarial Network

Generative adversarial network (GAN), first published in 2014 [94], is the most representative generative model in the field of DL. The GAN is only used in the de novo drug design and not in the DTI. Two DL modules, generator and discriminator, are included in pairs, and these two modules are trained adversarially with each other, and finally, the generator produces fake results that cannot be distinguished from the real ones by a discriminator. Although the GAN is used as a very powerful method for some data types such as image data, it has difficulty in generating large molecules compared to other generative models. In addition, the technical difficulty for training is somewhat higher than that of other models, and it involves problems such as mode collapse.

The GAN, combined with reinforcement learning (RL), is considered a very successful model for novel molecules generation. There are various GAN architecture applications used in drug discovery [95], however, we introduce only two very simple models. Objective-reinforced generative adversarial networks (ORGAN) [90], introduced by Guimaraes et al., and molecular GAN (MolGAN) [91], introduced by Cao and Kipf, are frequently cited successful models. Since ORGAN uses SMILES data as input, sequence GAN (seqGAN) [96] is used as the basic framework and RL is added. ORGAN showed good performance in drug likeliness and synthesizability except solubility compared to the naive RNN. The MolGAN is a similar concept to the ORGAN, but it applied the GCN

based on molecular graph representation and showed better performance than the ORGAN and the naive RNN.

3.7. Autoencoder

AE is a DL structure for basic unsupervised learning that consists of an encoder that compresses data and a decoder that reconstructs the data to their original shape. In this symmetric process, the dominant characteristic that distinguishes the data from each other is automatically extracted. A set of abstracted points compressed by the encoder, called a latent space, can be used in other models as new features. At the training stage, the encoder and decoder are trained simultaneously; however, after the training stage, only the encoder is separated and used for data embedding, dimension reduction, and visualization, or only the decoder is separated and used as a generation model. Because dimension reduction is possible without the need for data labels, it is good to use in combination with other DL models [97].

The points in the latent space created by the AE are very sparsely distributed; however, there is no continuous meaning between the points. Variational autoencoder (VAE) limits the latent space to a Gaussian-shaped stochastic fence. This increases the density of the latent space and makes the data continuous and smooth. Gómez-Bombarelli et al. showed that continuous search from one compound to another is possible in a smooth chemical latent space constructed using the SMILES [98]. AE is excellent in data compression and is used for the DTI, whereas VAE has lower compression performance than AE and is used for the de novo drug design due to the continuous and limited latent space characteristics described above.

Adversarial AE (AAE) [99] is a DL model that adds the GAN structure to the VAE, whose purpose is feature compression and generation. The VAE can compress the properties of compounds well; however, it exhibits inadequate performance to generate valid results. Conversely, the GAN can produce valid compounds and produce plausible results but can be biased with a single mode and have low diversity scores. Insilico Medicine first published the AAE [100,101] for the identification and generation of new compounds in 2016 [101], followed by an improved model named druGAN [102] in 2017. The AAE is a method that can show good performance in the generation of new compounds while compressing the data to the latent space. Polykovskiy et al. generated new compounds by changing the lipophilicity (logP) and synthetic accessibility of the input compound by adding the function to control the condition to AAE [103].

4. Deep Learning Methods for Drug–Target Interaction Prediction

DTI prediction using DL techniques incorporates both the chemical space of the compound and the genome space of the target protein into a pharmacological space, which is called as a chemogenomic (or proteochemometric, PCM) approach. This approach would ideally solve the DTI problem by building a chemogenomic matrix between the whole compounds and their biological proteins. Advances in the high-throughput screening (HTS) technology have enabled hundreds of thousands of compounds to be tested on biological targets in a very short time; however, it is practically impossible to obtain a complete chemogenomic matrix for a vast chemical space of 10^{60}.

As DL models predicting DTIs continue to apply state-of-the-art algorithms, combine multiple algorithms, and gradually shorten development periods, DL-based DTI prediction models have become so diverse that it is difficult to divide groups into appropriate categories [104]. Previous review papers have categorized the DTI prediction models into various groups [104–107]; however, none of these reviews have established a clearly distinguished classification scheme for DL methods. Another review categorized ML-based DTI prediction methods into docking simulation methods, ligand-based methods, GO-based methods and so on; however, this also did not provide detailed classification of the DL methods [108]. Therefore, we summarized the recent works using deep neural networks as prediction models for the DTIs. We describe the works by grouping them into

three branches according to their input features: (1) ligand-based approach, (2) structure-based approach, (3) relationship-based approach (Figure 2). Appendix A Tables A1–A3 summarize the studies involving each approach.

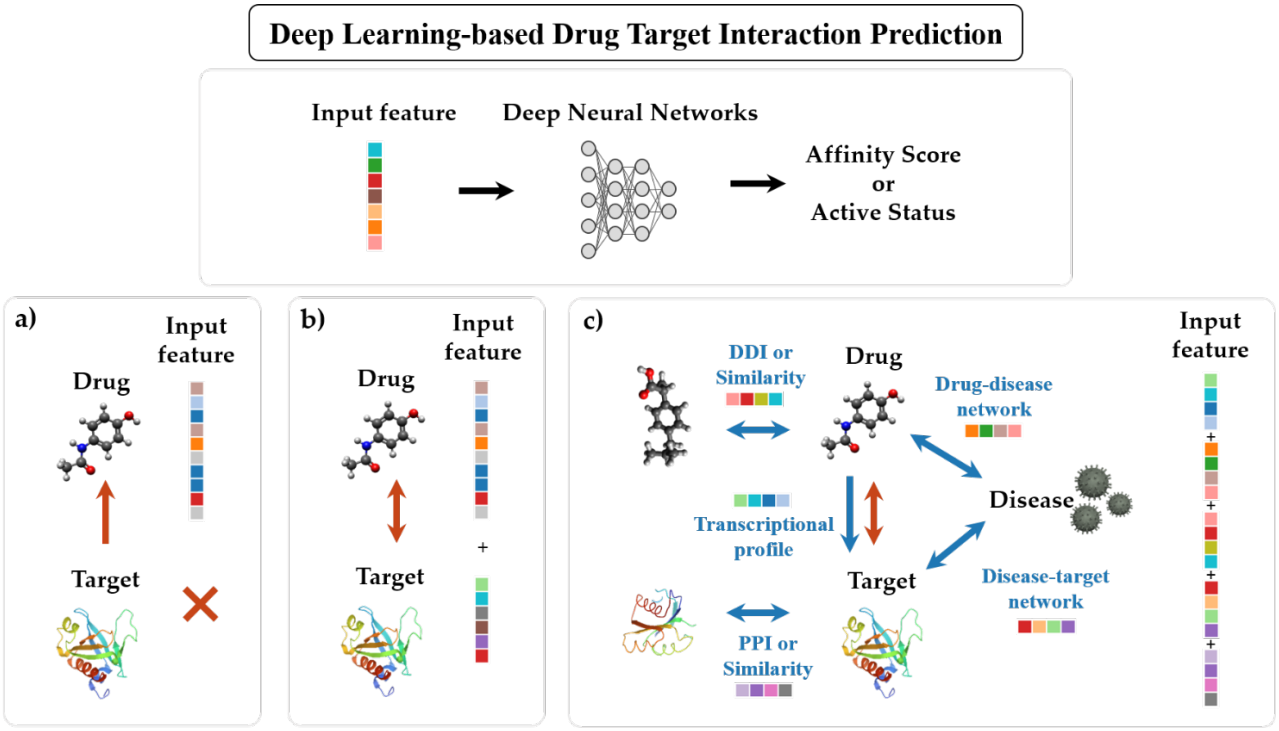

Figure 2. Deep learning-based drug–target interaction prediction. DL-based DTI prediction methods can be grouped based on their input features into three branches: (**a**) ligand-based approach, (**b**) structure-based approach, (**c**) relationship-based approach.

4.1. Ligand-Based Approach

The ligand-based approach is based on the hypothesis that a candidate ligand will be similar to the known ligands of the target proteins. It predicts the DTI via the ligand information of the target of interest. This approach includes similarity search methods that follow the assumption that structurally similar compounds usually have similar biological activities [6,109,110]. For decades, these VS methods have either prioritized compounds in large compound libraries through tremendous computing tasks or solved problems using manual formulas. The DL technology can shorten these cumbersome steps and manual tasks, and the difference between in silico prediction and empirical investigation is gradually narrowed through deep neural network models (Appendix A Table A1). Researchers have developed deep-learning-based VS for exploring compounds with desired characteristics, which has led to the revival of new drug designs, which will be detailed in the "De Novo Drug Designs" section.

With the development of benchmark packages such as MoleculeNet [111] and DeepChem [112], researchers can easily apply deep neural networks for analyzing ligands and predicting ligand-related properties, including bioactivities and physicochemical properties. Therefore, a number of ligand-based DL methods have adopted simple neural networks such as MLP and CNN [12,21,27,113]. In particular, ADMET studies tended to focus more on the representation power of the molecular descriptors than the model itself [27,34,35,114]. Hirohara et al. applied the SMILES string to a CNN model and detected motifs with important structures for protein-binding sites or unknown functional groups from learned features [25]. Wenzel et al. investigated multi-task deep neural networks using atom pairs and pharmacophoric donor–acceptor pairs as descriptors for predicting

microsomal metabolic liability [115]. Gao et al. employed several ML algorithms, including random forest, single-task deep neural network, and multi-task deep neural network models, in order to conduct comparisons of six types of 2D FPs in the protein–ligand binding affinity prediction [33]. Matsuzaka and Uesawa developed a CNN model that predicts agonists for constitutive androstane receptors by training 2D images of 3D chemical structures [109]. They optimized the best performance in snapshots at different angles or coordinates of a 3D ball-and-stick model, and as a result, the approach outperformed the predictions of typical 3D chemical structures.

Several studies applied state-of-art techniques such as graph convolution network and graph attention network for bioactivity or physicochemical property prediction. Since the introduction of the GCN was introduced, GCN models in drug-related applications constructed graph representations of a molecule that included information about the chemical substructures by summing up all the features of all the adjacent atoms [116]. Many studies have applied the GCNs as 3D descriptors instead of SMILES strings and evaluated that these learned descriptors outperformed in the prediction tasks and are more interpretable than the existing descriptors [23,24,86]. Chemi-net utilized the GCN models for molecular representation and compared performances between single-task and multi-task deep neural networks on their internal QSAR datasets [81]. Yang et al. proposed an advanced model, the directed message passing neural network (D-MPNN), by adopting a directed message-passing paradigm. They extensively compared their models on 19 public and 16 internal datasets and found that the D-MPNN models performed better or exhibited similar performance in most of the datasets [21]. They underperformed compared to traditional 3D descriptors in two datasets and were not robust when the dataset was small or extremely imbalanced. Then, another study group also practically used this D-MPNN model and successfully predicted an antibiotic, called halicin, which showed bactericidal efficacy in mice animal models [22]. This became the first case that led to antibiotic discovery by exploring a large-scale chemical space with DL methods that cannot be afforded by the current experimental approaches.

Another promising recent approach is the applications of attention-based graph neural networks [79]. Because the edge features can vary the graph representations for a molecule, the edge weights can be jointly learned with the node features. Thus, Shang et al. proposed an edge attention-based multi-relational GCN [11]. They built a dictionary of attention weights for each edge (i.e., individual bonds in the molecule), and as this dictionary is shared across the entire molecule, the model becomes robust to various input sizes. Consequently, the model can efficiently learn pre-aligned features from inherent properties of the molecular graph, and they evaluated that the performance of this model is better than that of the random forest model in Tox21 and HIV benchmark datasets. Withnall et al. [21] introduced a further augmentation with an attention mechanism to the MPNN model, called attention message passing neural network (AMPNN), which takes the weighted summation in the message passing stage [117]. They also extended the D-MPNN model (the reference by Yang et al. mentioned in the previous paragraph [21]) by attention mechanism in the same way as the AMPNN and called it the edge memory neural network (EMNN). This model outperformed other models on the standardized missing data from the maximum unbiased validation (MUV) benchmark set, although it is computationally more consumptive than other models.

4.2. Structure-Based Approach

Contrary to the ligand-based VS, structure-based VS uses both protein targets and their ligand information. Typical molecular docking simulation methods aimed at estimating geometrically feasible binding of ligands and proteins of a known tertiary structure [110]. While many ML methods for the DTI prediction utilize a variety of structural descriptors of ligands and targets as input features, several reviews separated these ML methods from typical structure-based approaches in methodology classification and classified them as feature-based methods [2,117,118]. However, we believe that the recent

studies incorporating DL of feature-based methods utilize the same method because the training is essentially performed with structural features [17,40,59,119,120]. Appendix A Table A2 shows the recent applications of structure-based DL methods for the DTI prediction.

One of the most commonly used DTI prediction methods in recent years is the use of 1D descriptors for drug and target [26,33,37,42,121]. As described in the previous representation section, drug and target can be expressed as sequences of atoms and amino acid residues, respectively, and the sequence-based descriptors have been preferred because DL models can be applied immediately without any tricky preprocessing of input features. DeepDTA proposed by Öztürk [24] applied only sequence information of the SMILES string and amino acid sequences to a CNN model and outperformed moderate ML methods such as KronRLS [122] and SimBoosts [123] on the Davis kinase binding affinity dataset [113] and KIBA dataset [114]. Wen et al. chose the common and simple features, such as ECFPs and protein sequence composition descriptors, and trained the features by a semi-supervised learning through deep belief-network [37]. This study suggested that in a problem where a very sparse set of total DTI pairs is used for the training, even a small dataset can be predicted more accurately by unsupervised pre-training. Another work called DeepConv-DTI constructed a deep convolution neural network model using only a type of RDKit Morgan FP and protein sequences [31]. They additionally captured local residue patterns of target protein sequences from the pooled convolution results, which can give high values to important protein regions such as actual binding sites.

A keystone of the structure-based regression model is the score function that ranks the binding potential of the protein–ligand 3D complexes and parametrizes the training data to predict the binding affinity values or binding pocket sites of the target proteins. AtomNet incorporated the 3D structural features of the protein–ligand complexes to the CNNs [124]. Their understanding is that the interactions between the chemical groups in the protein–ligand complexes are predominantly constrained in a local space; therefore, the CNN architecture is appropriate to learn local effects such as hydrogen bonding and π-bond stacking. They vectorized fixed-size 3D grids (i.e., voxel) over the protein–ligand complexes, and then each grid cell represents the structural features in that location. Since then, many researchers have investigated deep CNN models using voxels for binding affinity prediction or binding pocket site prediction [17,40,59,60,120], and these models have shown improved performance compared to the popular docking methods such as AutoDock Vina [125] or Smina [126]. This is because the CNN models are relatively resistant to noise in the input data and can be trained even when the input size is large.

Similar to the trends in the ligand-based methods, many DTI studies based on the structure-based methods using the GCNs have been published [8,127,128]. Feng et al. adopted both the ECFPs and GCNs as ligand features [8]. Compared to previous models such as KronRLS [122] and SimBoost [123], their models showed better performance on the Davis [113], Metz [129], and KIBA [114] benchmark datasets. However, they acknowledged that their GCN model could not beat their ECFP model because of difficulties in applying the GCN due to time and resource constraints. Another DTI prediction study by Torng et al. built an unsupervised graph-AE to learn the fixed-size representations of the protein-binding pockets [118]. Then, they used the initialized protein-pocket GCN in the pre-trained GCN model, while the ligand GCN model was trained using the automatically extracted features. They concluded that this model effectively captured the protein–ligand binding interactions without relying on the target–ligand complexes.

Attention-based DTI prediction methods have emerged because the attention mechanism-implemented models have key advantages that make the model interpretable [25,128,130]. Gao et al. used encoded vectors using the LSTM recurrent neural networks for the protein sequences and the GCN for ligand structures [119]. In particular, they focused on explaining the ability of their approach to provide biological insights to interpret the DTI predictions. To this end, two-way attention mechanisms were used to compute

the interaction of the drug–target pairs (DTPs) interact, enabling scalable interpretability to incorporate high-level information from the target proteins, such as GO terms. The Molecule transformer DTI (MT-DTI) method was proposed by Shin et al. using the self-attention mechanism for drug representations [23]. The pre-trained parameters from the publicly available 97 million compounds (PubChem) were transferred to the MT-DTI model, and it was fine-tuned and evaluated using two Davis [113] and KIBA [114] benchmark datasets. However, they did not apply the attention mechanism to represent the protein targets because the target sequence length was long, which takes a considerable amount of time to calculate, and there is not enough target information to pre-train. On the other hand, AttentionDTA presented by Zhao et al. combines an attention mechanism for the CNN models to determine the weight relationships between the compound and protein sequences [120]. They demonstrated that the affinity prediction tasks by the MLP model performed well on these attention-based drug and protein representations.

4.3. Relationship-Based Approach

According to polypharmacology, most compounds have more effects not only on their primary targets, but also on other targets. These effects depend on the dose of the drug and the related biological networks. Therefore, in silico proteochemometric modeling turned out to be useful, particularly when profiling selectivity or promiscuity of the ligands for proteins [121]. Moreover, multi-task learning neural networks are well suited for learning aspects of these different types of data simultaneously [131]. There are many applications of DL models that utilize relational information for multiple perspectives such as DTI-related heterogeneous networks and drug-induced gene-expression profiles. A network-based approach uses heterogeneous networks that integrate more than two types of nodes (drugs, target proteins/genes, diseases, or side effects) and various types of edges (similarities between drugs, similarities between proteins, drug–drug interactions (DDIs), PPIs, drug-disease association, protein-disease association, etc.) [132,133].

The key point of this approach is the use of local similarity between the nodes in the networks. For example, when a similarity network with drugs as nodes and drug–drug similarity values as the weights of the edges is considered, the DTIs can be predicted by utilizing their relationships and topological properties. It is based on the "guilt-by-association" theory that interacting entities are more likely to share functionalities [13]. Various ML methods that incorporate heterogeneous networks have been used as the prediction frameworks, e.g., support vector machine [134,135], regularized least square model (RLS) [127,128,136], and random walk with the restart algorithm [122,137].

With growing interest in the use of DL technologies, network-based DTI prediction studies using DL have been shown to improve the existing association prediction methods for measuring the topological similarities of bipartite (drug and target networks) and tripartite linked networks (drug, target, and disease networks) [15,69,73,74,104]. Zong et al. exploited the tripartite networks through the application of the DeepWalk method [130] to obtain the local latent information and compute topology-based similarities, and they demonstrated the potential of this method as a drug repurposing solution [13].

Some network-based DTI prediction studies used relationship-based features that were extracted by training the AE. A DTI-CNN prediction model devised by Zhao et al., used low-dimensional but rich depth features in a heterogeneous network trained by the stacked AE algorithm [68]. Moreover, in vivo experimental validation on atherosclerosis found that tetramethylpyrazine could attenuate atherosclerosis by inhibiting signal transductions in platelets. Other two studies [59,97] also applied the AE to capture the global structure information of the similarity measures. A study by Wang et al. applied a deep AE and introduced positive pointwise mutual information to compute the topological similarity matrix of drug and target [59]. Meanwhile, another study by Peng et al. utilized a denoising AE to select network-based features and reduce the dimensions of represen-

tations [97]. The denoising AE adds noise to high-dimensional, noisy, and incomplete input data and enables the encoder to learn more robustly by making the self-encoder learn to denoise.

However, these approaches have a limitation in that it is difficult to predict new drugs or targets, which is well known as the "cold start" problem of the recommendation systems [138]. These models are strongly influenced by the size and shape of the network; thus, if the network is not sufficiently comprehensive, they do not capture the properties of all the drugs or targets that may not appear in the network [13,139].

The other approach is using transcriptome data for DTI predictions, which measures the biological effect of drug action in in vitro experimental conditions. After the first release of the CMAP [69], a large-scale drug-induced transcriptome dataset, there have been many studies that succeeded in the identification of the drug repositioning candidates for a variety of diseases or the elucidation of the drug mode of action [140–143]. A number of studies have also employed the gene-expression profiles as the chemogenomic features for predicting DTIs. These studies are based on the assumption that drugs with similar expression profiles affect common targets [144,145].

Recent studies incorporated the updated version of CMAP, LINCS-L1000 database [70] into the DL DTI models [67,77,146]. Xie et al. built a binary classification model using a deep neural network based on the LINCS drug perturbation and gene knockout results [71]. On the other hand, Lee and Kim used the expression signature genes as the input drug and target features. They trained the rich information considering three distinct aspects of protein function, which included pathway-level memberships and PPI extracted using node2vec [63]. DTIGCCN by Saho and Zhang used a GCN model to extract the features of drug and target, respectively, from the LINCS data and CNN model to extract the latent features to predict DTPs [147]. In this hybrid model, they found that the Gaussian kernel function was helpful in building high-quality graphs; thus, their model showed better performance on classification tasks. The relationship-based DL methods described above are provided in Appendix A Table A3.

5. Deep Learning Methods for De Novo Drug Design

In general, when classifying the de novo drug designs, studies are classified based on the DL models [87]. However, in the case of actual implementation, it may be an appropriate classification; however, it may not be enough to understand the purpose of the model. In reflection of the recent trend changes, purpose, and usability, drug design using DL has been newly classified (Figure 3).

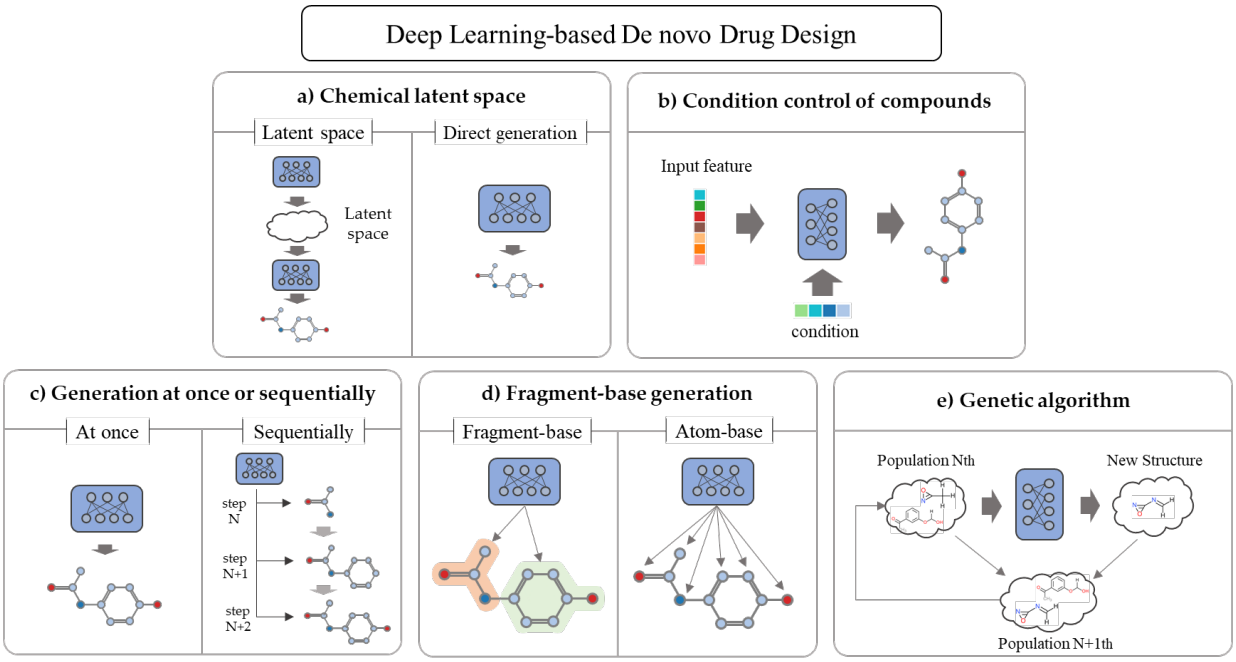

Figure 3. Deep learning-based de novo drug design. DL-based de novo drug design can be classified into five types according to function and method. (**a**) Classification according to the presence or absence of chemical latent space using manifold learning. (**b**) Classification according to the existence of condition control function. (**c**) Classification based on sequential generation. (**d**) Classification based on whether the molecule is produced in fragments or atoms. (**e**) Classification according to genetic algorithm using DL.

5.1. Chemical Latent Space

The manifold hypothesis states that there exists a low-dimensional subspace that explains specific data well within the original data space [148]. The latent space is mapped with a relatively low-dimensional vector space. The latent space created through manifold learning expresses the potential characteristics of the input data well. A representative model that performs manifold learning is the AE. Dimension reduction methods based on mathematical algorithms such as principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and singular value decomposition (SVD) have similar functions; however, they have limitations in determining complex manifolds compared to the ML methods.

Converting to a latent vector as an input feature has several advantages. First, the dimension of input data is reduced. Reducing the dimension reduces the risk of overfitting the module and makes learning easier with less data. Second, it becomes possible to search or optimize molecules in the latent space [149–152]. Since similar compounds or proteins are more densely expressed in a well-trained latent space [153,154], it is also possible to compare properties or calculate compounds according to compound structures [98]. A model named GENTRL proposed by Zhavoronkov et al. [7] generated compounds reflecting the latest trends that are indirectly extracted from patent dates using self-organizing map (SOM) from the chemical latent space. Instead of learning the encoder separately, if it is connected to a DL model and learned simultaneously during the training process, a latent space is produced that is more suitable for the purpose.

However, since the latent space is also data-dependent, it is produced differently each time, and it may be constructed differently or in an incomprehensible form unlike human intention. Recently, pre-trained encoders such as X-MOL [46] are provided as independent modules. By using transfer learning, described in Section 7.2, the performance can be improved by fine-tuning the latent space trained for a universal situation according to an appropriate purpose.

5.2. Condition Control of Compounds

There are two methods for studying new drug candidates in the de novo drug design using DL. The first method is generating as many arbitrary compounds as possible and filtering them through several steps according to the purpose and finally determining a small number of candidate drugs [7]. The second method is forcing conditions or properties to meet the purpose of the generation [45]. It cannot be said that either one is better; however, many random generation methods were used in the relatively early days, and recently there are many studies on controlling the condition. Condition-controllable models are useful when creating new drugs or optimizing the existing drugs because the condition control model can modify properties such as binding affinity, logP, molecular weight, side effects, and toxicity while maintaining the main structural characteristics of the molecule.

There are various locations and methods of applying the condition (Figure 4). Lim et al. [45] manipulated the properties of the compound to be produced using conditions in the VAE (Figure 5). This simple model is trained by adding molecular properties (MW, logP, HBD, HBA, topological polar surface area (TPSA)) to both the encoder and decoder of the VAE. When creating a new compound, the researcher only needs to add the desired property to the latent vector that determines the structure of the compound. The difference between the input value and the output value is approximately 10%. Lim et al. conducted an interesting experiment to observe the change in the condition. The first was to create a set of compounds with similar properties but different structures by putting the properties of aspirin in a random latent vector. Second, several compounds similar to aspirin were generated by adding the properties of aspirin to a latent vector near aspirin. Finally, similar to the second experiment, the properties of Tamiflu were added to the latent vector near Tamiflu; however, the structural change was observed while changing the logP to various values. Kang et al. [155] made a model almost similar to Lim et al.'s model [45]. However, while Lim et al.'s model can use only compounds with known properties, Kang et al.'s semi-supervised VAE (SSVAE) can use more compounds for training by inputting the predicted results from the property predictor. Hong et al. [156] suggested different structures in their two previous studies. They used the AAE model instead of the VAE and connected the property predictor from the latent vector to refine the latent space, and then they put the compound properties (logP, SAS, and TPSA) together with the latent vector in the decoder in the training process.

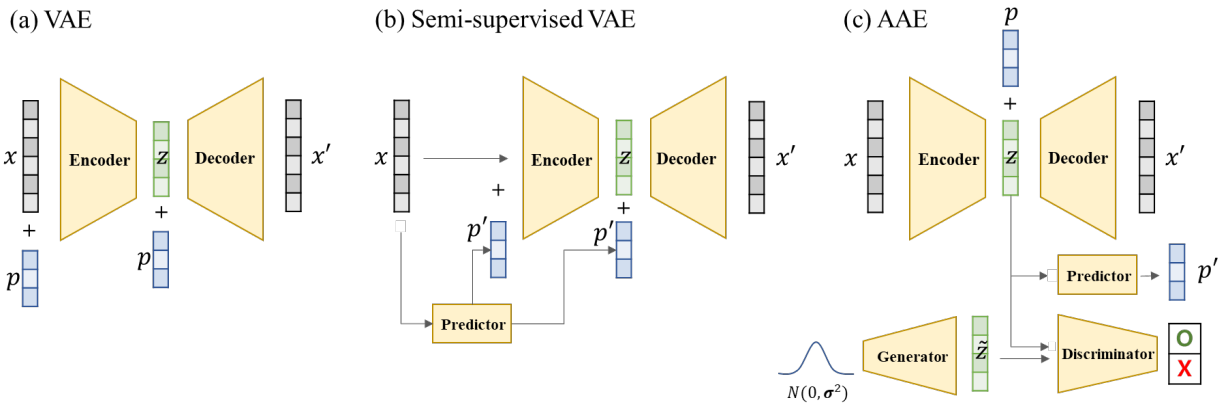

Figure 4. Three types of conditional de novo model using VAE. z is a latent vector and p is a molecular property. (**a**) Basic model for property control [45]. Concatenate the molecular property with the input value to the encoder and decoder. (**b**) Model with the property predictor [155] added to (**a**). It is possible to train even for molecules without the property data. (**c**) Model with condition control applied to the AAE [156]. Modify the latent space by predicting properties from the latent vector. This model does not add any properties to the encoder.

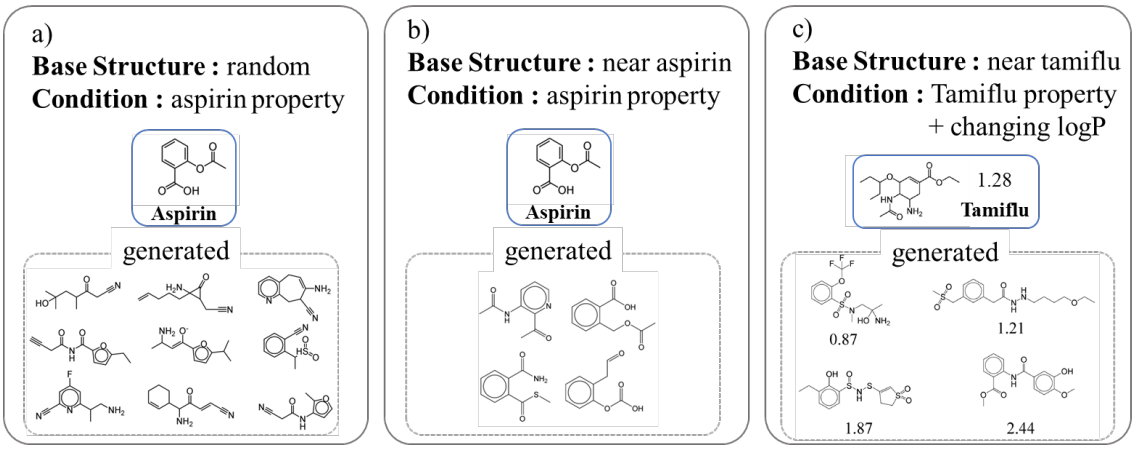

Figure 5. Generated compounds from conditional VAE. (**a**) Compounds created by adding aspirin condition to random points (compound structure) in the latent space. Although the form is different, they have similar properties (MW, logP, HBD, HBA, and TPSA). (**b**) Compounds produced by controlling the properties of aspirin at random points close to aspirin in the latent space. It looks very similar to aspirin. (**c**) Compounds produced by changing log P while maintaining the structure and other properties of Tamiflu. Only the desired properties in the reference compound can be controlled and improved. This figure is modified from [45].

5.3. Generation at Once or Sequentially

Basic generative models such as the AE or GAN form a compound from a corresponding input vector by a decoder or generator at once. In the case of the RNN using the SMILES, the word with the highest probability of matching the grammar is generated one by one from the start token until the end token appears, and finally, a large compound is completed. The one-time generation method is a method to create a new compound directly in the latent space, whereas the sequential method can start from nothing or a specific substructure and gradually complete the compound. The one-time method is simple and can provide more diverse results. Since the sequential method is generated while maintaining the active site or core scaffold with core characteristics, it can improve the binding score or properties; therefore, it can be used for fine lead optimization. Grisoni et al. [157] generated novel compounds using the SMILES and RNNs and Bongini et al. [158] using the GNNs. Lim et al. [84] created a graph-based sequential generative model from a specific scaffold rather than an atom, which has a property (MW, TPSA, logP) control function.

5.4. Fragment-Based Generation

A typical structure-based molecular representation, such as the SMILES or graph, consists of atoms and their junctions. However, compounds have more similar properties at the scaffold level than at the atomic level. The advantage of the fragment-based DL models is that when generating relatively large molecules, they output a product that is likely to exist in the natural state. For example, in the case of an atom-based model, the produced compound may include a ring consisting of 10 carbons, or a very long linear compound consisting of carbons, which are rare in nature. However, first, if the scaffold is used as a reference instead of the atom, it can be trained and created while maintaining the main substructure of the compound [84,146]. Second, it is easy to interpret and give feedback on the results based on the experts' existing knowledge [159]. For example, beta-lactam has a characteristic scaffold (Figure 6) [160] so when developing a new antibiotic, various types of drugs can be created while maintaining the scaffold. Jin et al. [161] used molecular graphs to generate compounds from fragments (they called them motifs), and Arús-Pous et al. [146] used the SMILES to model adding fragments (they called them decorators) from a core scaffold.

Fragment-based generation has a limitation in that it becomes difficult to find a new molecular entity because only compounds similar to the existing scaffold structure are

generated. In addition, there are very few types of atoms and bonds that are used in molecules, whereas the scaffolds have many types if not restricted by certain criteria.

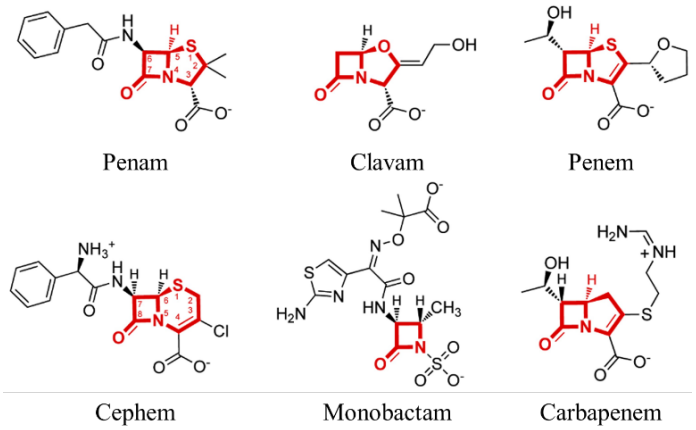

Figure 6. Chemical structures of the selected examples from six β-lactam structural categories. The core scaffolds (highlighted in red) are similar. Figure modified from [160].

5.5. Genetic Algorithm

Genetic algorithm is a method inspired by biogenetics, which has been traditionally used before DL [162], and it is mainly used to address optimization problems. The algorithm generates a random set of data called the initial generation and combines them to create a new generation. It repeats the process of intersecting some data with the highest score to create the next generation to obtain the most optimal result. If each data can be expressed in the form of a gene, and if there is a titration function that can be evaluated as a continuous value, it can be introduced relatively easily and produces a sufficiently acceptable result, although not perfect. De novo drug design using a genetic algorithm has been studied until recently [163,164], and the genetic algorithm method combined with DL [165] has been proposed in recent years.

6. Evaluation Method

6.1. Benchmarking Datasets and Tools

Over the decades, large amounts of repositories on the bioactivity, structure, and protein targets of small molecules have been accumulated in public databases such as PubChem [163], ChEMBL [164], and BindingDB [166]. These big datasets enable us to build predictive models for drug–target relationships via computational methods. To compare the performance of the models by evaluating their reproducibility for the prediction results, several datasets have been used. Appendix A Table A4 shows the list of benchmark datasets. These datasets consist of known active data and inactive compounds. In many datasets, an inactive compound is presumed to be inactive unless it is identified as active in an experimental biological assay, also referred to as "decoys" [167]. There are several benchmarking databases that provide refined decoy compounds [166,168]. They rationally selected inactive compounds to avoid false negatives because the false negative (i.e., active compounds are considered as inactive in the decoy sets) can underestimate the performance of the prediction methods. DUD-E [166], a gold standard dataset used for the evaluation of the VS methods, selected decoys based on the concept that the decoy compounds must be structurally different from the known ligands to reduce the false negative, whereas it must be similar to the known ligands with respect to physicochemical properties to reduce bias.

Chen et al. argued that the hidden bias in the widely used dataset (DUD-E database) may lead to misleading performance of the CNN models during the structure-based VS [169]. There were two remaining biases [170]. One is the limitation of exploring the decoy

restricted in the chemical space of reference compounds including active compounds (i.e., analogous bias). The other is the limitation of artificially good enrichment in evaluation because the physicochemical properties of the active compounds and decoy compounds (i.e., artificial enrichment bias) can be clearly distinguished. To overcome these limitations, the MUV datasets [168] and the demanding evaluation kits for objective in silico screening (DEKOIS) [171] were proposed. Until recently, fine-tuned benchmarking datasets have been consistently presented. Xia et al. proposed the unbiased ligand set (ULS) and unbiased decoy set (UDS) [172] for the G protein-coupled receptors. Another group used an asymmetric validation embedding procedure to design a novel dataset called LIT-PCBA dataset [173] for some PubChem bioassays.

Candidate drugs created using the de novo drug design cannot be measured for efficacy unless they are actually synthesized. The de novo drug design using DL has developed significantly in recent years [19], and designing a good generative model rather than generating an effective drug is a major evaluation criterion in the de novo drug design research field. MOlecular SEtS (MOSES) [174] and GuacaMol [175] are the most popular benchmarking tools. These two tools score and compare the performance of new DL models based on the base models. Both the tools use post-processing databases based on ZINC or ChEMBL, and include general representations such as FP, SMILES, and molecular graphs. The features and issues of both the tools for benchmarking in the de novo drug design are described in detail by Grant et al. [176] in their review paper.

6.2. Evaluation Metrics for DTI Prediction

The DTI prediction can be grouped into two types: 1) DTP prediction by a classification model that assigns a positive or negative (i.e., active or inactive) label to the DTP and 2) drug–target affinity (DTA) prediction by a regression model that estimates the binding affinity value between the drug and target. Because the evaluation metrics are different for each type, this section describes the performance metrics for each type of model.

6.2.1. Classification Metrics

The DTP prediction studies have adopted several common evaluation indicators including accuracy, precision, recall (also known as sensitivity), and specificity. These metrics are calculated from the confusion matrix. The most straightforward metric for classifier performance is accuracy. However, the accuracy metric does not work well in problems with skewness or class imbalance. For example, a prediction with a target that only affects less than 1% of all drugs is very easy to achieve 99% accuracy by obtaining the correct negatives even when few correct positives are predicted. For this reason, precision and recall are often quantified by many DL studies, which measure that the correctly predicted DTI with activity in practice is classified as positive repeatedly. Precision and recall can be measured simultaneously using two scores: F-score and precision–recall area under curve (AUPR). F-score (also called F measure) is a balance between precision and recall. For example, an F1 score is the weighted average value of precision and recall. The PR-AUC can show the tradeoff between precision and recall and reduce the impact of false positives. There are other useful metrics when the classes are imbalanced. Balanced accuracy applies the average of the sensitivity and specificity [177]. Matthews correlation coefficient (MCC) measures the correlation of the true classes with the predicted labels. Chicco and Jurman [178] showed that MCC is more informative in evaluating binary classifications than accuracy and F1 score.

Another commonly used evaluation indicator for DL methods is the area under the curve (AUC). The AUC means the area underneath a receiver–operator characteristic (ROC) curve that compares the performance of classifiers by distinguishing the two types of errors: false positives or negatives. The ROC curve is the plot of true positive rate (TPR, or sensitivity) against false positive rate (FPR, or 1-specificity), and the best classifier that will achieve perfection is the top-left of the plot (FPR = 0, and TPR = 1). The AUC value in the DTI prediction indicates how well the positive DTIs are ranked in the prediction. The

AUC is sensitive to the imbalanced DTI dataset, which are prone to a large number of false positives (i.e., few positive DTIs relative to negative DTIs) [179].

6.2.2. Regression Evaluation Metrics

Binding affinity scores such as IC50 and pKd predicted by DTA prediction models can be assessed by several evaluation indicators: mean square error (MSE), root mean square error (RMSE), Pearson's correlation coefficient (R), and squared correlation coefficient (R^2). These metrics have been implicated to determine the quality of predictive QSAR models. The MSE is defined as the average squared difference between the predicted and ground-truth binding affinity scores. The RMSE, as its name suggests, is the squared RMSE. R^2 measures how well the predicted values match the real values (i.e., goodness of fit). Some studies [23,120] applied the modified R^2 (r_m^2) to the test set prediction, which was introduced by Roy and Roy [180].

Other metrics such as concordance index (CI or C-index) and Spearman's correlation coefficient (ϱ) quantify the quality of rankings by comparing the order of the predictions and the order of the ground truths. A frequently used ranking metric in the DTA prediction is the CI [25,130,181]. When predicting the binding affinity values of two random DTPs, the CI measures whether those values were predicted in the same order as their actual values. The other metric, Spearman's correlation coefficient, measures the strength and direction of the association between two ranked variables. Several studies utilized Spearman's correlation coefficient with other metrics [17,59,60,182].

6.3. Evaluation Metrics for De Novo Drug Design

6.3.1. Generation Metrics

The generation index is meaningful in evaluating the performance of the DL generator model through the set of generated molecules rather than evaluating the generated compounds as drugs. This does not mean that models with better generative metrics make better drugs. The four generation indices are commonly used in the de novo drug design studies, and in the case of the SMILES data, an open library such as RDkit [119], GuacaMol [175], or MOSES [174] can be used to quickly measure the generation index (Figure 7).

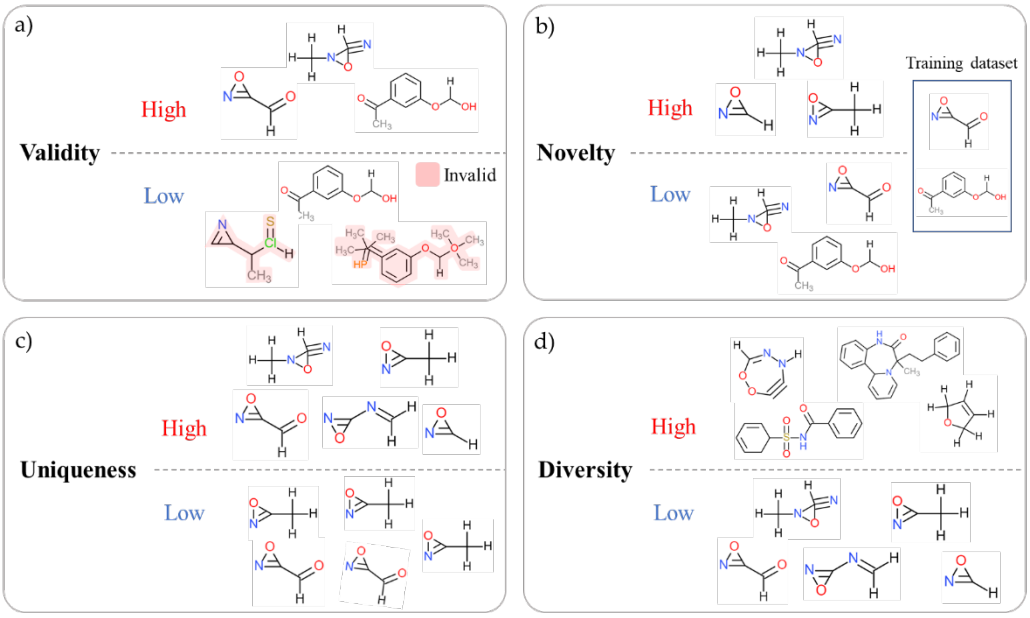

Figure 7. Generation metrics for de novo drug design. The four generation indices are commonly used in the de novo drug design studies: (**a**) validity, (**b**) novelty, (**c**) uniqueness, and (**d**) diversity. Unlike other metrics, a training dataset is required to measure novelty.

Validity index evaluates whether a generated compound can exist or not. For example, in the case of the SMILES expression method, if the grammar is not learned sufficiently, the valid molecules are not generated, or the parentheses do not match. Validity is the ratio of compounds that grammatically exist among all the compounds, and the closer to 1, the better the model. A higher validity may indicate a better model; however, from an industrial point of view of new drug development, low validity is not necessarily a problem. Even a model with low validity can increase the absolute number of valid compounds by securing a large population. This is because the additional cost of generating more compounds and filtering valid compounds in the VS stage is relatively low compared to the time and cost in other stages. Rather, if the novelty and uniqueness performance is lowered to increase the validity, it may not be suitable for the creation of new drug candidates.

Uniqueness is a number that determines whether the generator creates a new compound without duplication. Compared to other types of data such as images and sounds, a compound is a very discrete type of data. For this reason, even if a small change or noise is added to the input condition, the generated compound does not reflect the change, and the same compound may be created repeatedly. Uniqueness is evaluated by the number of generated products and the ratio of unique compounds with duplicates removed. If the uniqueness is 1, it means that all the generated compounds are different without duplicates.

While uniqueness measures the absence of overlap within the generated compound set, novelty measures the non-overlapping property by comparing the generated set with the existing dataset. That is, it evaluates whether the generator has created a new compound that does not exist in the training dataset. It is evaluated by the ratio of the subset with the training dataset compared to the generated compound. The closer the novelty is to 1, the more completely new compounds are created, and although it is used as an important indicator in the field of de novo drug design, it is not important if the existing compounds are also allowed.

$$\text{Validity}(N) = \frac{\#\ of\ valid\ compounds\ in\ N}{\#\ of\ compounds\ in\ N}, \quad (1)$$

$$\text{Uniqueness}(N) = \frac{\#\ of\ unique\ compounds\ in\ N}{\#\ of\ compounds\ in\ N}, \quad (2)$$

$$\text{Novelty}(N, T) = \frac{\#\ of\ intersection\ between\ N\ and\ T^c}{\#\ of\ compounds\ in\ N}. \quad (3)$$

N = generated compounds set, T = Training dataset, T^c = Complement set of T

Diversity or dissimilarity (or distance) is an indicator to determine how dissimilar and diverse the produced compounds are when only a few scaffolds or a small number of atoms are changed. Chemotype diversity can be measured as a value between 0 and 1 using scaled Shannon entropy. Similarity can also be measured using the distance between compounds the expressed in the FP or SMILES. As shown in Equations (4)–(6), the average distance within a set can be calculated using the Tanimoto coefficient. [183]

$$\text{Tanimoto}\ (x, y) = \left(\frac{x \cdot y^T}{x \cdot x^T + y \cdot y^T - x \cdot y^T}\right), \quad (4)$$

$$\text{Soergel}\ (x, y) = 1 - \text{Tanimoto}\ (x, y), \quad (5)$$

$$\text{Distance}\ (N) = \frac{2}{N_u^2} \sum_{i=1}^{N_u - 1} \sum_{j=i+1}^{N_u} \text{Soergel}\left(x_i^u, x_j^u\right). \quad (6)$$

Controllability is mainly used in the de novo models with the condition control functions [155,156]. It indicates how precisely the property value of the output compound is distributed for the input condition. Unlike other metrics, it is not expressed as a specific value, but is usually visualized using a histogram to evaluate the distribution compared to the target value. The smaller the variance, the better the performance.

6.3.2. Pharmacological Indicators

The pharmacological index measures whether the produced compounds have pharmacological effects, through a hypothetical method. Quantitative estimate of drug-likeness (QED), a representative pharmacological index, was introduced by Bickerton et al. [184], and it measures how similar the chemical properties of eight types of drugs are to those of the existing drug groups [185]: molecular weight (MW), lipophilicity (logP), number of hydrogen bond donors (HBD), number of hydrogen bond acceptors (HBA), polar surface area (PSA), number of rotatable bonds (ROTB), number of aromatic rings (AROM), and count of alerts for undesirable substructures. QED was inspired by Lipinski's rule and was standardized more quantitatively by including the insight. QED is widely used in the de novo drug design; however, usually, researchers evaluate only some of its metrics. Typically, the distributions of the MW and logP are often compared, and HBD and HBA are sometimes used. In particular, the MW and logP are often used to evaluate the control performance of a controllable de novo model [45] and can be considered to produce good performance when the variance is small.

Moreover, synthetic accessibility is measured to enhance the validity of real medicines from a practical industrial perspective. When building a model, synthetic accessibility can be optimized, or it can be filtered by removing compounds that are difficult to synthesize after being randomly generated. Currently, research is underway not only to measure composition difficulty through DL but also to propose suitable synthesis order. If this can be integrated, optimal drugs can be created from the early stages of new drug design considering molecular synthesis.

7. Limitation and Future Work

7.1. Current Challenges

7.1.1. Data Scarcity and Imbalance

The lack of labeled data is a major limitation to the use of DL-based drug discovery [186]. Data volumes resulting from drug discovery studies are small-scale because it requires expensive experiments and a long time to generate DTI data. For example, the most frequently used benchmark dataset for the DTI prediction is the Yamanishi_2008 dataset [187]. The dataset not only presents data on less than 1000 drugs, but also contains very limited DTI information with an average sparse rate of 3.6% [67].

Besides, the labeled data in drug discovery are extremely imbalanced. Since the HTS technique itself does not presuppose a high frequency of active responses, the HTS data consist of significantly fewer active responses than inactive responses. Consequently, there are often only a few validated drugs available for positive DTIs. In the PubChem Bioassay dataset, an active to inactive ratio of 1:40.92 (a hit rate of 2.385% of the total labeled activity values) indicates that most of the test results are inactive [188].

7.1.2. Absence of Standard Benchmark

In reality, the total number of drugs and proteins tested during the experiment is limited, making it imprecise to guarantee how a specific drug or target protein can work under the same experimental conditions. This problem is prominent in public databases that have accumulated data from the experimental results of numerous researchers around the world. However, big pharmaceutical companies can collect a large amount of data points by analyzing of constant conditions and well-characterized quality [133]. One research group built a model using the company's private data and the public ChEMBL [189] data and found that the predictive quality of the company model was higher than that of the public data model [115]. This demonstrates that the experimental conditions in the standardized datasets can affect the DNN prediction quality. Therefore, the necessity of data standardization and curation prior to building a predictive model are indispensable. Many public databases, including PubChem [190], ChEMBL, and ExCAPE-DB [191], aimed to standardize and integrate multiple-sourced datasets to facilitate computational

drug discovery. However, many DTI prediction models use only a small benchmarking dataset and use the train data and test data from the same source. This shows that many DTI models do not properly validate their generalization performance, demonstrating their inability to predict new DTIs in practical drug development.

7.2. Promising Method

7.2.1. Transfer Learning

As mentioned in the previous section, one of the biggest problems in drug discovery using AI is the lack of data. When targeting a specific disease or newly discovered target, the amount of data is so small that it is difficult to train. Moreover, it is difficult to easily apply augmentation to all the data. In such a situation, transfer learning is an excellent alternative [186,192]. Transfer learning, as part of lifelong learning, is inspired by how quickly humans acquire new knowledge from other similar experiences in the past. Transfer learning can improve many problems of insufficient data by fine-tuning a pre-trained model with a large dataset in another or a general field to an actual small-scale dataset [181]. Bonggun et al. [23] imported a molecule representation model learned from the PubChem database and applied it to their DTI model to improve performance. Panagiotis et al. reported that the transfer learning method exhibited improved performance in CHEMBL25 or DRD2 in the de novo study using conditional RNN [182].

Multi-task learning is also frequently used in drug discovery [186]. If transfer learning is to take the weights of a well-initialized DL model using a large dataset and use it for the target model, multi-task learning trains multiple tasks with many common parts at the same time (Figure 8). With multi-task learning, intrinsic features that are difficult to train with small datasets can be trained using different tasks. Steven et al. showed that using multi-task learning increased the AUC compared to the conventional random forest method or logistic regression method. When using multi-task learning, some datasets exhibited slightly decreased AUC, but for most datasets, AUC increased significantly. In particular, it is noteworthy that the performance of the datasets with a relatively smaller amount of data improved significantly. Using a pre-trained model improves the performance [47]; however, it has the advantage of significantly reducing the training time and computing power from an industrial and practical point of view. Therefore, we recommend using transfer learning for representation learning.

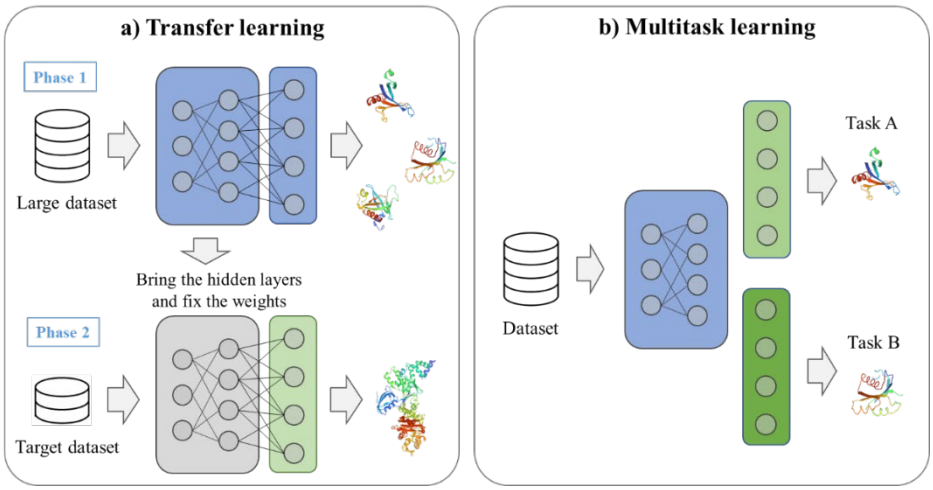

Figure 8. Simple example of transfer learning and multi-task learning.

7.2.2. Data Augmentation

There is a method of supplementing the data by incorporating small modifications in the existing data or changing the expression rule, which is called data augmentation. Data augmentation reduces model overfitting and improves the general performance. For data such as the voxel, a common image data augmentation method called geometric transformations can be applied [193]. Alternatively, there is a data augmentation method that adds a small amount of noise that does not affect the performance of the data. Isidro et al. [194] improved the predictive performance of the model by adding Gaussian noise to bioactivities and compound descriptors.

Another popular data augmentation method in drug discovery is randomized SMILES [146,182]. One compound can be written in various SMILES according to the starting point and direction. In the early stage of drug discovery using DL, a canonical SMILES was used for consistent expression; however, in the field of de novo drug design, randomized SMILES is used in a more general way [195]. Josep et al. [182] revealed that the quality of the generative model was better when using randomized SMILES than when using canonical SMILES. Randomized SMILES is mainly used for the de novo drug design [146]; however, Esben [30] showed that randomized SMILES trained more reliably and performed better than the canonical form even when predicting IC50. Unlike de novo, where the number of possible representations of a molecule is important, DTI requires information on the relationship between the ligand and target; therefore, it is not widely used.

7.2.3. Uncertainty and Interpretation

DL is a very powerful tool. It gives us hope that problems that were difficult to address using the classical ML methods can be solved with good performance if high-quality data are supplied abundantly. However, problems arose as the field of application of DL was expanded to a specialized area rather than an easy task. Since the parameters in the model are fixed and the operation process can be known, it is not actually a black box; however, it is treated as a black box because it is difficult for a human to interpret the process of deriving the result [196]. The non-transparency of this interpretation makes it difficult to accurately understand the reasoning process or an obstacle to decision-making. In particular, in areas such as drug discovery or disease diagnosis, where a wrong decision is costly and time consuming, sufficient evidence is needed to accept the result. Therefore, there is a growing need for explainable AI. An explainable AI review paper in the field of drug discovery by Jiménez-Luna et al. describes this well [197].

Although 'explainable' is defined in many ways, we will describe only two of the most commonly used concepts [198]. The first is 'uncertainty estimation'. Uncertainty can be thought of as the opposite of reliability of AI. In the case of the classification model, the weight for each class is output in the last layer, and the class with the highest value is selected using a function such as softmax. However, sometimes, the model outputs completely different results even with very small changes in the weight of the data or hidden layer. From this point of view, uncertainty can be interpreted as a measure of robustness against noise in the training process or model parameters when a certain result is output. Uncertainty leads researchers to make safer and more efficient decisions by estimating risks that will occur during drug development [199]. The second is 'interpretation'. Interpretation is often used interchangeably with 'transparency' depending on the paper [3,198]. Xuhong et al. [64] redefined 'interpretability' as follows in their paper: "The model interpretability is the ability (of the model) to explain or to present in understandable terms to a human." The initial concept of an interpretable DL model was to create class activation maps [200] from the convolution layer of the CNN to visualize the reason for prediction by matching the input result. In the recent drug discovery field, attention-based explainable models dominate. The increased use of attention-based models such as the transformer is also because the performance is better than the other methods at sequential

data; however, the reason can be inferred indirectly from attention. Gao et al. [119] created an attention matrix from the results of embedded protein (LSTM) and molecule (GCN). The attention matrix visualized contributing weights of atoms in molecule and residues that affect the DTI, thereby helping researchers to understand the process in a transparent manner and gain new insights. As a solubility prediction method, but not that of DTI, Karpov et al. [77] used a transformer-CNN model from the SMILES data, and Liu et al. [201] used the GCN from a molecule graph to predict the positive or negative contribution of the atoms to solubility. Chen et al. created a model to interpret the atoms contributing to the interaction in the prediction of the DDI [80].

The advantage of interpretability is that it gives the researchers confidence in the results. When the reason for drawing a conclusion is consistent with prior knowledge, the expert can accept the decision with high confidence [3]. It can also provide new inductive inspiration to experts [198]. Finally, it can provide another channel to discover problems when the performance of the DL models is poor.

Author Contributions: Note that this is a review article. Investigation, visualization, and writing—original draft preparation, S.P., J.K. Writing—review and editing, S.P., J.K., D.M., W.K. Supervision—D.M., W.K. All authors have read and agreed to the published version of the manuscript.

Funding: WK was funded by National Research Foundation of South Korea (2017M3C9A5028690).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AI	Artificial Intelligence
ADMET	Absorption, distribution, metabolism, excretion, and toxicity
AUC	Area Under the Curve
AUPR	Area Under the Precision–Recall Curve
AE	Autoencoder
CNN	Convolutional Neural Networks
CI	Concordance Index
DDI	Drug–Drug Interaction
MAE	Mean Absolute Error
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MLP	Multi-Layer Perceptron
DL	Deep learning
DTA	Drug–Target Affinity
DTI	Drug–Target Interaction
DTP	Drug–Target Pair
FP	Fingerprint
FPR	False Positive Rate
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
HTS	High-Throughput Screening
GAN	Generative Adversarial Networks
GCN	Graph Convolutional Networks
GO	Gene Ontology
LINCS	Library of Integrated Network-based Cellular Signatures
LSTM	Long Short-Term Memory
PPI	Protein–Protein Interaction
QSAR	Quantitative Structure-Activity Relationship

RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks
SMILES	Simplified Molecular-Input Line-Entry System
TPR	True Positive Rate
TPSA	Topological Polar Surface Area
VAE	Variational AutoEncoder
VS	Virtual Screening

Appendix A

Table A1. Ligand-based DL methods for DTI prediction.

Reference	Models	Input Drug Type	Datasets	Algorithm Type	Year	Evaluation Metrics
Gao et al. [33]	MLP; Multi-task	Fingerprint (ECFP; FP2; Estate1; Estate2; MACCS; ERG)	PDBbind	Regression	2019	Pearson correlation coefficient (R); RMSE
Wenzel et al. [115]	MLP; Multi-task	Atom pair; pharmacophoric donor–acceptor pairs;	ChEMBL	Regression	2019	R^2
Xie et al. [32]	MLP; LSTM	Fingerprint (MACCS+ECFP)	DrugBank; ChEMBL; PDBbind	Regression	2020	Pearson correlation coefficient (R); RMSE
Hirohara et al. [25]	CNN	SMILES convolution fingerprint	Tox21	Classification	2018	AUC
Matsuzaka et al. [109]	CNN	2D image	Tox21	Classification	2019	AUC; Balanced accuracy; F-score; MCC
Rifaioglu et al. [19]	CNN	2D image	ChEMBL; MUV; DUD-E	Classification	2020	AUC; Accuracy; Precision; Recall; F1-score; MCC
Liu et al. [81]	GCN; Multi-task	3D molecular graph	Amgen's internal dataset; ChEMBL	Regression	2019	R^2; Accuracy
Yang et al. [21]	GCN	SMILES	PDBbind; ChEMBL; PubChem Bioassay; MUV; Tox21; ToxCast; SIDER etc.	Classification; Regression	2019	MAE; RMSE; AUC; AUPR
Shang et al. [11]	GCN; Attention-based	Molecular graph	Tox21; HIV; Freesolv; Lipophilicity (MoleculeNet)	Regression	2018	AUC; RMSE

Table A2. Structure-based DL methods for DTI prediction.

Reference	Models	Input Drug Type	Input Target Type	Datasets	Algorithm Type	Evaluation Metrics	Year
Wen et al. [37]	MLP	Fingerprint (ECFP)	PSC (protein sequence composition descriptor)	DrugBank	Classification	TPR; TNR; Accuracy; AUC	2017
Chen et al. [57]	MLP	Fingerprint (PubChemFP)	Various protein features *	DrugBank; Yamanishi	Classification	AUC; AUPR	2020
Öztürk et al. [24]	CNN	SMILES	Sequence	Davis; KIBA	Regression	CI; MSE	2018
Shin et al. [23]	CNN; attention	SMILES	Sequence	Davis; KIBA;	Regression	CI; RMSE; r_m^2; AUPR	2019
Zhao et al. [120]	CNN; attention	SMILES	Sequence	Davis; KIBA	Regression	CI; RMSE; r_m^2; AUPR	2019
Gonczarek et al. [202]	CNN	Atom pair	Atom pair	DUD-E; PDBBind	Regression	AUC	2016
Ragoza et al. [203]	CNN	Voxel	Voxel	DUD-E; CSAR	Regression; Classification	AUC	2017
Jiménez et al. [204]	CNN	Voxel	Voxel	PDBbind; CSAR2012	Regression	RMSE; R^2	2018
Kwon et al. [75]	CNN	Voxel	Voxel	CASF-2016 [205]	Regression	MAE; RMSE	2020
Pu et al. [51]	CNN; multi-classification	Voxel	Voxel	PDB; TOUGH-M1 [206]	Classification	MCC; AUC; Accuracy	2019
Lee et al. [31]	CNN	Fingerprint	Sequence	DrugBank; KEGG; IUPHAR; MATADOR; PubChem Bioassay; KinaseSARfari [189]	Classification	AUC; AUPR; Sensitivity; Specificity; Precision; Accuracy; F1-score	2019
Hasan Mahmud et al. [207]	CNN	SMILES; 193 features by Rcpi	Sequence; 1290 features by PROFEAT	DrugBank; Yamanishi	Regression	AUC; Accuracy; Sensitivity; Precision; F1 score; AUPR	2020
Wang et al. [34]	LSTM	Fingerprint (PubChemFP)	PSSM; Legendre Moment [208]	DrugBank; Yamanishi; KEGG; SuperTarget	Classification	AUC; Accuracy; TPR; Specificity; Precision; MCC	2020
Tsubaki et al. [209]	GNN; CNN; attention	Fingerprint (PubChemFP)	Sequence; Pfam domain	DUD-E; DrugBank; MATADOR	Classification	AUC; Precision; Recall	2019
Torng and Altman [118]	GCN	Molecular graph	Molecular graph	DUD-E; MUV	Classification	AUC	2019

Feng et al. [8]	GCN	Fingerprint (ECFP); 3D molecular graph	PSC (protein sequence composition descriptor)	Davis; Metz; KIBA; ToxCast	Regression	R^2	2019
Jiang et al. [210]	GNN	3D molecular graph	3D molecular graph	KIBA; Davis	Regression	r_m^2; CI; MSE; Pearson correlation coefficient; Accuracy	2020

* CTD; CT; Pseudo AAC; Pseudo PSSM; NMBroto; Structure feature from SPIDER.

Table A3. Relationship-based DL methods for DTI prediction.

Reference	Models	Relationship Data Type	Datasets	Algorithm Type	Evaluation Metrics	Year
Xie et al. [71]	MLP	LINCS signature	DrugBank; CTD; DGIdb; STITCH	Regression	Accuracy; F-score; TPR	2018
Lee and Kim [63]	MLP; node2vec	LINCS signature; PPI (Protein-protein interaction); Pathway	LINCS; ChEMBL; TTD; MATADOR; KEGG; IUPHAR; PharmGKB; KiDB	Classification	AUC; Precision	2019
Gao et al. [119]	CNN; LSTM	LINCS signature; GO term	BindingDB	Regression	Accuracy; AUC; AUPR	2018
Shao and Zhang [147]	CNN; GCN	LINCS signature	LINCS; DrugBank	Classification	Accuracy; AUC	2020
Thafar et al. [67]	node2vec	Drug similarity (structure, side effects); Target similarity (sequence, GO); PPI	Yamanishi; KEGG; BRENDA; SuperTarget; DrugBank; BioGRID; SIDER	Classification	AUPR; AUC	2020
Zong et al. [13]	DeepWalk [130]	Drug-target association; Drug-disease association; Disease-target association	DrugBank; Human diseasome [211]	Classification	AUC	2017
Mongia and Majumdar [212]	Multi-graph deep matrix factorization	Drug similarity (structure); Target similarity (sequence)	Yamanishi; KEGG; BRENDA; SuperTarget; DrugBank	Classification	AUPR; AUC	2020
Wang et al. [59]	AE	Drug similarity (structure, side effects); Target similarity (sequence, GO); PPI	Yamanishi; KEGG; BRENDA; SuperTarget; DrugBank; SIDER	Classification	AUPR; AUC	2020
Zhao et al. [68]	CNN; AE	Drug similarity (structure); Target similarity (sequence); PPI	DrugBank; STRING	Classification	Accuracy; AUPR; AUC	2020
Peng et al. [97]	CNN; AE	Drug-target association; Drug-disease association; Disease-target association; Drug similarity (structure, side effects); Target similarity (sequence, GO); PPI	DrugBank; Human Protein Reference Database [2009]; CTD; SIDER;	Classification	AUPR; AUC	2020

Zhong et al. [213]	GCN	LINCS signature; PPI	ChEMBL; LINCS; STRING	Classification	Accuracy; F-score; AUPR; Precision; Recall; AUC	2020

Table A4. Benchmark datasets for DTIs.

Dataset	No. of DTIs	No. of Target	No. of Drug	Year	Availability*
PharmGKB [214]	777	1030	4078	2020	https://www.pharmgkb.org/downloads
Yamanishi [215]	5127	989	932	2008	https://members.cbio.mines-paristech.fr/~yyamanishi/pharmaco/
DrugBank [216]	6566	4844	18,734	2020	https://go.drugbank.com/releases/latest
IUPHAR [217]	6605	1577	14,981	2020	https://www.guidetopharmacology.org/download.jsp
SuperTarget/MATADOR [218]	8936	1799	719	2008	http://matador.embl.de/
DGIdb [219]	11,137	3820	58,555	2020	https://www.dgidb.org/downloads
CTD [220]	17,814	46,364	2,521,525	2020	http://ctdbase.org/downloads/
TTD [221]	18,351	1814	29,388	2020	http://db.idrblab.net/ttd/full-data-download
Davis [113]	27,621	379	68	2011	https://tdcommons.ai/multi_pred_tasks/dti/#davis
Tox21	77,946	12	7831	2014	https://deepchemdata.s3-us-west-1.amazonaws.com/datasets/tox21.csv.gz
Metz [129]	103,920	172	3858	2011	https://www.nature.com/articles/nchembio.530
KIBA [222]	118,036	229	2068	2014	https://tdcommons.ai/multi_pred_tasks/dti/#kiba
MUV [168]	249,886	17	93,087	2009	https://deepchemdata.s3-us-west-1.amazonaws.com/datasets/muv.csv.gz
BindingDB [223]	456,248	3716	747,066	2020	https://www.bindingdb.org/bind/
ToxCast [224]	530,605	335	7675	2007	https://www.epa.gov/chemical-research
ExCAPE-DB [191]	582,724	1667	1,361,473	2017	https://solr.ideaconsult.net/search/excape/
DUD-E [166]	1,167,186	102	22,886	2012	http://dude.docking.org/

* Site accessed date: 14 September, 2021.

References

1. Reddy, A.S.; Zhang, S. Polypharmacology: Drug discovery for the future. *Expert Rev. Clin. Pharmacol.* 2013, *6*, 41–47.
2. Sachdev, K.; Gupta, M.K. A comprehensive review of feature based methods for drug target interaction prediction. *J. Biomed. Inform.* **2019**, *93*, 103159, doi:10.1016/j.jbi.2019.103159.
3. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* 2019, *18*, 463–477.
4. Kimber, T.B.; Chen, Y.; Volkamer, A. Deep learning in virtual screening: Recent applications and developments. *Int. J. Mol. Sci.* **2021**, *22*, doi:10.3390/ijms22094435.
5. Lipinski, C.F.; Maltarollo, V.G.; Oliveira, P.R.; Silva, A.B.F. da; Honorio, K.M. Advances and Perspectives in Applying Deep Learning for Drug Design and Discovery. *Front. Robot. AI* **2019**, *6*, 108, doi:10.3389/FROBT.2019.00108.

6. Rifaioglu, A.S.; Atas, H.; Martin, M.J.; Cetin-Atalay, R.; Atalay, V.; Doğan, T. Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools and databases. *Brief. Bioinform.* 2019, *20*, 1878–1912.
7. Zhavoronkov, A.; Ivanenkov, Y.A.; Aliper, A.; Veselov, M.S.; Aladinskiy, V.A.; Aladinskaya, A. V.; Terentiev, V.A.; Polykovskiy, D.A.; Kuznetsov, M.D.; Asadulaev, A.; et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **2019**, *37*, 1038–1040, doi:10.1038/s41587-019-0224-x.
8. Feng, Q.; Dueva, E.; Cherkasov, A.; Ester, M. PADME: A Deep Learning-based Framework for Drug-Target Interaction Prediction;
9. Skalic, M.; Varela-Rial, A.; Jiménez, J.; Martínez-Rosell, G.; De Fabritiis, G. LigVoxel: Inpainting binding pockets using 3D-convolutional neural networks. *Bioinformatics* **2019**, *35*, 243–250, doi:10.1093/bioinformatics/bty583.
10. Winter, R.; Montanari, F.; Noé, F.; Clevert, D.A. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chem. Sci.* **2019**, *10*, 1692–1701, doi:10.1039/c8sc04175j.
11. Shang, C.; Liu, Q.; Chen, K.-S.; Sun, J.; Lu, J.; Yi, J.; Bi, J. Edge Attention-based Multi-Relational Graph Convolutional Networks. **2018**, 27–29.
12. Zeng, X.; Zhu, S.; Lu, W.; Liu, Z.; Huang, J.; Zhou, Y.; Fang, J.; Huang, Y.; Guo, H.; Li, L.; et al. Target identification among known drugs by deep learning from heterogeneous networks. *Chem. Sci.* **2020**, *11*, 1775–1797, doi:10.1039/c9sc04336e.
13. Zong, N.; Kim, H.; Ngo, V.; Harismendy, O. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. *Bioinformatics* **2017**, *33*, 2337–2344, doi:10.1093/bioinformatics/btx160.
14. Husic, B.E.; Charron, N.E.; Lemm, D.; Wang, J.; Pérez, A.; Majewski, M.; Krämer, A.; Chen, Y.; Olsson, S.; De Fabritiis, G.; et al. Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys.* **2020**, *153*, doi:10.1063/5.0026133.
15. Hassan-Harrirou, H.; Zhang, C.; Lemmin, T. RosENet: Improving Binding Affinity Prediction by Leveraging Molecular Mechanics Energies with an Ensemble of 3D Convolutional Neural Networks. *J. Chem. Inf. Model.* **2020**, *60*, 2791–2802, doi:10.1021/acs.jcim.0c00075.
16. Gentile, F.; Agrawal, V.; Hsing, M.; Ton, A.-T.; Ban, F.; Norinder, U.; E. Gleave, M.; Cherkasov, A. Deep Docking: A Deep Learning Platform for Augmentation of Structure Based Drug Discovery. *ACS Cent. Sci.* **2020**, *6*, 939–949, doi:10.1021/acscentsci.0c00229.
17. Xue, L.; Bajorath, J. Molecular Descriptors in Chemoinformatics, Computational Combinatorial Chemistry, and Virtual Screening. *Comb. Chem. High Throughput Screen.* **2012**, *3*, 363–372, doi:10.2174/1386207003331454.
18. Redkar, S.; Mondal, S.; Joseph, A.; Hareesha, K.S. A Machine Learning Approach for Drug-target Interaction Prediction using Wrapper Feature Selection and Class Balancing. *Mol. Inform.* **2020**, *39*, doi:10.1002/minf.201900062.
19. Rifaioglu, A.S.; Atalay, V.; Martin, M.J.; Cetin-Atalay, R.; Doğan, T. DEEPScreen: High performance drug-target interaction prediction with convolutional neural networks using 2-D structural compound representations. *bioRxiv* 2018, 491365.
20. David, L.; Thakkar, A.; Mercado, R.; Engkvist, O. Molecular representations in AI-driven drug discovery: a review and practical guide. *J. Cheminform.* **2020**, *12*, 1–22, doi:10.1186/s13321-020-00460-5.
21. Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; et al. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59*, 3370–3388, doi:10.1021/acs.jcim.9b00237.
22. Stokes, J.M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N.M.; MacNair, C.R.; French, S.; Carfrae, L.A.; Bloom-Ackerman, Z.; et al. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *180*, 688-702.e13, doi:10.1016/j.cell.2020.01.021.
23. Shin, B.; Park, S.; Kang, K.; Ho, J.C. Self-Attention Based Molecule Representation for Predicting Drug-Target Interaction. *Proc. Mach. Learn. Res.* **2019**, *106*, 1–18.
24. Öztürk, H.; Özgür, A.; Ozkirimli, E. DeepDTA: Deep drug-target binding affinity prediction. *Bioinformatics* **2018**, *34*, i821–i829, doi:10.1093/bioinformatics/bty593.
25. Hirohara, M.; Saito, Y.; Koda, Y.; Sato, K.; Sakakibara, Y. Convolutional neural network based on SMILES representation of compounds for detecting chemical motif. *BMC Bioinformatics* **2018**, *19*, doi:10.1186/s12859-018-2523-5.
26. Tetko, I. V.; Karpov, P.; Van Deursen, R.; Godin, G. State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nat. Commun.* **2020**, *11*, 1–11, doi:10.1038/s41467-020-19266-y.
27. Liu, B.; Ramsundar, B.; Kawthekar, P.; Shi, J.; Gomes, J.; Luu Nguyen, Q.; Ho, S.; Sloane, J.; Wender, P.; Pande, V. Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models. *ACS Cent. Sci.* **2017**, *3*, 1103–1113, doi:10.1021/acscentsci.7b00303.
28. Bai, R.; Zhang, C.; Wang, L.; Yao, C.; Ge, J.; Duan, H. molecules Transfer Learning: Making Retrosynthetic Predictions Based on a Small Chemical Reaction Dataset Scale to a New Level. *Molecules* **2020**, *25*, doi:10.3390/molecules25102357.
29. Arús-Pous, J.; Johansson, S.V.; Prykhodko, O.; Bjerrum, E.J.; Tyrchan, C.; Reymond, J.L.; Chen, H.; Engkvist, O. Randomized SMILES strings improve the quality of molecular generative models. *J. Cheminform.* **2019**, *11*, 1–13, doi:10.1186/s13321-019-0393-0.
30. Bjerrum, E.J. SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules. **2017**.
31. Lee, I.; Keum, J.; Nam, H. DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Comput. Biol.* **2019**, *15*, 1–21, doi:10.1371/journal.pcbi.1007129.
32. Xie, L.; Xu, L.; Kong, R.; Chang, S.; Xu, X. Improvement of Prediction Performance With Conjoint Molecular Fingerprint in Deep Learning. *Front. Pharmacol.* **2020**, *11*, 1–15, doi:10.3389/fphar.2020.606668.

33. Gao, K.; Duy Nguyen, D.; Sresht, V.; Mathiowetz, A.M.; Tu, M.; Wei, G.-W. *Are 2D fingerprints still valuable for drug discovery?*; 2019;
34. Wang, Y. Bin; You, Z.H.; Yang, S.; Yi, H.C.; Chen, Z.H.; Zheng, K. A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 1–9, doi:10.1186/s12911-020-1052-0.
35. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213, doi:10.1093/nar/gkv951.
36. Morgan, H.L. The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113, doi:10.1021/c160017a018.
37. Wen, M.; Zhang, Z.; Niu, S.; Sha, H.; Yang, R.; Yun, Y.; Lu, H. Deep-Learning-Based Drug–Target Interaction Prediction. **2017**, doi:10.1021/acs.jproteome.6b00618.
38. Moumbock, A.F.A.; Li, J.; Mishra, P.; Gao, M.; Günther, S. Current computational methods for predicting protein interactions of natural products. *Comput. Struct. Biotechnol. J.* 2019, *17*, 1367–1376.
39. Wood, D.J.; De Vlieg, J.; Wagener, M.; Ritschel, T. Pharmacophore Fingerprint-Based Approach to Binding Site Subpocket Similarity and Its Application to Bioisostere Replacement. **2012**, doi:10.1021/ci3000776.
40. Zhang, Y.F.; Wang, X.; Kaushik, A.C.; Chu, Y.; Shan, X.; Zhao, M.Z.; Xu, Q.; Wei, D.Q. SPVec: A Word2vec-Inspired Feature Representation Method for Drug-Target Interaction Prediction. *Front. Chem.* **2020**, *7*, 1–11, doi:10.3389/fchem.2019.00895.
41. Goh, G.B.; Hodas, N.O.; Siegel, C.; Vishnu, A. SMILES2Vec: An Interpretable General-Purpose Deep Neural Network for Predicting Chemical Properties. **2017**, doi:10.475/123.
42. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed representations ofwords and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **2013**, 1–9.
43. Asgari, E.; Mofrad, M.R.K. Continuous distributed representation of biological sequences for deep proteomics and genomics. *PLoS One* **2015**, *10*, 1–15, doi:10.1371/journal.pone.0141287.
44. Jaeger, S.; Fulle, S.; Turk, S. Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. *J. Chem. Inf. Model.* **2018**, *58*, 27–35, doi:10.1021/acs.jcim.7b00616.
45. Lim, J.; Ryu, S.; Kim, J.W.; Kim, W.Y. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *J. Cheminform.* **2018**, *10*, 1–9, doi:10.1186/s13321-018-0286-7.
46. Xue, D.; Zhang, H.; Xiao, D.; Gong, Y.; Chuai, G.; Sun, Y.; Tian, H. X-MOL : large-scale pre-training for molecular understanding and diverse molecular analysis. **2021**.
47. Li, P.; Wang, J.; Qiao, Y.; Chen, H.; Yu, Y. Learn molecular representations from large-scale unlabeled molecules for drug discovery. 1–20.
48. Kuzminykh, D.; Polykovskiy, D.; Kadurin, A.; Zhebrak, A.; Baskov, I.; Nikolenko, S.; Shayakhmetov, R.; Zhavoronkov, A. 3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks. *Mol. Pharm.* **2018**, *15*, 4378–4385, doi:10.1021/acs.molpharmaceut.7b01134.
49. Li, Z.; Yang, S.; Song, G.; Cai, L. HamNet: Conformation-Guided Molecular Representation with Hamiltonian Neural Networks. **2021**, 1–11.
50. Amidi, A.; Amidi, S.; Vlachakis, D.; Megalooikonomou, V.; Paragios, N.; Zacharaki, E.I. EnzyNet: Enzyme classification using 3D convolutional neural networks on spatial representation. *PeerJ* **2018**, *2018*, 1–11, doi:10.7717/peerj.4750.
51. Pu, L.; Govindaraj, R.G.; Lemoine, J.M.; Wu, H.-C.; Brylinski, M. DeepDrug3D: Classification of ligand-binding pockets in proteins with a convolutional neural network. *PLOS Comput. Biol.* **2019**, *15*, e1006718, doi:10.1371/JOURNAL.PCBI.1006718.
52. Gainza, P.; Sverrisson, F.; Monti, F.; Rodolà, E.; Boscaini, D.; Bronstein, M.M.; Correia, B.E. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* **2020**, *17*, 184–192, doi:10.1038/s41592-019-0666-6.
53. Wang, Y.; Wu, S.; Duan, Y.; Huang, Y. A Point Cloud-Based Deep Learning Strategy for Protein-Ligand Binding Affinity Prediction.
54. Lim, J.; Ryu, S.; Park, K.; Choe, Y.J.; Ham, J.; Kim, W.Y. Predicting Drug-Target Interaction Using a Novel Graph Neural Network with 3D Structure-Embedded Graph Representation. *J. Chem. Inf. Model.* **2019**, *59*, 3981–3988, doi:10.1021/acs.jcim.9b00387.
55. Coley, C.W.; Barzilay, R.; Green, W.H.; Jaakkola, T.S.; Jensen, K.F. Convolutional Embedding of Attributed Molecular Graphs for Physical Property Prediction.
56. Zhu, L.; Davari, M.D.; Li, W. Recent advances in the prediction of protein structural classes: Feature descriptors and machine learning algorithms. *Crystals* **2021**, *11*, 1–16, doi:10.3390/cryst11040324.
57. Chen, C.; Shi, H.; Han, Y.; Jiang, Z.; Cui, X.; Yu, B. DNN-DTIs: Improved drug-target interactions prediction using XGBoost feature selection and deep neural network. *bioRxiv* **2020**, doi:10.1101/2020.08.11.247437.
58. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402, doi:10.1093/NAR/25.17.3389.
59. Wang, H.; Wang, J.; Dong, C.; Lian, Y.; Liu, D.; Yan, Z. A novel approach for drug-target interactions prediction based on multimodal deep autoencoder. *Front. Pharmacol.* **2020**, *10*, 1–19, doi:10.3389/fphar.2019.01592.
60. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: tool for the unification of biology. *Nat. Genet. 2000 251* **2000**, *25*, 25–29, doi:10.1038/75556.
61. Kuhlman, B.; Bradley, P. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol. 2019 2011* **2019**, *20*, 681–697, doi:10.1038/s41580-019-0163-x.

62. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nat. 2021* **2021**, 1–11, doi:10.1038/s41586-021-03819-2.
63. Lee, H.; Kim, W. Comparison of target features for predicting drug-target interactions by deep neural network based on large-scale drug-induced transcriptome data. *Pharmaceutics* **2019**, *11*, doi:10.3390/pharmaceutics11080377.
64. Li, X.; Xiong, H.; Li, X.; Wu, X.; Zhang, X.; Liu, J.; Bian, J.; Dou, D. Interpretable Deep Learning: Interpretation, Interpretability, Trustworthiness, and Beyond. **2021**.
65. A, L.; C, B.; H, T.; M, G.; JP, M.; P, T. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425, doi:10.1016/J.CELS.2015.12.004.
66. Yang, F.; Fan, K.; Song, D.; Lin, H. Graph-based prediction of Protein-protein interactions with attributed signed graph embedding. *BMC Bioinformatics* **2020**, *21*, 1–16, doi:10.1186/s12859-020-03646-8.
67. Thafar, M.A.; Thafar, M.A.; Olayan, R.S.; Olayan, R.S.; Ashoor, H.; Ashoor, H.; Albaradei, S.; Albaradei, S.; Bajic, V.B.; Gao, X.; et al. DTiGEMS+: Drug-target interaction prediction using graph embedding, graph mining, and similarity-based techniques. *J. Cheminform.* **2020**, *12*, 1–17, doi:10.1186/s13321-020-00447-2.
68. Zhao, Y.; Zheng, K.; Guan, B.; Guo, M.; Song, L.; Gao, J.; Qu, H.; Wang, Y.; Shi, D.; Zhang, Y. DLDTI: a learning-based framework for drug-target interaction identification using neural networks and network representation. *J. Transl. Med.* **2020**, *18*, doi:10.1186/s12967-020-02602-7.
69. J, L.; ED, C.; D, P.; JW, M.; IC, B.; MJ, W.; J, L.; JP, B.; A, S.; KN, R.; et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006**, *313*, 1929–1935, doi:10.1126/SCIENCE.1132939.
70. A, S.; R, N.; SM, C.; DD, P.; TE, N.; X, L.; J, G.; JF, D.; AA, T.; JK, A.; et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **2017**, *171*, 1437-1452.e17, doi:10.1016/J.CELL.2017.10.049.
71. Xie, L.; He, S.; Song, X.; Bo, X.; Zhang, Z. Deep learning-based transcriptome data classification for drug-target interaction prediction. *BMC Genomics* **2018**, *19*, 13–16, doi:10.1186/s12864-018-5031-0.
72. Zhu, J.; Wang, J.; Wang, X.; Gao, M.; Guo, B.; Gao, M.; Liu, J.; Yu, Y.; Wang, L.; Kong, W.; et al. Prediction of drug efficacy from transcriptional profiles with deep learning. *Nat. Biotechnol.* **2021**, doi:10.1038/s41587-021-00946-z.
73. Korkmaz, S. Deep learning-based imbalanced data classification for drug discovery. *J. Chem. Inf. Model.* **2020**, *60*, 4180–4190, doi:10.1021/acs.jcim.9b01162.
74. Townshend, R.J.L.; Powers, A.; Eismann, S.; Derry, A. ATOM3D: Tasks On Molecules in Three Dimensions. *arXiv* **2021**.
75. Kwon, Y.; Shin, W.H.; Ko, J.; Lee, J. AK-score: Accurate protein-ligand binding affinity prediction using an ensemble of 3D-convolutional neural networks. *Int. J. Mol. Sci.* **2020**, *21*, 1–16, doi:10.3390/ijms21228424.
76. Huang, K.; Fu, T.; Glass, L.M.; Zitnik, M.; Xiao, C.; Sun, J. DeepPurpose: a deep learning library for drug–target interaction prediction. *Bioinformatics* **2020**, 1–3, doi:10.1093/bioinformatics/btaa1005.
77. Karpov, P.; Godin, G.; Tetko, I. V. Transformer-CNN: Swiss knife for QSAR modeling and interpretation. *J. Cheminform.* **2020**, *12*, 17, doi:10.1186/s13321-020-00423-w.
78. Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The graph neural network model. *IEEE Trans. Neural Networks* **2009**, *20*, 61–80, doi:10.1109/TNN.2008.2005605.
79. Sun, M.; Zhao, S.; Gilvary, C.; Elemento, O.; Zhou, J.; Wang, F. Graph convolutional networks for computational drug development and discovery. *Brief. Bioinform.* **2020**, *21*, 919–935, doi:10.1093/BIB/BBZ042.
80. Chen, X.; Liu, X.; Wu, J. GCN-BMP: Investigating graph representation learning for DDI prediction task. *Methods* **2020**, *179*, 47–54, doi:10.1016/j.ymeth.2020.05.014.
81. Liu, K.; Sun, X.; Jia, L.; Ma, J.; Xing, H.; Wu, J.; Gao, H.; Sun, Y.; Boulnois, F.; Fan, J. Chemi-net: A molecular graph convolutional network for accurate drug property prediction. *Int. J. Mol. Sci.* **2019**, *20*, doi:10.3390/ijms20143389.
82. Long, Y.; Wu, M.; Liu, Y.; Kwoh, C.K.; Luo, J.; Li, X. Ensembling graph attention networks for human microbe-drug association prediction. *Bioinformatics* **2020**, *36*, I779–I786, doi:10.1093/bioinformatics/btaa891.
83. Zhang, J.; Jiang, Z.; Hu, X.; Song, B. A novel graph attention adversarial network for predicting disease-related associations. *Methods* **2020**, *179*, 81–88, doi:10.1016/J.YMETH.2020.05.010.
84. Lim, J.; Hwang, S.Y.; Moon, S.; Kim, S.; Kim, W.Y. Scaffold-based molecular design with a graph generative model. *Chem. Sci.* **2020**, *11*, 1153–1164, doi:10.1039/c9sc04503a.
85. Hochreiter, S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertainty, Fuzziness Knowlege-Based Syst.* **1998**, *6*, 107–116, doi:10.1142/S0218488598000094.
86. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: Lstm cells and network architectures. *Neural Comput.* **2019**, *31*, 1235–1270, doi:10.1162/NECO_A_01199.
87. Mouchlis, V.D.; Afantitis, A.; Serra, A.; Fratello, M.; Papadiamantis, A.G.; Aidinis, V.; Lynch, I.; Greco, D.; Melagraki, G. Advances in de novo drug design: From conventional to machine learning methods. *Int. J. Mol. Sci.* **2021**, *22*, 1–22, doi:10.3390/ijms22041676.
88. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.* **2014**, 1724–1734, doi:10.3115/v1/d14-1179.
89. Jastrzębski, S.; Leśniak, D.; Czarnecki, W.M. Learning to SMILE(S). **2016**.

90. Guimaraes, G.L.; Sanchez-Lengeling, B.; Outeiral, C.; Farias, P.L.C.; Aspuru-Guzik, A. Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models. **2017**.
91. De Cao, N.; Kipf, T. MolGAN: An implicit generative model for small molecular graphs. **2018**.
92. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.* **2019**, *1*, 4171–4186.
93. Mark Lennox, Neil M. Robertson, B.D. Modelling Drug-Target Binding Affinity using a BERT based Graph Neural network. *Annu. Rev. Biochem.* **2021**, *68*, 559–581.
94. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144, doi:10.1145/3422622.
95. Eugene Lin, Chieh-Hsin Lin, and H.-Y.L. Relevant Applications of Generative Adversarial Networks in Drug Design and Discovery : *Molecules* **2020**, *25*, 1–25.
96. Yu, L.; Zhang, W.; Wang, J.; Yu, Y. SeqGAN: Sequence generative adversarial nets with policy gradient. *31st AAAI Conf. Artif. Intell. AAAI 2017* **2017**, 2852–2858.
97. Peng, J.; Li, J.; Shang, X. A learning-based method for drug-target interaction prediction based on feature representation learning and deep neural network. *BMC Bioinforma. 2020 2113* **2020**, *21*, 1–13, doi:10.1186/S12859-020-03677-1.
98. Gómez-Bombarelli, R.; Wei, J.N.; Duvenaud, D.; Hernández-Lobato, J.M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T.D.; Adams, R.P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276, doi:10.1021/acscentsci.7b00572.
99. Makhzani, A.; Shlens, J.; Jaitly, N.; Goodfellow, I.; Frey, B. Adversarial Autoencoders. **2015**.
100. Vanhaelen, Q.; Lin, Y.C.; Zhavoronkov, A. The Advent of Generative Chemistry. *ACS Med. Chem. Lett.* **2020**, *11*, 1496–1505, doi:10.1021/acsmedchemlett.0c00088.
101. Kadurin, A.; Aliper, A.; Kazennov, A.; Mamoshina, P.; Vanhaelen, Q.; Khrabrov, K.; Zhavoronkov, A. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* **2017**, *8*, 10883–10890, doi:10.18632/oncotarget.14073.
102. Kadurin, A.; Nikolenko, S.; Khrabrov, K.; Aliper, A.; Zhavoronkov, A. DruGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. *Mol. Pharm.* **2017**, *14*, 3098–3104, doi:10.1021/acs.molpharmaceut.7b00346.
103. Polykovskiy, D.; Zhebrak, A.; Vetrov, D.; Ivanenkov, Y.; Aladinskiy, V.; Mamoshina, P.; Bozdaganyan, M.; Aliper, A.; Zhavoronkov, A.; Kadurin, A. Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery. *Mol. Pharm.* **2018**, *15*, 4398–4405, doi:10.1021/acs.molpharmaceut.8b00839.
104. J, V.; M, L.; E, G.; E, H.; FJ, L. Merging Ligand-Based and Structure-Based Methods in Drug Discovery: An Overview of Combined Virtual Screening Approaches. *Molecules* **2020**, *25*, doi:10.3390/MOLECULES25204723.
105. Abbasi, K.; Razzaghi, P.; Poso, A.; Ghanbari-Ara, S.; Masoudi-Nejad, A. Deep Learning in Drug Target Interaction Prediction: Current and Future Perspectives. *Curr. Med. Chem.* **2020**, *28*, 2100–2113, doi:10.2174/0929867327666200907141016.
106. D'Souza, S.; Prema, K. V.; Balaji, S. Machine learning models for drug–target interactions: current knowledge and future directions. *Drug Discov. Today* **2020**, *25*, 748–756, doi:10.1016/j.drudis.2020.03.003.
107. Bagherian, M.; Sabeti, E.; Wang, K.; Sartor, M.A.; Nikolovska-Coleska, Z.; Najarian, K. Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. *Brief. Bioinform.* **2020**, doi:10.1093/bib/bbz157.
108. Thafar, M.; Raies, A. Bin; Albaradei, S.; Essack, M.; Bajic, V.B. Comparison Study of Computational Prediction Tools for Drug-Target Binding Affinities. *Front. Chem.* **2019**, *7*, 1–19, doi:10.3389/fchem.2019.00782.
109. Matsuzaka, Y.; Uesawa, Y. Prediction model with high-performance constitutive androstane receptor (CAR) using DeepSnap-deep learning approach from the tox21 10K compound library. *Int. J. Mol. Sci.* **2019**, *20*, 4855, doi:10.3390/ijms20194855.
110. Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288, doi:10.1016/0022-2836(82)90153-X.
111. Wu, Z.; Ramsundar, B.; Feinberg, E.N.; Gomes, J.; Geniesse, C.; Pappu, A.S.; Leswing, K.; Pande, V. MoleculeNet: A benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9*, 513–530, doi:10.1039/c7sc02664a.
112. deepchem/deepchem: Democratizing Deep-Learning for Drug Discovery, Quantum Chemistry, Materials Science and Biology Available online: https://github.com/deepchem/deepchem (accessed on Jul 13, 2021).
113. Davis, M.I.; Hunt, J.P.; Herrgard, S.; Ciceri, P.; Wodicka, L.M.; Pallares, G.; Hocker, M.; Treiber, D.K.; Zarrinkar, P.P. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2011**, *29*, 1046–1051, doi:10.1038/nbt.1990.
114. J, T.; A, S.; S, S.; T, X.; P, H.; K, W.; T, A. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J. Chem. Inf. Model.* **2014**, *54*, 735–743, doi:10.1021/CI400709D.
115. Wenzel, J.; Matter, H.; Schmidt, F. Predictive Multitask Deep Neural Network Models for ADME-Tox Properties: Learning from Large Data Sets. *J. Chem. Inf. Model.* **2019**, *59*, 1253–1268, doi:10.1021/ACS.JCIM.8B00785.
116. Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; Vinyals, O.; Dahl, G.E. Neural Message Passing for Quantum Chemistry. *34th Int. Conf. Mach. Learn. ICML 2017* **2017**, *3*, 2053–2070.
117. Withnall, M.; Lindelöf, E.; Engkvist, O.; Chen, H. Building attention and edge message passing neural networks for bioactivity and physical-chemical property prediction. *J. Cheminform.* **2020**, *12*, 1, doi:10.1186/s13321-019-0407-y.

118. Torng, W.; Altman, R.B. Graph Convolutional Neural Networks for Predicting Drug-Target Interactions. *J. Chem. Inf. Model.* **2019**, doi:10.1021/acs.jcim.9b00628.
119. Gao, K.Y.; Fokoue, A.; Luo, H.; Iyengar, A.; Dey, S.; Zhang, P. *Interpretable Drug Target Prediction Using Deep Neural Representation*; 2017;
120. Zhao, Q.; Xiao, F.; Yang, M.; Li, Y.; Wang, J. AttentionDTA: Prediction of drug-target binding affinity using attention model. In Proceedings of the Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019; Institute of Electrical and Electronics Engineers Inc., 2019; pp. 64–69.
121. Cortés-Ciriano, I.; Ain, Q.U.; Subramanian, V.; Lenselink, E.B.; Méndez-Lucio, O.; IJzerman, A.P.; Wohlfahrt, G.; Prusis, P.; Malliavin, T.E.; Westen, G.J.P. van; et al. Polypharmacology modelling using proteochemometrics (PCM): recent methodological developments, applications to target families, and future prospects. *Medchemcomm* **2015**, *6*, 24–50, doi:10.1039/C4MD00216D.
122. Nascimento, A.C.A.; Prudêncio, R.B.C.; Costa, I.G. A multiple kernel learning algorithm for drug-target interaction prediction. *BMC Bioinforma. 2016 171* **2016**, *17*, 1–16, doi:10.1186/S12859-016-0890-3.
123. He, T.; Heidemeyer, M.; Ban, F.; Cherkasov, A.; Ester, M. SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines. *J. Cheminform.* **2017**, *9*, 1–14, doi:10.1186/s13321-017-0209-z.
124. Wallach, I.; Dzamba, M.; Heifets, A. AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. **2015**, 1–11.
125. Trott, O.; Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **2010**, *31*, 455, doi:10.1002/JCC.21334.
126. DR, K.; MP, B.; CJ, C. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904, doi:10.1021/CI300604Z.
127. Z, X.; LY, W.; X, Z.; ST, W. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst. Biol.* **2010**, *4 Suppl 2*, S6, doi:10.1186/1752-0509-4-S2-S6.
128. Liu, Y.; Wu, M.; Miao, C.; Zhao, P.; Li, X.-L. Neighborhood Regularized Logistic Matrix Factorization for Drug-Target Interaction Prediction. *PLOS Comput. Biol.* **2016**, *12*, e1004760, doi:10.1371/JOURNAL.PCBI.1004760.
129. Metz, J.T.; Johnson, E.F.; Soni, N.B.; Merta, P.J.; Kifle, L.; Hajduk, P.J. Navigating the kinome. *Nat. Chem. Biol.* **2011**, *7*, 200–202, doi:10.1038/nchembio.530.
130. Perozzi, B.; Al-Rfou, R.; Skiena, S. DeepWalk: Online Learning of Social Representations. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **2014**, 701–710, doi:10.1145/2623330.2623732.
131. Camacho, D.M.; Collins, K.M.; Powers, R.K.; Costello, J.C.; Collins, J.J. Next-Generation Machine Learning for Biological Networks. *Cell* **2018**, *173*, 1581–1592, doi:10.1016/j.cell.2018.05.015.
132. Luo, Y.; Zhao, X.; Zhou, J.; Yang, J.; Zhang, Y.; Kuang, W.; Peng, J.; Chen, L.; Zeng, J. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.* **2017**, *8*, doi:10.1038/s41467-017-00680-8.
133. David, L.; Arús-Pous, J.; Karlsson, J.; Engkvist, O.; Bjerrum, E.J.; Kogej, T.; Kriegl, J.M.; Beck, B.; Chen, H. Applications of Deep-Learning in Exploiting Large-Scale and Heterogeneous Compound Data in Industrial Pharmaceutical Research. *Front. Pharmacol.* **2019**, *10*, doi:10.3389/FPHAR.2019.01303.
134. K, B.; Y, Y. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics* **2009**, *25*, 2397–2403, doi:10.1093/BIOINFORMATICS/BTP433.
135. Keum, J.; Nam, H. SELF-BLM: Prediction of drug-target interactions via self-training SVM. *PLoS One* **2017**, *12*, e0171839, doi:10.1371/JOURNAL.PONE.0171839.
136. Hao, M.; Wang, Y.; Bryant, S.H. Improved prediction of drug-target interactions using regularized least squares integrating with kernel fusion technique. *Anal. Chim. Acta* **2016**, *909*, 41, doi:10.1016/J.ACA.2016.01.014.
137. Chen, X.; Liu, M.-X.; Yan, G.-Y. Drug–target interaction prediction by random walk on the heterogeneous network. *Mol. Biosyst.* **2012**, *8*, 1970–1978, doi:10.1039/C2MB00002D.
138. Bedi, P.; Sharma, C.; Vashisth, P.; Goel, D.; Dhanda, M. Handling cold start problem in Recommender Systems by using Interaction Based Social Proximity factor. *2015 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2015* **2015**, 1987–1993, doi:10.1109/ICACCI.2015.7275909.
139. Yu, H.; Choo, S.; Park, J.; Jung, J.; Kang, Y.; Lee, D. Prediction of drugs having opposite effects on disease genes in a directed network. *BMC Syst. Biol. 2015 101* **2016**, *10*, 17–25, doi:10.1186/S12918-015-0243-2.
140. Lee, H.; Kang, S.; Kim, W. Drug Repositioning for Cancer Therapy Based on Large-Scale Drug-Induced Transcriptional Signatures. *PLoS One* **2016**, *11*, e0150460, doi:10.1371/JOURNAL.PONE.0150460.
141. Duda, M.; Zhang, H.; Li, H.-D.; Wall, D.P.; Burmeister, M.; Guan, Y. Brain-specific functional relationship networks inform autism spectrum disorder gene prediction. *Transl. Psychiatry 2018 81* **2018**, *8*, 1–9, doi:10.1038/s41398-018-0098-6.
142. Liu, T.-P.; Hsieh, Y.-Y.; Chou, C.-J.; Yang, P.-M. Systematic polypharmacology and drug repurposing via an integrated L1000-based Connectivity Map database mining. *R. Soc. Open Sci.* **2018**, *5*, doi:10.1098/RSOS.181321.
143. Gao, Y.; Kim, S.; Lee, Y.-I.; Lee, J. Cellular Stress-Modulating Drugs Can Potentially Be Identified by in Silico Screening with Connectivity Map (CMap). *Int. J. Mol. Sci. 2019, Vol. 20, Page 5601* **2019**, *20*, 5601, doi:10.3390/IJMS20225601.
144. Hizukuri, Y.; Sawada, R.; Yamanishi, Y. Predicting target proteins for drug candidate compounds based on drug-induced gene expression data in a chemical structure-independent manner. *BMC Med. Genomics* **2015**, *8*, doi:10.1186/S12920-015-0158-1.

145. R, S.; M, I.; Y, T.; H, Y.; Y, Y. Predicting inhibitory and activatory drug targets by chemically and genetically perturbed transcriptome signatures. *Sci. Rep.* **2018**, *8*, doi:10.1038/S41598-017-18315-9.
146. Arús-Pous, J.; Patronov, A.; Bjerrum, E.J.; Tyrchan, C.; Reymond, J.L.; Chen, H.; Engkvist, O. SMILES-based deep generative scaffold decorator for de-novo drug design. *J. Cheminform.* **2020**, *12*, 1–18, doi:10.1186/s13321-020-00441-8.
147. Shao, K.; Zhang, Z.; He, S.; Bo, X. DTIGCCN: Prediction of drug-target interactions based on GCN and CNN. **2020**, 337–342, doi:10.1109/ictai50040.2020.00060.
148. Fefferman, C.; Mitter, S.; Narayanan, H. Testing the manifold hypothesis. *J. Am. Math. Soc.* **2016**, *29*, 983–1049, doi:10.1090/jams/852.
149. Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning:Generative models for matter engineering. *Science (80-.).* **2018**, *361*, 360–365, doi:10.1126/science.aat2663.
150. Li, X.; Xu, Y.; Yao, H.; Lin, K. Chemical space exploration based on recurrent neural networks: Applications in discovering kinase inhibitors. *J. Cheminform.* **2020**, *12*, 1–13, doi:10.1186/s13321-020-00446-3.
151. It, B.; Learning, M.; Autoen-, V.; Autoencoder, M.V.; Vae, E.; Vae, E.; Generative, N. LATENT OPTIMIZATION VARIATIONAL AUTOENCODER FOR CONDITIONAL MOLECULE GENERATION. **2020**, 1–44.
152. Aumentado-Armstrong, T. Latent molecular optimization for targeted therapeutic design. *arXiv* **2018**, 1–14.
153. Koge, D.; Ono, N.; Huang, M.; Altaf-Ul-Amin, M.; Kanaya, S. Embedding of Molecular Structure Using Molecular Hypergraph Variational Autoencoder with Metric Learning. *Mol. Inform.* **2021**, *40*, 1–7, doi:10.1002/minf.202000203.
154. Blaschke, T.; Olivecrona, M.; Engkvist, O.; Bajorath, J.; Chen, H. Application of Generative Autoencoder in De Novo Molecular Design. *Mol. Inform.* **2018**, *37*, 1–11, doi:10.1002/minf.201700123.
155. Kang, S.; Cho, K. Conditional Molecular Design with Deep Generative Models. *J. Chem. Inf. Model.* **2019**, *59*, 43–52, doi:10.1021/acs.jcim.8b00263.
156. Hong, S.H.; Ryu, S.; Lim, J.; Kim, W.Y. Molecular Generative Model Based on an Adversarially Regularized Autoencoder. *J. Chem. Inf. Model.* **2020**, *60*, 29–36, doi:10.1021/acs.jcim.9b00694.
157. Grisoni, F.; Moret, M.; Lingwood, R.; Schneider, G. Bidirectional Molecule Generation with Recurrent Neural Networks. *J. Chem. Inf. Model.* **2020**, *60*, 1175–1183, doi:10.1021/acs.jcim.9b00943.
158. Bongini, P.; Bianchini, M.; Scarselli, F. Molecular graph generation with Graph Neural Networks. **2020**, 1–21, doi:10.1016/j.neucom.2021.04.039.
159. Bian, Y.; Xie, X.Q. (Sean) Computational Fragment-Based Drug Design: Current Trends, Strategies, and Applications. *AAPS J.* **2018**, *20*, 1–11, doi:10.1208/s12248-018-0216-7.
160. F. Mojica, M.; A. Bonomo, R.; Fast, W. B1-Metallo-β-Lactamases: Where Do We Stand? *Curr. Drug Targets* **2015**, *17*, 1029–1050, doi:10.2174/1389450116666151001105622.
161. Jin, W.; Barzilay, R.; Jaakkola, T. Hierarchical Generation of Molecular Graphs using Structural Motifs. *arXiv* **2020**.
162. Kawai, K.; Nagata, N.; Takahashi, Y. De novo design of drug-like molecules by a fragment-based molecular evolutionary approach. *J. Chem. Inf. Model.* **2014**, *54*, 49–56, doi:10.1021/ci400418c.
163. Spiegel, J.O.; Durrant, J.D. AutoGrow4: An open-source genetic algorithm for de novo drug design and lead optimization. *J. Cheminform.* **2020**, *12*, 1–16, doi:10.1186/s13321-020-00429-4.
164. Leguy, J.; Cauchy, T.; Glavatskikh, M.; Duval, B.; Da Mota, B. EvoMol: a flexible and interpretable evolutionary algorithm for unbiased de novo molecular generation. *J. Cheminform.* **2020**, *12*, 1–19, doi:10.1186/s13321-020-00458-z.
165. Anonymous Deep Evolutionary Learning for Molecular Design. **2020**, *ICLR*, 1–12.
166. Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594, doi:10.1021/JM300687E.
167. JJ, I. Community benchmarks for virtual screening. *J. Comput. Aided. Mol. Des.* **2008**, *22*, 193–199, doi:10.1007/S10822-008-9189-4.
168. Rohrer, S.G.; Baumann, K. Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data. *J. Chem. Inf. Model.* **2009**, *49*, 169–184, doi:10.1021/ci8002649.
169. Chen, L.; Cruz, A.; Ramsey, S.; Dickson, C.J.; Duca, J.S.; Hornak, V.; Koes, D.R.; Kurtzman, T. Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening. *PLoS One* **2019**, *14*, e0220113, doi:10.1371/journal.pone.0220113.
170. Réau, M.; Langenfeld, F.; Zagury, J.F.; Lagarde, N.; Montes, M. Decoys selection in benchmarking datasets: Overview and perspectives. *Front. Pharmacol.* **2018**, *9*, doi:10.3389/fphar.2018.00011.
171. MR, B.; TM, I.; SM, V.; FM, B. Evaluation and optimization of virtual screening workflows with DEKOIS 2.0--a public library of challenging docking benchmark sets. *J. Chem. Inf. Model.* **2013**, *53*, 1447–1462, doi:10.1021/CI400115B.
172. Xia, J.; Jin, H.; Liu, Z.; Zhang, L.; Wang, X.S. An Unbiased Method To Build Benchmarking Sets for Ligand-Based Virtual Screening and its Application To GPCRs. **2014**, doi:10.1021/ci500062f.
173. Tran-Nguyen, V.K.; Jacquemard, C.; Rognan, D. LIT-PCBA: An unbiased data set for machine learning and virtual screening. *J. Chem. Inf. Model.* **2020**, *60*, 4263–4273, doi:10.1021/acs.jcim.0c00155.
174. Polykovskiy, D.; Zhebrak, A.; Sanchez-Lengeling, B.; Golovanov, S.; Tatanov, O.; Belyaev, S.; Kurbanov, R.; Artamonov, A.; Aladinskiy, V.; Veselov, M.; et al. Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models. *Front. Pharmacol.* **2018**, *11*.
175. Brown, N.; Fiscato, M.; Segler, M.H.S.; Vaucher, A.C. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* **2019**, *59*, 1096–1108, doi:10.1021/acs.jcim.8b00839.

176. Grant, L.L.; Sit, C.S. De novo molecular drug design benchmarking . *RSC Med. Chem.* **2021**, doi:10.1039/d1md00074h.
177. García, V.; Mollineda, R.A.; Sánchez, J.S. Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2009**, *5524 LNCS*, 441–448, doi:10.1007/978-3-642-02172-5_57.
178. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics 2019 211* **2020**, *21*, 1–13, doi:10.1186/S12864-019-6413-7.
179. Carrington, A.M.; Fieguth, P.W.; Qazi, H.; Holzinger, A.; Chen, H.H.; Mayr, F.; Manuel, D.G. A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. *BMC Med. Informatics Decis. Mak. 2020 201* **2020**, *20*, 1–12, doi:10.1186/S12911-019-1014-6.
180. Roy, P.P.; Roy, K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* **2008**, *27*, 302–313, doi:10.1002/qsar.200710043.
181. Playe, B.; Stoven, V. Evaluation of network architecture and data augmentation methods for deep learning in chemogenomics. **2019**, doi:10.1101/662098.
182. Kotsias, P.-C.; Arús-Pous, J.; Chen, H.; Engkvist, O.; Tyrchan, C.; Bjerrum, E.J. Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks. *Nat. Mach. Intell.* **2020**, *2*, 254–265, doi:10.1038/s42256-020-0174-5.
183. González-Medina, M.; Owen, J.R.; El-Elimat, T.; Pearce, C.J.; Oberlies, N.H.; Figueroa, M.; Medina-Franco, J.L. Scaffold diversity of fungal metabolites. *Front. Pharmacol.* **2017**, *8*, doi:10.3389/fphar.2017.00180.
184. Karimi, M.; Wu, D.; Wang, Z.; Shen, Y. Explainable Deep Relational Networks for Predicting Compound-Protein Affinities and Contacts. *J. Chem. Inf. Model.* **2021**, *61*, 46–66, doi:10.1021/acs.jcim.0c00866.
185. Bickerton, G.R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A.L. Quantifying the chemical beauty of drugs. *Nat. Chem. 2011 42* **2012**, *4*, 90–98, doi:10.1038/nchem.1243.
186. Cai, C.; Wang, S.; Xu, Y.; Zhang, W.; Tang, K.; Ouyang, Q.; Lai, L.; Pei, J. Transfer Learning for Drug Discovery. *J. Med. Chem.* **2020**, *63*, 8683–8694, doi:10.1021/acs.jmedchem.9b02147.
187. Y, Y.; M, A.; A, G.; W, H.; M, K. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* **2008**, *24*, doi:10.1093/BIOINFORMATICS/BTN162.
188. Tran-Nguyen, V.K.; Rognan, D. Benchmarking data sets from pubchem bioassay data: Current scenario and room for improvement. *Int. J. Mol. Sci.* 2020, *21*, 1–22.
189. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrián-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954, doi:10.1093/NAR/GKW1074.
190. Wang, Y.; Bryant, S.H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B.A.; Thiessen, P.A.; He, S.; Zhang, J. PubChem BioAssay: 2017 update. *Nucleic Acids Res.* **2017**, *45*, D955, doi:10.1093/NAR/GKW1118.
191. Sun, J.; Jeliazkova, N.; Chupakhin, V.; Golib-Dzib, J.-F.; Engkvist, O.; Carlsson, L.; Wegner, J.; Ceulemans, H.; Georgiev, I.; Jeliazkov, V.; et al. ExCAPE-DB: an integrated large scale dataset facilitating Big Data analysis in chemogenomics. *J. Cheminform.* **2017**, *9*, 17, doi:10.1186/s13321-017-0203-5.
192. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A survey on deep transfer learning. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2018**, *11141 LNCS*, 270–279, doi:10.1007/978-3-030-01424-7_27.
193. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, doi:10.1186/s40537-019-0197-0.
194. Cortes-Ciriano, I.; Bender, A. Improved Chemical Structure-Activity Modeling Through Data Augmentation. *J. Chem. Inf. Model.* **2015**, *55*, 2682–2692, doi:10.1021/acs.jcim.5b00570.
195. Arús-Pous, J.; Awale, M.; Probst, D.; Reymond, J.L. Exploring chemical space with machine learning. *Chimia (Aarau).* **2019**, *73*, 1018–1023, doi:10.2533/chimia.2019.1018.
196. Cho, Y.R.; Kang, M. Interpretable machine learning in bioinformatics. *Methods* **2020**, *179*, 1–2, doi:10.1016/j.ymeth.2020.05.024.
197. Doshi-Velez, F.; Kim, B. Towards A Rigorous Science of Interpretable Machine Learning 22. **2017**, 1–13.
198. Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2020**, *2*, 573–584, doi:10.1038/s42256-020-00236-4.
199. Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C.A.; Bekas, C.; Lee, A.A. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent. Sci.* **2019**, *5*, 1572–1583, doi:10.1021/acscentsci.9b00576.
200. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2016**, *2016-Decem*, 2921–2929, doi:10.1109/CVPR.2016.319.
201. Liu, H.; Lee, C.-W.; Su, B.-H.; Tseng, Y.J. A new explainable graph convolution network based on discrete method: using water solubility as an example Abstract. **2015**, 2232.
202. Gonczarek, A.; Tomczak, J.M.; Zaręba, S.; Kaczmar, J.; Dąbrowski, P.; Walczak, M.J. Learning Deep Architectures for Interaction Prediction in Structure-based Virtual Screening. **2016**, 2–6, doi:10.1016/j.compbiomed.2017.09.007.
203. Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D.R. Protein-Ligand Scoring with Convolutional Neural Networks. *J. Chem. Inf. Model.* **2017**, *57*, 942–957, doi:10.1021/acs.jcim.6b00740.
204. Jiménez, J.; Škalič, M.; Martínez-Rosell, G.; De Fabritiis, G. KDEEP: Protein-Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. *J. Chem. Inf. Model.* **2018**, *58*, 287–296, doi:10.1021/acs.jcim.7b00650.
205. Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf. Model.* **2018**, *59*, 895–913, doi:10.1021/ACS.JCIM.8B00545.

206. Govindaraj, R.G.; Brylinski, M. Comparative assessment of strategies to identify similar ligand-binding pockets in proteins. *BMC Bioinforma. 2018 191* **2018**, *19*, 1–17, doi:10.1186/S12859-018-2109-2.
207. Hasan Mahmud, S.M.; Chen, W.; Jahan, H.; Dai, B.; Din, S.U.; Dzisoo, A.M. DeepACTION: A deep learning-based method for predicting novel drug-target interactions. *Anal. Biochem.* **2020**, *610*, 113978, doi:10.1016/j.ab.2020.113978.
208. Chong, C.W.; Raveendran, P.; Mukundan, R. Translation and scale invariants of Legendre moments. *Pattern Recognit.* **2004**, *37*, 119–129, doi:10.1016/J.PATCOG.2003.06.003.
209. Tsubaki, M.; Tomii, K.; Sese, J. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics* **2019**, *35*, 309–318, doi:10.1093/BIOINFORMATICS/BTY535.
210. Jiang, M.; Li, Z.; Zhang, S.; Wang, S.; Wang, X.; Yuan, Q.; Wei, Z. Drug-target affinity prediction using graph neural network and contact maps. *RSC Adv.* **2020**, *10*, 20701–20712, doi:10.1039/d0ra02297g.
211. Goh, K.-I.; Cusick, M.E.; Valle, D.; Childs, B.; Vidal, M.; Barabási, A.-L. The human disease network. *Proc. Natl. Acad. Sci.* **2007**, *104*, 8685–8690, doi:10.1073/PNAS.0701361104.
212. Mongia, A.; Majumdar, A. Drug-Target Interaction prediction using Multi-Graph Regularized Deep Matrix Factorization., doi:10.1101/774539.
213. Zhong, F.; Wu, X.; Li, X.; Wang, D.; Fu, Z.; Liu, X.; Wan, X.; Yang, T.; Luo, X.; Chen, K.; et al. Computational target fishing by mining transcriptional data using a novel Siamese spectral-based graph convolutional network. **2020**, 1–29, doi:10.1101/2020.04.01.019166.
214. CF, T.; TE, K.; RB, A. PharmGKB: the Pharmacogenomics Knowledge Base. *Methods Mol. Biol.* **2013**, *1015*, 311–320, doi:10.1007/978-1-62703-435-7_20.
215. Backman, T.W.H.; Evans, D.S.; Girke, T. Large-scale bioactivity analysis of the small-molecule assayed proteome. *PLoS One* **2017**, *12*, e0171413, doi:10.1371/journal.pone.0171413.
216. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082, doi:10.1093/NAR/GKX1037.
217. Armstrong, J.F.; Faccenda, E.; Harding, S.D.; Pawson, A.J.; Southan, C.; Sharman, J.L.; Campo, B.; Cavanagh, D.R.; Alexander, S.P.H.; Davenport, A.P.; et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. *Nucleic Acids Res.* **2020**, *48*, D1006–D1021, doi:10.1093/NAR/GKZ951.
218. S, G.; M, K.; M, D.; M, C.; C, S.; E, P.; J, A.; EG, U.; A, G.; LJ, J.; et al. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.* **2008**, *36*, doi:10.1093/NAR/GKM862.
219. Wagner, A.H.; Coffman, A.C.; Ainscough, B.J.; Spies, N.C.; Skidmore, Z.L.; Campbell, K.M.; Krysiak, K.; Pan, D.; McMichael, J.F.; Eldred, J.M.; et al. DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res.* **2016**, *44*, D1036--44, doi:10.1093/nar/gkv1165.
220. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; Wiegers, J.; Wiegers, T.C.; Mattingly, C.J. Comparative Toxicogenomics Database (CTD): update 2021. *Nucleic Acids Res.* **2021**, *49*, D1138–D1143, doi:10.1093/NAR/GKAA891.
221. Chen, X.; Ji, Z.L.; Chen, Y.Z. TTD: Therapeutic Target Database. *Nucleic Acids Res.* **2002**, *30*, 412–415.
222. KiBA - a benchmark dataset for drug target prediction — Helsingin yliopisto Available online: https://researchportal.helsinki.fi/fi/datasets/kiba-a-benchmark-dataset-for-drug-target-prediction (accessed on Aug 12, 2021).
223. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045--53, doi:10.1093/nar/gkv1072.
224. Exploring ToxCast Data: Citing ToxCast Data | US EPA Available online: https://www.epa.gov/chemical-research/exploring-toxcast-data-citing-toxcast-data (accessed on Aug 12, 2021).