From machine learning to deep learning: progress in machine intelligence for rational drug discovery

Lu Zhang[1], Jianjun Tan[1], Dan Han[1] and Hao Zhu[1,2,3]

[1] College of Life Science and Bio-engineering, Beijing University of Technology, Beijing, 100124, China
[2] Department of Chemistry, Rutgers University, Camden, NJ 08102, USA
[3] The Rutgers Center for Computational and Integrative Biology, Camden, NJ 08102, USA

Machine intelligence, which is normally presented as artificial intelligence, refers to the intelligence exhibited by computers. In the history of rational drug discovery, various machine intelligence approaches have been applied to guide traditional experiments, which are expensive and time-consuming. Over the past several decades, machine-learning tools, such as quantitative structure–activity relationship (QSAR) modeling, were developed that can identify potential biological active molecules from millions of candidate compounds quickly and cheaply. However, when drug discovery moved into the era of 'big' data, machine learning approaches evolved into deep learning approaches, which are a more powerful and efficient way to deal with the massive amounts of data generated from modern drug discovery approaches. Here, we summarize the history of machine learning and provide insight into recently developed deep learning approaches and their applications in rational drug discovery. We suggest that this evolution of machine intelligence now provides a guide for early-stage drug design and discovery in the current big data era.

Introduction

Computational tools have been developed and applied to drug discovery as cost-effective alternatives to traditional experiment protocols. The accurate identification of new hits from large chemical libraries by computational models is desirable for the pharmaceutical industry because it can reduce the costs and time associated with experiments needed to obtain new drug candidates with optimized pharmacodynamics and pharmacokinetic (PK/PD) properties [1]. Virtual screening (VS), which is a standard computational approach, is widely used to guide rational drug discovery [2]. Historically, machine-learning approaches, which are one of the most important components of machine intelligence, have been used to generate various QSAR models for VS over the past few decades [3]. The resulting models are based on molecular structures and target activities, such as physicochemical properties, therapeutic activities, and PK properties [4], which can vary in the different stages of drug discovery.

The QSAR modeling procedure has been standardized across rational drug discovery processes [5]. Given the improvements in modeling approaches and the generation of descriptors, QSAR is widely applied at all stages of preclinical studies. The original hypothesis of QSAR ('similar compounds have similar activities') remains the foundation of all QSAR models developed so far. However, although different types of descriptor and different machine-learning methods used for QSAR modeling have their own pros and cons, the resulting models still suffer same issues, such as overfitting and active cliffs, which leads to the failure of predicting new compounds, especially those with chemical structures that different compared with those in the training sets used to develop QSAR models. Thus, new efforts are underway to make QSAR more applicable for drug discovery by integrating new modeling techniques. For example, currently the application of an applicability domain is a necessary step in QSAR modeling and

Corresponding authors: Tan, J. (tanjianjun@bjut.edu.cn), Zhu, H. (hao.zhu99@rutgers.edu)

http://dx.doi.org/10.1016/j.drudis.2017.08.010

the use of combinatorial QSAR avoids the potential problems caused by using an individual approach [6].

Over the past decade, rapid advances in high-throughput screening (HTS) techniques and relevant data-sharing protocols have moved modern drug discovery into the big data era. In addition, improvements in computational power resulting from the development of computer science hardware make big data modeling possible. The need for novel techniques, including data mining/generation, curation, storage, management, and modeling, results in both new challenges in, and opportunities for, the use of machine intelligence. Traditional QSAR approaches are not suitable for use with big data, which are characterized by volume (scale of data), velocity (growth of data), variety (diversity of sources), and veracity (uncertainty of data). Compared with machine learning, deep learning approaches that explain the vanishing effects of gradients [7] are more suitable to use with raw high-dimensional data. Therefore, as the result of data-driven and computational power-driven studies, machine intelligence has moved to a new position in drug discovery (Fig. 1). Here, we review current state-of-the-art deep-learning studies in the context of drug discovery and provide a brief summary of their advantages and future perspectives.

QSAR approaches

QSAR approaches used in drug discovery can be classified as linear and nonlinear techniques. Linear discriminant analysis (LDA), for example, introduced by Belhumeur in 1996 for pattern recognition and artificial intelligence [8], is a supervised machine-learning method that is suitable for dealing with small data sets. LDA is a classifier that considers a linear equation to maximize the between-class distance and minimize the within-class distance. LDA has been used to predict drug–drug interactions [9], identify new compounds [10], and detect adverse drug events [11], among others. Although LDA is a simple approach, the combination of LDA and novel descriptors is still considered a powerful modeling method. For example, Marrero et al. used a LDA algorithm combined with topologic, 3D-chiral, topographic, and geometric descriptors [10] to predict the antifungal activity of drugs and yielded a higher accuracy compared with other nonlinear approaches.

Support vector machines (SVMs) were proposed by Vapnik and colleagues [12] for their ability to deal with high-dimensional variables in small data sets. For linear problems, the SVM model separates different categories by mapping points in space to maximize the margin between different classes of point [13]. For nonlinear problems, SVMs use kernel mapping and transform nonlinear data sets into a high-dimension feature space for linear classification purposes. SVM has been widely applied for various modeling purposes in drug discovery [14]. For example, Poorinmohammad et al. [13] combined the SVM approach with pseudo amino acid composition descriptors to classify anti-HIV peptides, with a prediction accuracy of 96.76%.

Decision trees (DTs) are a transparent and interpretable machine-learning approach. Generally, there are two essential steps for the construction of decision trees: selecting attributes and pruning. First, molecule attributes are selected as a 'test' on a molecule (e.g., whether the partition coefficient of the molecule is >5). The selected attributes are viewed as internal nodes (including the root node and nonleaf nodes); the branch represents the outcome of the 'test' and the leaf node represents a classifica-

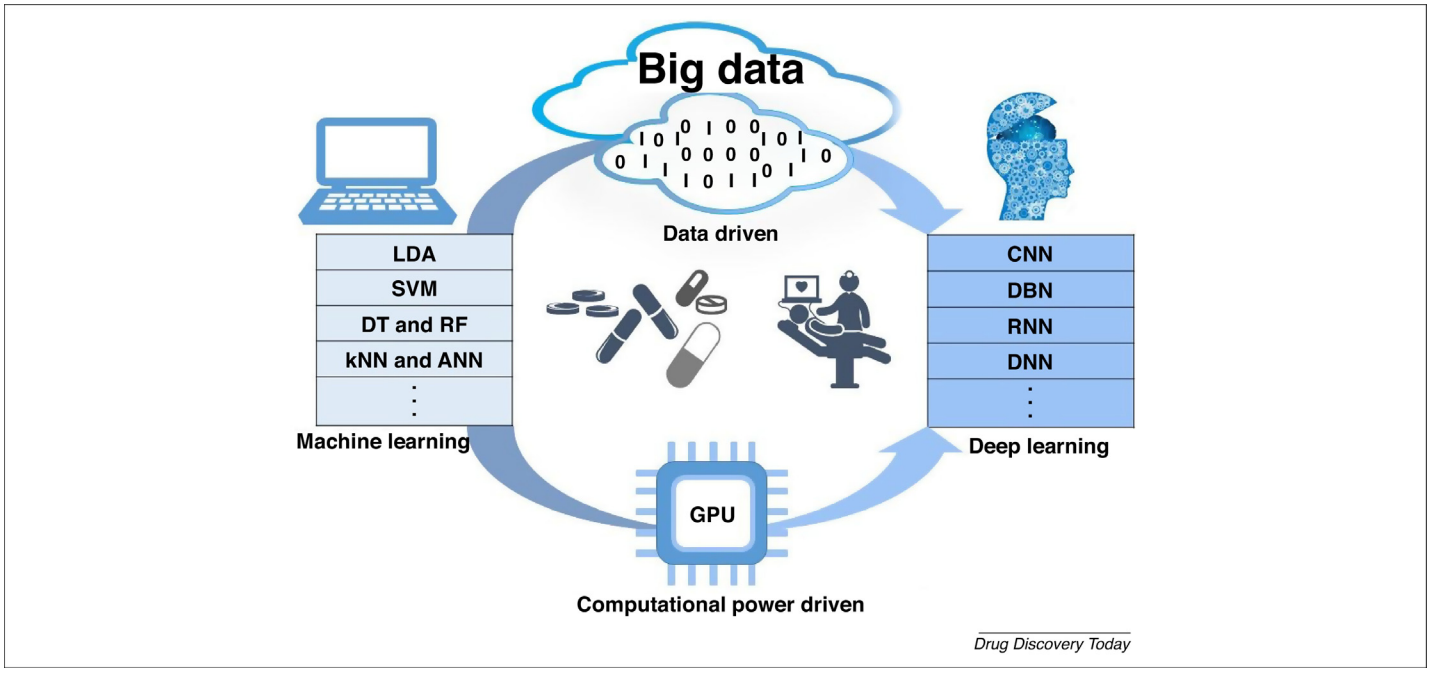

FIGURE 1

Advances in machine intelligence for drug discovery. Driven by massive data and powerful parallel computing capacity, traditional machine-learning methods have evolved into deep learning approaches. These methods have been shown to be useful for drug discovery and clinical medicine. Please see main text for definitions of abbreviations.

tion label. Second, to avoid overfitting and to decrease the complexity of the tree, pruning algorithms are used to trim the generated tree. Recently, DTs were used to model absorption, distribution, and metabolism properties [15] of drugs as well as their toxicity [16]. For example, to evaluate the toxicity of volatile organic compounds, Gupta and co-workers [16] used DT forest and DT boost algorithms to model the sensory irritation potency of volatile organic compounds. The former algorithm combined DTs with the bagging technique, whereas the latter integrated DTs with a gradient boosting algorithm; both models showed improvement over standard DTs.

Random forest (RF) is an ensemble modeling approach that operates by constructing multiple DTs as base learners. By introducing a random selection of features and the 'bagging' idea [17], each base learner further increases the 'test' nodes and is trained by random sampled subsets instead of by the original data set. The final outcome is a consensus score from all individual DT outputs. Compared with DTs, RF is less likely to overfit the data. RF has been widely used for bioactivity classification [18], toxicity modeling [19], protein–ligand binding affinity prediction [20], and drug target identification [21], among others. For example, Mistry et al. [19] used RF and DTs to model the drug–vehicle toxicity relationship for the first time. Their data set included 227 093 potential drug candidates and 39 potential vehicles. The resulted model predicted the toxicity relief of drugs by specific vehicles. Wang et al. [20] used thee RF method to model the protein–ligand binding affinity between 170 complexes of HIV-1 proteases, 110 complexes of trypsin, and 126 complexes of carbonic anhydrase. Furthermore, Kumari et al. improved RF by integrating bootstrap and rotation feature matrix components and successfully discriminated human drug targets from nondrug targets [21].

The k nearest neighbor (kNN) is an unsupervised algorithm for classification and regression [22]. In most cases, kNN is used for classifications that operate by counting the class of k nearest neighbors in the feature space. Thus, the kNN algorithm is one of the most simple and easy to perform of all machine-learning algorithms, and is normally integrated with other feature-selection algorithms. To identify antiviral drugs, Weidlich et al. [23] applied kNN integrated with a simulated annealing method and RF for 679 drug-like molecules. Their results showed that this improved kNN model outperformed the RF models. Meanwhile, the kNN approach was also applied to predict other bioactivities of drug-like molecules [24].

Artificial neural networks (ANN), which simulate brain function, are an attractive and powerful modeling approach widely used in recent drug discovery research. Based on their topological structure, ANN approaches can be classified into four types: forward, backward, random, and self-organized networks. Among these architectures, back propagation neural networks (BPNNs) are one of the most popular ANN methods. BPNN, proposed by Rumelhart and MeClelland, is a forward neural network with multilayered perception [25]. It is a gradient-descendent method that minimizes the mean-square errors of the difference between the network outputs and the experimental data in the training set. BPNN is characterized by robustness, superior fault tolerance, parallel co-processing, self-organizing, and self-learning abilities. BPNN has been applied not only in QSAR studies [26], but also in chemometric analysis [27].

Combinatorial QSAR and hybrid QSAR

Although QSAR approaches have been developed for decades, common issues remain that have not been solved by using any existing approaches. For example, most traditional QSAR studies have used a single modeling approach to develop a single model based on one type of descriptor. In other cases, when multiple QSAR models have been available, the model selection has always been performed based on statistics obtained from training sets (e.g., cross-validation). However, previous studies have shown that model performance based on training sets has a poor correlation with the external predictions of new compounds [28]. Thus, traditional QSARs that aim to develop a single model and/or select a single model based on training set performance for prediction purpose are questionable.

Compared with traditional QSAR modeling procedures (e.g., modeling by using one statistical tool and one type of descriptors), recent modeling studies in drug discovery have focused on predictions based on a combination of various types of model (by using different statistical tools and different types of chemical descriptor). Normally for a data set containing enormous and diverse compounds, an individual model would only cover part of its chemical and/or biological diversity. Consensus modeling based on a combinatorial QSAR (combi-QSAR) workflow take advantages of the output information obtained from various available individual models and fully explores the diverse chemical and/or biological information provided by a large training set. The combi-QSAR strategy has been applied to model various absorption, distribution, metabolism, and excretion (ADME) properties [29], the toxicity of drug molecules [30], and to select and design new drug candidates [31]. As a tradeoff, combi-QSAR modeling is more time-consuming than the development of a single model.

Another common issue related to QSAR modeling is the existence of 'active cliffs' [32]. QSAR modeling cannot deal with the situation whereby two compounds have similar structures but different activities, because it is against the basic hypothesis that 'similar compounds have similar activities'. In some early QSAR studies, additional physicochemical properties, such as partition coefficients (logP) [33], water solubility [34], and melting point [35], were used successfully to augment computed chemical descriptors and improve the predictive power of QSAR models. These studies suggest that experimental results obtained from low-cost experimental testing can be used as extra biological descriptors in QSAR modeling to help resolve the 'active cliffs' issue. Over the past decade, the rapidly expanding HTS data sets available for large and diverse chemical libraries make it possible to extend the scope of conventional chemical descriptors in QSAR modeling to new hybrid descriptors, including both chemical descriptors and biological descriptors. Therefore, in recent drug discovery studies, models were generated based on new hybrid descriptors. For example, Kim et al. and Wang et al. showed that the oral bioavailability and blood–brain barrier (BBB) models can be improved by including biological descriptors of membrane transportations [36,37]. In this new modeling strategy, the target properties of modeling are still biological activities in drug discovery, but the content and interpretation of 'descriptors' and the resulting models are different. This modeling focuses on the prediction of the same target property from different (chemical, biological, and

genomic) characteristics of drugs and provides a unique opportunity to take advantage of both chemical and biological information relating to drug molecules.

Advances in rational drug discovery resulting from deep learning

In the current era of big data and combined with the development of advanced screening protocols (e.g., HTS) and large chemical libraries, the amount of biological data is increasing dramatically. The availability of large data sets and their processing using graphics processing units (GPU) have promoted the development of new modeling approaches. In 2006, Hinton et al. introduced the deep belief networks that made it possible to construct nets with many hidden layers [7]. This resulted in a new theory and caught the attention of many researchers and leading pharmaceutical companies. The concept of deep learning originated from the ANN approach, in which feedforward neural networks combined with many hidden layers are thought of as deep neural networks [38]. Deep learning comprises simple but nonlinear processing units that each transform the representations or features at one level (starting with the raw input) into a representation at a higher, more representative level [39]. Thus, the deep learning approach is a representation-learning method that results in learning multiple levels of representations from low- to high-level features. For example, to recognize images, deep learning networks can learn color information from raw pixel inputs in the first layer and then transform color information to edges of objects in the next layer. Without manually selecting the molecular descriptors, deep learning methods automatically select representations from raw, high-dimension, and heterogeneous data, which is exactly what big data modeling requires [40]. Thus, this is likely to result in deep learning being widely used in various aspects of research, such as image recognition, speech recognition, video games, as well as model development in drug discovery [41]. The most commonly used networks are convolutional neural networks (CNN), stacked autoencoders, deep belief networks (DBN), and restricted Boltzmann machines. As a relatively new approach, its applications in drug discovery can be summarized as follows: (i) new drug molecule identification; (ii) protein engineering; (iii) gene expression data analysis; and (iv) pharmacodynamics modeling.

New drug molecule identification

Identifying new drug candidates from large chemical libraries with computational models (e.g., VS) is an effective and feasible way to facilitate the drug discovery process. Generally, deep learning methods can also be used in this approach [42,43] to perform VS [44–46]. For example, Pereira et al. introduced a novel deep learning-based VS method, called DeepVS [44]. They performed docking with 40 receptors and 2950 ligands, and compared the results with 95 316 decoys. The docking outputs were used to train deep CNN that could rank the list of ligands for each receptor. The results showed that DeepVS achieved the best performance reported for the VS of these 40 receptors. Similarly, deep learning can also be used to generate focused molecule libraries [47] or new molecular fingerprints [48] and to model PK properties of potential drugs [49].

Protein engineering

Protein engineering involves developing and simulating proteins using computers. Recently, researchers used deep learning approaches to explore and discover protein structures and functions. To uncover protein functions, many efforts have been made to simulate interactions between proteins and other biological molecules (e.g., DNA). For example, Hassanzadeh et al. used a recurrent convolutional network to predict the binding specificity of proteins to different DNA loci. They utilized data from *in vitro* high-throughput experiments to evaluate their modeling. This modeling approach was shown to be the most accurate for detecting the binding preference between two proteins and individual DNA subregions [50]. Deep learning methods can also be used to predict biological functions of proteins directly from their raw 3D electron density and electrostatic potential fields [51].

Gene expression data analysis

With the emergence of next-generation sequencing technology, massive amounts of heterogeneous genomics data can fit well with the requirements of deep learning methods. Thus, deep learning methods have been used in precision medicine development [52], sequence specification prediction [53], and genomics modeling for drug repurposing [54]. For example, Aliper et al. [54] used transcriptional response data to predict the therapeutic categories of drugs. In their study, they used gene-level data of 26 420 drug perturbation samples belonging to 12 therapeutic categories across three cell lines. They integrated the gene expression profiles and pathway activation scores as new features into a deep neural network (DNN) modeling approach, which generated the highest classification accuracy compared with other traditional approaches. They also showed that DNN can accurately predict the category of drugs with different PK and PD conditions.

TABLE 1

Examples of commercial drugs and drug candidates discovered by computational methods.

Year	Drugs	Function	Computational method	Developer	Refs
2012	CCT244747	Inhibits Checkpoint kinase 1	Docking	The Institute of Cancer Research, UK; in preclinical phase	[57]
2014	PTC725	Inhibits hepatitis C RNA replication	SAR/QSAR	PTC Therapeutics, Merck Research Laboratories, USA; in preclinical phase	[58]
2016	RG7800	Treats spinal muscular atrophy	SAR/QSAR	Pharma Research & Early Development, PTC Therapeutics and SMA Foundation, USA; in Phase I clinical trials	[59]
2015	GDC-0941	Inhibits phosphatidylinositol-3-kinase	Molecular modeling	The Royal Marsden National Health Service Foundation Trust and The Institute of Cancer Research, UK; in Phase I clinical trials	[60]

Pharmacodynamics modeling

PD modeling is vital to determine the interactions between drugs and their associated targets. Given the diversity of drug molecules and their targets, the potential drug–protein interactions are also complex and have many potential conformations. Recently, deep learning methods were used to predict the interactions of different complexes, such as drug–protein [55] and homogenous complexes [56]. In a recent report, Wen et al. [55] used DBN to predict drug–target interactions. To identify new drug–target interaction pairs, they used 2 146 240 drug–protein interaction pairs that contained approved drugs and targets without separating them into different classes. The resulted recall (predicted positive pairs/number of samples) obtained by DBN and RF were 13.6% and 1.1%, respectively, which highlighted the improved potential of deep learning methods in finding new drug–target interactions compared with QSAR approaches.

Concluding remarks and future perspectives

Machine intelligence has been applied in the drug discovery field for decades. Traditional machine-learning modeling has evolved into a variety of new methods, such as combi-QSAR and hybrid QSAR, and remains a popular approach to study various drug-related topics. There are various drugs on the market and/or in clinical trials that have been designed by machine-learning or other computational methods (Table 1) [57–60].

Despite the advantages and popularity of using machine-learning approaches (e.g., QSAR) in modeling studies, machine intelligence has, in some instances, been replaced by deep learning in recent years. The development of deep learning methods is driven by the accumulation of massive amounts of biomedical data and the powerful parallel computing capacity of GPUs. Importantly, deep learning methods can deal with complex tasks based on large, heterogeneous, and high-dimensional data sets without the need for human input. These methods have been shown to be useful in many practical and commercial applications, including drug discovery studies.

Although deep learning is a promising new technique in machine intelligence, deep learning methods and their related studies still have some limitations. First, the availability of a large amount of high-quality data will affect the performance and reliability of deep learning modeling. The massive amounts of biomedical data generated by pharmaceutical companies are normally not available to the public but are kept as expensive private commercial assets. Second, a lack of rational interpretations of associated biological mechanisms is another limitation of models generated by deep learning methods. Although they have been shown to have high prediction accuracies, deep learning models still perform as 'black boxes' that are difficult to use to reveal the biological mechanisms integrated in the data used for modeling.

Overall, as a newly developed machine intelligence technique, deep learning has demonstrated the potential for use in the new big data era of drug discovery. With more data becoming available and new approaches being developed, deep learning methods will become a major computer-aided drug design (CADD) approach in the near future.

Acknowledgment

The authors thank the Chinese Natural Science Foundation Project (No. 21173014) for financial support.

References

1 Danishuddin, M. and Khan, A.U. (2015) Structure based virtual screening to discover putative drug candidates: necessary considerations and successful case studies. *Methods* 71, 135–145
2 Scior, T. *et al.* (2012) Recognizing pitfalls in virtual screening: a critical review. *J. Chem. Inf. Model.* 52, 867–881
3 Wang, T. *et al.* (2015) Quantitative structure-activity relationship: promising advances in drug discovery platforms. *Expert Opin. Drug Dis.* 10, 1283–1300
4 Lavecchia, A. (2015) Machine-learning approaches in drug discovery: methods and applications. *Drug Discov. Today* 20, 318–331
5 Patani, G.A. and LaVoie, E.J. (1996) Bioisosterism: a rational approach in drug design. *Chem. Rev.* 96, 3147–3176
6 Zhu, H. *et al.* (2009) Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure. *Chem. Res. Toxicol.* 22, 1913–1921
7 Hinton, G.E. *et al.* (2006) pp. 1527–1554, MIT Press
8 Belhumeur, P.N. *et al.* (1996) Eigenfaces vs: Fisherfaces: recognition using class specific linear projection. *Eur. Conf. Comput. Vis.* 19, 45–58
9 Vilar, S. *et al.* (2015) Improving detection of arrhythmia drug-drug interactions in pharmacovigilance data through the implementation of similarity-based modeling. *PLoS One* 10, e0129974
10 Medina Marrero, R. *et al.* (2015) QuBiLs-MAS method in early drug discovery and rational drug identification of antifungal agents. *SAR QSAR Environ. Res.* 26, 943–958
11 Vilar, S. *et al.* (2015) 3D pharmacophoric similarity improves multi adverse drug event identification in pharmacovigilance. *Sci. Rep.* 5 8809
12 Vapnik, V.N. and Vapnik, V. (1998) *Statistical Learning Theory*. Wiley
13 Poorinmohammad, N. *et al.* (2015) Computational prediction of anti HIV-1 peptides and *in vitro* evaluation of anti HIV-1 activity of HIV-1 P24-derived peptides. *J. Pept. Sci.* 21, 10–16
14 Jain, N. *et al.* (2015) In silico de novo design of novel NNRTIs: a bio-molecular modelling approach. *RSC Adv.* 5, 14814–14827
15 Newby, D. *et al.* (2015) Decision trees to characterise the roles of permeability and solubility on the prediction of oral absorption. *Eur. J. Med. Chem.* 90, 751–765
16 Gupta, S. *et al.* (2015) Estimating sensory irritation potency of volatile organic chemicals using QSARs based on decision tree methods for regulatory purpose. *Ecotoxicology* 24, 873–886
17 Breiman, L. (2001) Random Forests. *Mach. Learn.* 45, 5–32
18 Singh, H. *et al.* (2015) QSAR based model for discriminating EGFR inhibitors and non-inhibitors using Random forest. *Biol Direct* 10, 1745–6150
19 Mistry, P. *et al.* (2016) Using random forest and decision tree models for a new vehicle prediction approach in computational toxicology. *Soft. Comput.* 20, 2967–2979
20 Wang, Y. *et al.* (2015) A comparative study of family-specific protein-ligand complex affinity prediction based on random forest approach. *J. Comput. Aid. Mol. Des.* 29, 349–360
21 Kumari, P. *et al.* (2015) Identification of human drug targets using machine-learning algorithms. *Comput. Biol. Med.* 56, 175–181
22 Altman, N.S. (1992) An introduction to kernel and nearest–neighbor nonparametric regression. *Am. Stat.* 46, 175–185
23 Weidlich, I.E. *et al.* (2013) Inhibitors for the hepatitis C virus RNA polymerase explored by SAR with advanced machine learning methods. *Bioorg. Med. Chem.* 21, 3127–3137
24 Jaradat, N.J. *et al.* (2015) Combining docking-based comparative intermolecular contacts analysis and k-nearest neighbor correlation for the discovery of new check point kinase 1 inhibitors. *J. Comput. Aid. Mol. Des.* 29, 561–581
25 Rumelhart, D.E. *et al.* (1986) Learning representations by back-propagating errors. *Nature* 323, 533–536
26 Zhang, T. *et al.* (2015) Activity prediction and molecular mechanism of bovine blood derived angiotensin I-converting enzyme inhibitory peptides. *PLoS One* 10, e0119598

27 Jain, N. *et al.* (2015) In silico de novo design of novel NNRTIs: a bio-molecular modelling approach. *RSC Adv.* 5, 14814–14827
28 Johnson, S.R. (2008) The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *J. Chem. Inf. Model.* 48, 25–26
29 Zhang, L.Y. *et al.* (2008) QSAR modeling of the blood–brain barrier permeability for diverse organic compounds. *Pharm. Res.* 25, 1573–1904
30 Solimeo, R. *et al.* (2012) Predicting chemical ocular toxicity using a combinatorial QSAR approach. *Chem. Res. Toxicol.* 25, 2763–2769
31 Sprague, B. *et al.* (2014) Design, synthesis and experimental validation of novel potential chemopreventive agents using random forest and support vector machine binary classifiers. *J. Comput. Aid. Mol. Des.* 28, 631–646
32 Maggiora, G.M. (2006) On outliers and activity cliffs—why QSAR often disappoints. *J. Chem. Inf. Model.* 46, 1535
33 Klopman, G. *et al.* (2003) MCASE study of the multidrug resistance reversal activity of propafenone analogs. *J. Comput. Aid. Mol. Des.* 17, 291–297
34 Stoner, C.L. *et al.* (2004) Implementation of an ADME enabling selection and visualization tool for drug discovery. *J. Pharm. Sci.* 93, 1131–1141
35 Mayer, P. and Reichenberg, F. (2006) Can highly hydrophobic organic substances cause aquatic baseline toxicity and can they contribute to mixture toxicity? *Environ. Toxicol. Chem.* 25, 2639
36 Kim, M. *et al.* (2014) Critical evaluation of human oral bioavailability for pharmaceutical drugs by using various cheminformatics approaches. *Pharm. Res.* 31, 1002–1014
37 Wang, W. *et al.* (2015) Developing enhanced blood-brain barrier permeability models: Integrating external bio-assay data in QSAR modeling. *Pharm. Res.* 32, 3055–3065
38 Mamoshina, P. *et al.* (2016) Applications of deep learning in biomedicine. *Mol. Pharm.* 13 1445-1445
39 Lecun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444
40 Chen, X.W. and Lin, X. (2014) Big data deep learning: challenges and perspectives. *IEEE Access.* 2, 514–525
41 Schmidhuber, J. (2015) Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117
42 Zhang, R. *et al.* (2016) Using deep learning for compound selectivity prediction. *Curr. Comput. Aid. Drug* 12, 5
43 Ma, J. *et al.* (2015) Deep neural nets as a method for quantitative structure–activity relationships. *J. Chem. Inf. Model.* 55, 263–274
44 Pereira, J.C. *et al.* (2016) Boosting docking-based virtual screening with deep learning. *J. Chem. Inf. Model.* 56, 2495–2506
45 Unterthiner, T. *et al.* (2014) Deep learning as an opportunity in virtual screening. *Proc. Deep Learning Workshop NIPS*
46 Gonczarek, A. *et al.* (2016) Learning deep architectures for interaction prediction in structure-based virtual screening. *arXiv* 1610, 07187
47 Segler, M.H.S. *et al.* (2017) Generating focussed molecule libraries for drug discovery with recurrent neural networks. *arXiv* 1701, 01329
48 Kadurin, A. *et al.* (2016) The cornucopia of meaningful leads: applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* 8, 10883–10890
49 Hughes, T.B. *et al.* (2015) Modeling epoxidation of drug-like molecules with a deep machine learning network. *ACS Central Sci.* 1, 168–180
50 Wang, M.D. and Hassanzadeh, H.R. (2017) DeeperBind: enhancing prediction of sequence specificities of DNA binding proteins. *arXiv* 1611, 05777
51 Golkov, V. *et al.* (2017) 3D deep learning for biological function prediction from physical fields. *arXiv* 1704.04039
52 Liang, M. *et al.* (2015) Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach: IEEE/ACM Trans. *Comp. Biol. Bioinformatics* 12, 928–937
53 Alipanahi, B. *et al.* (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* 33, 831–838
54 Aliper, A. *et al.* (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharmaceut.* 13, 2524–2530
55 Wen, M. *et al.* (2017) Deep-learning-based drug-target interaction prediction. *J. Proteome. Res.* 16, 1401–1409
56 Kwon, S. and Yoon, S. (2017) DeepCCI: end-to-end deep learning for chemical-chemical interaction prediction. *arXiv* 1704.08432
57 Walton, M.I. *et al.* (2012) CCT244747 is a novel potent and selective chk1 inhibitor with oral efficacy alone and in combination with genotoxic anticancer drugs. *Clin. Cancer Res.* 18, 5650–5661
58 Zhang, N. *et al.* (2014) Structure–activity relationship (SAR) optimization of 6-(indol-2-yl)pyridine-3-sulfonamides: identification of potent, selective, and orally bioavailable small molecules targeting hepatitis c (HCV) ns4b. *J. Med. Chem.* 57, 2121–2135
59 Ratni, H. *et al.* (2016) Specific correction of alternative survival motor neuron 2 splicing by small molecules: discovery of a potential novel medicine to treat spinal muscular atrophy. *J. Med. Chem.* 59, 6086–6100
60 Sarker, D. *et al.* (2015) First-in-human phase i study of pictilisib (gdc-0941), a potent pan-class i phosphatidylinositol-3-kinase (pi3k) inhibitor, in patients with advanced solid tumors. *Clin. Cancer Res.* 21, 77–86