

## New Methods for Predicting Drug Molecule Activity Using Deep Learning

Zhang Jie ✉

Institute of Life Science, Jiyang College of Zhejiang A&F University, Zhuji, 311800, China

✉ Corresponding author email: [jessi.j.zhang@foxmail.com](mailto:jessi.j.zhang@foxmail.com)

Bioscience Method, 2024, Vol.15, No.1 doi: [10.5376/bm.2024.15.0004](https://doi.org/10.5376/bm.2024.15.0004)

Received: 12 Jan., 2024

Accepted: 13 Feb., 2024

Published: 25 Feb., 2024

**Copyright** © 2024 Zhang, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Preferred citation for this article:**

Zhang J., 2024, New methods for predicting drug molecule activity using deep learning, Bioscience Method, 15(1): 28-36 (doi: [10.5376/bm.2024.15.0004](https://doi.org/10.5376/bm.2024.15.0004))

**Abstract** With the rapid development of deep learning technology, its application in predicting drug molecule activity is becoming increasingly widespread. This study reviews the latest progress and applications of deep learning in the field of drug discovery, especially in predicting drug molecule activity. It focuses on discussing several major deep learning models, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Graph Neural Networks (GNN), and how they help improve the accuracy and efficiency of drug activity prediction. Additionally, the importance of interdisciplinary collaboration in promoting the application of deep learning in drug discovery is explored, and directions for future research are proposed, including improving model interpretability, optimizing data quality, and expanding the application of deep learning technology. This study aims to provide researchers and drug development experts with a comprehensive and in-depth perspective on the potential and challenges of deep learning in predicting drug molecule activity, while also offering insights and references for research and development in related fields.

**Keywords** Deep learning; Drug molecule activity; Drug discovery; Graph neural networks; Interdisciplinary collaboration

Over the past few decades, drug discovery has been a central focus of medical research, not only because it provides new therapies to improve patient quality of life, but also because it plays a critical role in public health and global health. However, despite the increasing importance of drug discovery, the traditional drug discovery process faces numerous challenges, including high R&D costs, lengthy development cycles, and low success rates (Noe and Peakman, 2017). Every new drug from conceptualization to market must undergo a lengthy R&D journey, covering stages from early basic research and molecular screening to clinical trials (Schneider, 2017). Additionally, traditional methods depend on limited chemical and biological knowledge, making it particularly difficult to predict the biological activity and safety of molecules, leading to many potential drug candidates failing in clinical trial stages.

With the rapid development of computing technology, deep learning, as an advanced form of artificial intelligence, has shown great potential in learning and simulating human cognitive processes (Walters and Barzilay, 2020). Deep learning technologies, particularly Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Graph Neural Networks (GNNs), have made revolutionary advances in fields such as image and speech recognition, natural language processing, and gaming strategies (Jiménez-Luna, 2020). These technologies can process and analyze large amounts of unstructured data, uncover deep patterns in complex data, thereby providing profound insights into human intelligence. Therefore, deep learning has not only changed the landscape of data science but also provided new perspectives and methods for medical research, especially drug discovery (Tran et al., 2023).

Given the tremendous success of deep learning in other fields, this study aims to explore how deep learning technologies can be applied to predict drug molecule activity, and how these new methods can help overcome the challenges of the traditional drug discovery process. This research will thoroughly review the latest advances and applications of deep learning technologies in drug molecule design, activity prediction, toxicity assessment, and drug-target interaction prediction. By deeply analyzing how deep learning can improve the efficiency and accuracy of drug discovery, this study will demonstrate the potential of these technologies in accelerating new drug development, reducing R&D costs, and enhancing drug safety.

Furthermore, this study will also discuss the challenges and future directions of applying deep learning in the field of drug discovery, aiming to provide valuable insights and guidance for drug researchers and developers. Through this research, it is expected to drive innovation in the field of drug discovery, bring more effective and safer new therapies to patients, provide a solid foundation for the further application and development of deep learning in drug discovery, and inspire more researchers and developers to explore this promising field.

## 1 Overview of Deep Learning Methods

### 1.1 Convolutional neural networks (CNN)

Convolutional Neural Networks (CNNs) are a fundamental and widely used network architecture within deep learning technologies, particularly suited for image processing and recognition tasks. In the field of drug discovery, CNNs are employed to process molecular images and structural data to identify and predict the biological activity of compounds. Molecular structures can be represented visually, where different atoms and chemical bonds are depicted using various colors and shapes. CNNs can extract important features from these images through their convolutional layers, learning about the molecule's intrinsic properties and activity correlations. This approach has shown superior performance in predicting drug molecule solubility, toxicity, and affinity towards specific proteins.

Jones et al. (2020) introduced a hybrid model that combines features from different representations, such as three-dimensional CNNs (3D-CNNs) and spatial graph CNNs (SG-CNNs), to enhance the accuracy of binding affinity predictions. They compared the performance of these models with traditional methods, demonstrating that their hybrid model surpasses both individual neural network models and conventional scoring methods in terms of accuracy and computational efficiency. Improving the prediction precision of protein-ligand binding affinity is crucial in drug discovery.

Hentabli et al. (2022) focused on developing a deep learning approach to predict the biological activity of compounds, introducing a novel technique using a convolutional neural network (CNN) model. The model was evaluated using standard datasets with homogenous and heterogenous activity categories (MDL Drug Data Report and Sutherland). By leveraging deep learning techniques, they advanced the computational prediction of compound biological activity, which is vital for the drug discovery and development process.

Yaseen (2023) proposed an innovative method using artificial intelligence, particularly convolutional neural networks (CNNs), to predict drug-target interactions (DTIs). This study developed a system based on machine learning and deep learning to classify drug-target interactions of different drug combinations. The results indicated that the new method significantly enhances the prediction accuracy of DTIs, which could accelerate the drug discovery and development processes.

These examples underscore the transformative impact of CNNs in the realm of drug discovery, highlighting their capability to significantly refine the prediction of molecular activities and interactions.

### 1.2 Recurrent neural networks (RNN)

Recurrent Neural Networks (RNNs) are another type of deep learning model, designed to process sequential data such as text or time series. In drug discovery, RNNs are utilized to handle molecular sequences, specifically the one-dimensional Simplified Molecular Input Line Entry System (SMILES) representations of compounds. By leveraging RNNs to process these sequences, researchers can capture long-term dependencies and patterns in molecular structures, thereby predicting the biological activities of molecules. RNNs are particularly adept at managing dynamically lengthed molecular sequences, learning complex chemical information from intricate molecular structures.

Zhang et al. (2019) introduced a deep learning-based method, DLBSS, which predicts transcription factor binding sites (TFBSs) by integrating DNA sequences with DNA shape features. This method employs a shared Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) to discover common patterns from DNA sequences and their corresponding shape characteristics.

Amabilino et al. (2020) discovered that Recurrent Neural Networks (RNNs) could serve as SMILES string

generators for automated drug design. By using transfer learning, an initial model is trained on a large generic molecular dataset to learn the general syntax of SMILES, and then fine-tuned on a smaller set of molecules, which enhances the effectiveness of molecule generation. This approach reduces the need for extensive post-screening and minimizes selection bias in drug candidate identification.

Kayama et al. (2021) utilized Recurrent Neural Networks (RNNs) to predict the success rates of PCR amplification for specific primer sets and DNA templates. This indicates that RNNs can learn the relationships between primers and template sequences and use this knowledge to predict outcomes of chemical reactions, suggesting a potential for RNNs to predict synthetic pathways for drug molecules. Although this study is not directly related to the synthesis of drug molecules, it demonstrates the potential of RNNs in predicting chemical reaction outcomes, which could accelerate the drug development process.

### 1.3 Graph neural networks (GNN)

Graph Neural Networks (GNN) have gained significant attention in the field of drug discovery in recent years. Unlike CNNs and RNNs, GNNs are specifically designed to process graph data, making them particularly suitable for molecular structure analysis, as molecules can be naturally represented as graphs—with atoms as nodes and chemical bonds as edges. GNNs capture complex interactions between nodes by updating the state of nodes, allowing the model to learn the overall structural information of molecules and interactions between atoms. This approach has shown high accuracy in predicting molecular activity, especially when considering the three-dimensional structure of molecules.

Wieder et al. (2021) introduced a new GNN architecture called Directed Edge Graph Isomorphism Network (D-GIN), which is composed of two different sub-architectures and can improve the accuracy of predicting the lipophilicity and solubility of molecules. They argue that combining models of different key aspects can make graph neural networks more insightful while enhancing their predictive ability.

Xiong et al. (2021) discussed the integration of artificial intelligence technologies, especially Graph Neural Networks (GNNs), in the field of new drug design. They introduced the applications of GNNs in new drug design from three main perspectives: molecular scoring, molecule generation and optimization, and synthesis planning. The goal of new drug design is to create new chemical entities with desired biological activity and pharmacokinetic properties. Furthermore, the study pointed out that data-driven methods have rapidly gained popularity in drug design in recent years, with GNNs receiving wide attention due to their effective processing of graph-structured data.

Low et al. (2022) proposed a GNN for predicting the Gibbs free energy of molecular dissolution ( $\Delta G_{\text{sol}}^{\text{v}}$ ), which, in addition to encoding typical atom and bond-level features, also incorporated chemically intuitive solvent-related parameters, such as semi-empirical local atomic charges and solvent dielectric constants. This work allows for the examination of interactions that enhance or reduce solubility through visualization of the learned model.

### 1.4 Self-supervised learning

Self-supervised learning is a machine learning technique that learns data representations without the need for externally annotated data. In drug discovery, self-supervised learning is employed to learn effective molecular representations from unlabeled molecular data. Through self-supervised learning frameworks, models are capable of learning general representations of molecules by predicting certain internal features of the molecules, such as parts of the molecule or its chemical properties. This method allows researchers to utilize a vast amount of unannotated compound data, thereby eliminating the dependence on expensive or hard-to-obtain labeled datasets. This way, self-supervised learning helps improve the generalization capability of models, enabling them to better predict the activity of new molecules.

Chen et al. (2021) developed a self-supervised learning method that pre-trained models from over seven hundred

million unlabeled molecules. This intrinsic learning of chemical logics enables the extraction of predictive representations from specific molecular sequences. To validate the proposed method, ten benchmark and thirty-eight virtual screening datasets were considered. Extensive validation showed that the method performed exceptionally well, confirming the capability of self-supervised learning to extract useful information from large-scale unlabeled datasets.

Self-supervised frameworks can efficiently utilize a large volume of non-annotated data to compensate for the lack of labeled data, especially in scenarios where data is scarce. A specific self-supervised framework designed for predicting molecular properties, improves the performance of graph neural networks in molecular property prediction by employing multiple pretext tasks across different scales of molecules (atoms, fragments, and whole molecules). These deep learning methods each have their unique advantages and play roles in different aspects of predicting drug molecule activity. By integrating these techniques, researchers can understand and predict the biological activity of molecules from various perspectives, providing powerful tools for the discovery and development of new drugs.

## 2 Application Cases for Predicting Drug Molecule Activity

### 2.1 Prediction of molecular properties

Deep learning technologies, especially Convolutional Neural Networks (CNNs) and Graph Neural Networks (GNNs), have been utilized to predict the physicochemical properties of drug molecules, such as solubility, lipophilicity, and molecular weight. These properties are crucial for assessing the pharmacological, toxicological, and pharmacokinetic characteristics of drug molecules. For instance, the ImageMol framework uses self-supervised learning to enhance the accuracy of predicting these molecular properties, demonstrating the potential of deep learning in this field.

Studies such as Tang et al. (2020) have established a Self-Attention Message Passing Neural Network (SAMPN) based on the graph neural network framework. This framework, which directly utilizes chemical graphs, effectively improves the prediction accuracy for molecular properties such as lipophilicity and solubility. Additionally, its attention mechanism allows for an intuitive display of each atom's contribution to the molecular properties, aiding researchers in visually understanding the relationship between molecular properties and structure (Tang et al., 2020).

Zeng et al. (2022) introduced a self-supervised pre-training deep learning framework called ImageMol, which extracts chemical representations from unlabeled drug samples to predict the molecular targets of candidate compounds. This framework has shown high performance in evaluating molecular properties such as the metabolism, brain permeability, and toxicity of drugs.

Wang et al. (2023) proposed a novel multimodal molecular pre-training framework, MolIG, for predicting molecular properties based on images and graph structures. The MolIG model effectively integrates the advantages of molecular graphs and images through self-supervised tasks, capturing key molecular structural features and high-level semantic information to enhance the prediction performance of molecular properties.

### 2.2 Prediction of drug-target interactions

Deep learning models, particularly GNNs, have been applied to predict the interactions between molecules and specific biological targets, which is vital for identifying new drug candidates and understanding their mechanisms of action. By learning from extensive drug-target interaction data, these models can predict the binding affinity of unknown molecules with targets, thereby accelerating the drug screening and optimization process.

The development of the structure-based deep convolutional neural network AtomNet was designed to predict the biological activity of small molecules suitable for drug discovery applications. By applying the concept of convolution, AtomNet successfully predicted new active molecules for targets previously lacking known modulators. Moreover, compared to traditional docking methods, AtomNet has demonstrated superior performance on multiple benchmark datasets, achieving an AUC of over 0.9 for 57.8% of targets on the DUDE

benchmark set.

Karimi et al. (2018) in their study, they proposed a semi-supervised deep learning model that combines recurrent neural networks and convolutional neural networks. This model can use both unlabeled and labeled data for joint encoding of molecular representation and affinity prediction. Their approach achieved a relative error within a factor of five in test cases, and up to a factor of twenty for proteins not included during training, showcasing its high accuracy and interpretative capability.

Öztürk et al. (2018) introduced the DeepDTA model, which uses convolutional neural networks (CNNs) to predict the binding affinity between drugs and proteins. This method utilizes the one-dimensional sequence information of drugs and targets, displaying good predictive performance. The main innovation of the DeepDTA model lies in its ability to effectively predict their binding affinity without the need for three-dimensional structural information of the drug and target. This method not only achieved performance superior to existing techniques on a larger benchmark dataset but also its use of drug and target sequence information allows the model to be widely applied to drugs and targets with unknown three-dimensional structures.

### **2.3 Toxicity prediction**

Deep learning is also employed to predict the potential toxicity of drug molecules, a crucial step in the drug development process. By analyzing the relationship between chemical structures and known toxicities, deep learning models can forecast potential toxicity issues of new molecules, aiding researchers to sidestep compounds likely to cause severe side effects at early stages.

The Tox21 program, initiated by multiple federal agencies in the United States, aims to predict the toxicity of chemicals through high-throughput screening (HTS) techniques combined with deep learning models, reducing the need for traditional animal toxicity studies.

According to Mayr et al. (2016), the DeepTox framework was developed to predict compound toxicity directly from molecular structures. In the Tox21 Challenge, DeepTox demonstrated superior performance. The DeepTox framework standardizes chemical characterizations of compounds, calculates a plethora of chemical descriptors as inputs for machine learning methods, trains models, and combines the best models into an ensemble to predict the toxicity of new compounds.

In the Tox21 Data Challenge, the GGL-Tox model showcased its accuracy and efficiency in toxicity analysis and prediction by integrating geometric deep learning with gradient boosting decision tree algorithms (Jiang et al., 2021).

Jimenez-Carretero et al. (2018) explored the potential of using deep convolutional neural networks (CNNs) to predict toxicity from pre-processed DAPI-stained cellular microscopy images of drugs. They found that the Tox-CNN model, based on nuclear profiling, outperformed other models in classifying cells by health status. This study validated the sensitivity and broad specificity of deep learning methods in predicting drug-induced toxicity from cellular images.

### **2.4 Drug design**

The application of deep learning in de novo drug design is particularly exciting. Using techniques such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), researchers can design new molecular structures that theoretically possess high activity and low toxicity. These methods allow scientists to explore and generate novel molecules with specific biological activity features from a vast chemical space, significantly accelerating the drug discovery process.

Kadurin et al. (2017) proposed an advanced autoencoder model, druGAN, for de novo generation of new molecular fingerprints with predefined anti-cancer properties. Compared to Variational Autoencoders (VAEs), druGAN offers advantages in the tunability of generated molecular fingerprints, capability in handling large molecular datasets, and efficiency in unsupervised pre-training for regression models.

Popova et al. (2017) introduced a novel computational strategy, ReLeaSE (Reinforcement Learning for Structural Evolution), which combines generative models with predictive models to generate new chemical structures aimed at compounds with desired physical and/or biological properties. This method can produce chemically feasible molecules and predict the required properties of newly generated compounds, aiding in the design of chemical libraries with specific physical properties such as melting points or hydrophobicity, or targeting specific biological markers like inhibitors against Janus kinase 2.

Yu and Welch (2021) developed MichiGAN, a novel neural network that combines the advantages of VAEs and GANs to sample single-cell RNA-seq datasets. This method allows manipulation of semantically distinct aspects of cell identity, predicting single-cell gene expression responses to drug treatments.

Through these application cases, we see the tremendous potential and diverse applications of deep learning technology in predicting drug molecule activity and drug design, showcasing the immense potential of artificial intelligence in modern pharmaceutical research. These technologies not only enable more accurate prediction of molecular pharmacological properties but also allow for innovation in drug design stages with unprecedented speed and efficiency. As deep learning algorithms and computational capabilities continue to advance, we can anticipate further breakthroughs in the fields of drug discovery and molecular design.

### 3 Challenges and Limitations

Despite the significant potential demonstrated by deep learning in predicting drug molecule activity, there are still challenges related to data quality and availability, model interpretability, generalization ability, and computational resource demands that need to be addressed through continuous research and technological innovation. These challenges involve not only improvements to the data and models themselves but also a deeper understanding, evaluation, and enhancement of existing methods.

#### 3.1 Data quality and availability

In the process of using deep learning to predict drug molecule activity, the quality, size, and diversity of datasets are key factors that affect model performance. High-quality datasets are essential for building accurate predictive models; however, many publicly available chemical and biological datasets suffer from issues of mislabeling, incompleteness, and insufficient updates (Jiménez-Luna et al., 2020). Additionally, data diversity is crucial for enhancing model generalization capabilities, but acquiring broad and varied data is often challenging, especially for rare or novel compound categories (Cai et al., 2020). Therefore, researchers need to invest significant efforts in data cleaning and preprocessing to enhance data quality and continually seek new data sources to increase dataset diversity.

#### 3.2 Model interpretability

Deep learning models, especially complex neural networks, are often seen as "black boxes" because their decision-making processes are difficult to interpret (Li et al., 2021). This characteristic is particularly problematic in scientific research and clinical applications, where decisions often require clear explanations and justifications. The lack of model interpretability limits the application of deep learning models in drug discovery, as researchers and clinicians need to understand the basis of model predictions to make informed decisions. Although recent years have seen some techniques aimed at improving model interpretability, this remains an active area of research that needs further exploration and innovation.

#### 3.3 Generalization ability

The generalization ability of deep learning models, or their capability to predict unseen data, is a critical metric for evaluating their performance. In scenarios of predicting drug molecule activity, models need to accurately predict the activity of molecules across different chemical spaces and biological environments. However, due to the vast and complex nature of chemical space, models might perform well on training sets but poorly on new, unseen molecules (Liu et al., 2019). Enhancing model generalization requires training with high-quality and diverse datasets, as well as employing advanced model architectures and regularization techniques to prevent overfitting.

### 3.4 Computational resource requirements

Deep learning models, particularly those with many parameters, demand substantial computational resources. This includes significant processor (CPU or GPU) capabilities and extensive memory requirements. For some research institutions and small businesses, the high computational costs may limit their ability to use deep learning for predicting drug molecule activity. Moreover, the training and optimization processes of complex models can be time-consuming, which might become a bottleneck in research and development environments that require rapid iterations and experiments. Therefore, finding more efficient models and algorithms with lower computational costs is a crucial direction in current deep learning research.

By addressing these challenges, the field can better leverage deep learning technologies to advance drug discovery processes effectively.

## 4 Conclusion and Prospects

Deep learning technology has shown tremendous potential and significant contributions in predicting the activity of drug molecules. By utilizing advanced algorithms such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Graph Neural Networks (GNNs), researchers can extract valuable features from complex molecular structures to predict the activity of drug molecules. These methods play a crucial role in accelerating drug discovery, reducing research and development costs, and enhancing prediction accuracy. In particular, the development of self-supervised learning frameworks like ImageMol has provided a new approach to processing unlabeled data, further expanding the application range of deep learning in drug design.

To fully leverage the potential of deep learning in drug discovery, interdisciplinary collaboration is strongly recommended. The close cooperation between chemists, biologists, and computer scientists can accelerate the process of discovering new drugs by sharing knowledge, data, and resources to jointly address challenges in drug design. Such collaboration can help develop more accurate and interpretable deep learning models, thereby improving the practicality and transparency of the models. Promoting communication and collaboration among scientists from different backgrounds will provide new perspectives and solutions for solving complex problems in drug discovery.

Future research should focus on enhancing the interpretability, generalization ability, and data efficiency of deep learning models. Specifically, developing new model interpretation tools will help researchers understand the molecular features and biological mechanisms behind model predictions, thereby increasing the transparency and trustworthiness of the models. Improving the models' generalization capability will ensure that deep learning algorithms maintain high performance across different chemical spaces and biological environments.

Technically, combining deep learning with cutting-edge technologies like quantum computing and augmented reality may open up new research directions in simulating complex molecular dynamics, exploring unknown chemical spaces, and designing personalized drugs. Developing algorithms that can effectively utilize small or imbalanced data will also be a key focus of future research, which is particularly important for accelerating the development of drugs for rare diseases and personalized medicine.

Deep learning offers new tools and methods for predicting drug molecule activity and discovering new drugs. Through interdisciplinary collaboration, continuous improvement of deep learning technologies, and exploration of its new applications in drug discovery, a more efficient, precise, and personalized drug development process is expected in the future.

## References

- Amabilino S., Pogány P., Pickett S., and Green D., 2020, Guidelines for RNN transfer learning based molecular generation of focussed libraries, *Journal of Chemical Information and Modeling*, 60(12): 5699-5713.  
<https://doi.org/10.1021/acs.jcim.0c00343>  
PMid:32659085

- Cai C., Wang S., Xu Y., Zhang W., Tang K., Qi O., Lai L., and Pei J., 2020, Transfer learning for drug discovery, *Journal of Medicinal Chemistry*, 63(16): 8683-8694.  
<https://doi.org/10.1021/acs.jmedchem.9b02147>  
PMid:32672961
- Chen D., Zheng J., Wei G., and Pan F., 2021, Extracting predictive representations from hundreds of millions of molecules, *The Journal of Physical Chemistry Letters*, 12(44): 10793-10801.  
<https://doi.org/10.1021/acs.jpcclett.1c03058>  
PMid:34723543 PMCID:PMC9358546
- Öztürk H., Özgür A., and Ozkirimli E., 2018, DeepDTA: deep drug-target binding affinity prediction, *Bioinformatics*, 34(17): i821-i829.  
<https://doi.org/10.1093/bioinformatics/bty593>  
PMid:30423097 PMCID:PMC6129291
- Hentabli H., Bengherbia B., Saeed F., Salim N., Nafea I., Toubal A., and Nasser M., 2022, Convolutional Neural Network Model Based on 2D Fingerprint for Bioactivity Prediction, *International Journal of Molecular Sciences*, 23(21): 13230.  
<https://doi.org/10.3390/ijms232113230>  
PMid:36362018 PMCID:PMC9657591
- Jiang J., Wang R., and Wei G., 2021, GGL-Tox: Geometric Graph Learning for Toxicity Prediction, *Journal of Chemical Information and Modeling*, 61(4): 1691-1700.  
<https://doi.org/10.1021/acs.jcim.0c01294>  
PMid:33719422 PMCID:PMC8155789
- Jimenez-Carretero D., Abrishami V., Fernandez-de-Manuel L., Palacios I., Quilez-Álvarez A., Díez-Sánchez A., Pozo M., and Montoya M., 2018, Tox\_(R)CNN: Deep learning-based nuclei profiling tool for drug toxicity screening, *PLoS Computational Biology*, 14(11): e1006238.  
<https://doi.org/10.1371/journal.pcbi.1006238>  
PMid:30500821 PMCID:PMC6291153
- Jiménez-Luna J., Grisoni F., and Schneider G., 2020, Drug discovery with explainable artificial intelligence, *Nat. Mach. Intell.*, (2): 573-584.  
<https://doi.org/10.1038/s42256-020-00236-4>
- Jones D., Kim H., Zhang X., Zemla A., Stevenson G., Bennett W., Kirshner D., Wong S., Lightstone F., and Allen J., 2020, Improved Protein-ligand binding affinity prediction with structure-based deep fusion inference, *Journal of Chemical Information and Modeling*, 61(4): 1583-1592.  
<https://doi.org/10.1021/acs.jcim.0c01306>  
PMid:33754707
- Kadurin A., Nikolenko S., Khrabrov K., Aliper A., and Zhavoronkov A., 2017, druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico, *Molecular Pharmaceutics*, 14(9): 3098-3104.  
<https://doi.org/10.1021/acs.molpharmaceut.7b00346>  
PMid:28703000
- Karimic M., Wu D., Wang Z.Y., and Shen Y., 2018, DeepAffinity: Interpretable Deep Learning of Compound-Protein Affinity through Unified Recurrent and Convolutional Neural Networks, *bioRxiv*, 35(18): 3329-3338.  
<https://doi.org/10.1093/bioinformatics/btz111>  
PMid:30768156 PMCID:PMC6748780
- Kayama K., Kanno M., Chisaki N., Tanaka M., Yao R., Hanazono K., Camer G., and Endoh D., 2021, Prediction of PCR amplification from primer and template sequences using recurrent neural network, *Scientific Reports*, 11: 7493.  
<https://doi.org/10.1038/s41598-021-86357-1>  
PMid:33820936 PMCID:PMC8021588
- Li X.H., Xiong H.Y., Li X.J., Wu X.Y., Zhang X., Liu J., Bian J., and Dou D.J., 2021, Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond, *Knowledge and Information Systems*, 64: 3197-3234.  
<https://doi.org/10.1007/s10115-022-01756-8>
- Liu J., Yang Y.H., Lv S., Wang J., and Chen H., 2019, Attention-based BiGRU-CNN for Chinese question classification, *Journal of Ambient Intelligence and Humanized Computing*, 10:1-12.  
<https://doi.org/10.1007/s12652-019-01344-9>
- Low K., Coote M., and Izgorodina E., 2022, Explainable Solvation Free Energy Prediction Combining Graph Neural Networks with Chemical Intuition, *Journal of Chemical Information and Modeling*, 62(22): 5457-5470.  
<https://doi.org/10.1021/acs.jcim.2c01013>  
PMid:36317829
- Mayr A., Klambauer G., Unterthiner T., and Hochreiter S., 2016, DeepTox: Toxicity Prediction using Deep Learning, *Frontiers in Environmental Science*, 3(8): 1-8.  
<https://doi.org/10.3389/fenvs.2015.00080>
- Noe M., and Peakman M., 2017, Drug Discovery Technologies, *Current and Future Trends*, 2: 1-32.  
<https://doi.org/10.1016/B978-0-12-409547-2.12312-1>  
PMid:28289437 PMCID:PMC5310094



- Popova M., Isayev O., and Tropsha A., 2017, Deep reinforcement learning for de novo drug design, *Science Advances*, 4(7): e7885.  
<https://doi.org/10.1126/sciadv.aap7885>  
PMid:30050984 PMCID:PMC6059760
- Schneider G., 2017, Automating drug discovery, *Nature Reviews Drug Discovery*, 17: 97-113.  
<https://doi.org/10.1038/nrd.2017.232>  
PMid:29242609
- Tang B., Kramer S., Fang M., Qiu Y., Wu Z., and Xu D., 2020, A self-attention based message passing neural network for predicting molecular lipophilicity and aqueous solubility, *Journal of Cheminformatics*, 12: 15.  
<https://doi.org/10.1186/s13321-020-0414-z>  
PMid:33431047 PMCID:PMC7035778
- Tran T., Tayara H., and Chong K., 2023, Artificial Intelligence in Drug Metabolism and Excretion Prediction: Recent Advances, Challenges, and Future Perspectives, *Pharmaceutics*, 15(4): 1260.  
<https://doi.org/10.3390/pharmaceutics15041260>  
PMid:37111744 PMCID:PMC10143484
- Walters W.P., and Barzilay R., 2021, Applications of Deep Learning in Molecule Generation and Molecular Property Prediction, *Acc. Chem. Res.*,54(2): 263-270.  
<https://doi.org/10.1021/acs.accounts.0c00699>  
PMid:33370107
- Wang Z., Mi J., Lu S., and He J., 2023, MultiModal-Learning for Predicting Molecular Properties: A Framework Based on Image and Graph Structures, *ArXiv*, abs/2311.16666.
- Wieder O., Kuenemann M., Wieder M., Seidel T., Meyer C., Bryant S., and Langer T., 2021, Improved Lipophilicity and Aqueous Solubility Prediction with Composite Graph Neural Networks, *Molecules*, 26(20): 6185.  
<https://doi.org/10.3390/molecules26206185>  
PMid:34684766 PMCID:PMC8539502
- Xiong J., Xiong Z., Chen K., Jiang H., and Zheng M., 2021, Graph neural networks for automated de novo drug design, *Drug Discovery Today*, 26(6): 1382-1393.  
<https://doi.org/10.1016/j.drudis.2021.02.011>  
PMid:33609779
- Yaseen B., 2023, Drug Target Interaction Prediction Using Convolutional Neural Network (CNN), 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2(2): 1-5.  
<https://doi.org/10.1109/HORA58378.2023.10156717>
- Yu H., and Welch J., 2021, MichiGAN: sampling from disentangled representations of single-cell data using generative adversarial networks, *Genome Biology*, 22: 158.  
<https://doi.org/10.1186/s13059-021-02373-4>  
PMid:34016135 PMCID:PMC8139054
- Zeng X., Xiang H., Yu L., Wang J., Li K., Nussinov R., and Cheng F., 2022, Accurate prediction of molecular targets using a self-supervised image representation learning framework, *Research Square*, 4: 1004-1016.  
<https://doi.org/10.1038/s42256-022-00557-6>
- Zhang Q., Shen Z., and Huang D., 2019, Predicting in-vitro transcription factor binding sites using DNA Sequence+Shape, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(2): 667-676.  
<https://doi.org/10.1109/TCBB.2019.2947461>  
PMid:31634140