CHEMBIOINFO-09: Cheminfo., Chemom., Bioinfo., Comput. & Quantum Chem. Congress München, GR-Cambridge, UK-Ch. Hill, USA, 2023.

Toward Artificial Intelligence Era in Drug Discovery and Design

Shan He

Department of Organic and Inorganic Chemistry, Faculty of Science and Technology, University of the Basque Country UPV/EHU, P.O.Box 644, 48080 Bilbao, Spain.
IKERDATA S.L INVESTIGO Next-generation Researcher.

Abstract. In the last decades, we have experienced a revolution in data science in terms of the huge amount of data to be analyzed (era of big data) and the availability of high-performance processors. In drug discovery, this scenario is not different: the large volume of data (chemical, biological, etc.) along with the automation of techniques have generated a fertile ground for the use of artificial (or computational) intelligence/Machine Leaning (AI/ML). This powerful tool helped the researchers to achieve several major theoretical and applied breakthroughs. In this mini-review, recent research work of AI/ML in drug discovery and design will be introduced.

In the last decades, we have experienced a revolution in data science in terms of the huge amount of data to be analyzed (era of big data) and the availability of high-performance processors. In drug discovery, this scenario is not different: the large volume of data (chemical, biological, etc.) along with the automation of techniques have generated a fertile ground for the use of artificial (or computational) intelligence/Machine Leaning (AI/ML). Furthermore, new drugs discovery includes many stages, which makes this process slow, costly, and leading to failures at the end.[1] The scenario gets worse if we take into account the thousands or millions of molecules capable of being synthesized in each development stage.[2] In this sense, AI/ML is gaining more and more attention in drug discovery and design because it can speed up this procedure and makes it less time and source-consuming. In this mini-review, recent research work of AI/ML in drug discovery and design will be introduced.

Pereira *et al.* proposed a deep learning approach to improve docking-based virtual screening. The deep neural network that is introduced, DeepVS, uses the output of a docking program and learns how to extract relevant features from basic data such as atom and residues types obtained from protein−ligand complexes. They used atom and amino acid embeddings and implemented an effective way of creating distributed vector representations of protein−ligand complexes by modeling the compound as a set of atom contexts that is further processed by a convolutional layer. Also, they evaluated DeepVS on the Directory of Useful Decoys (DUD), using the output of two docking programs: Autodock Vina1.1.2 and Dock 6.6. Using a strict evaluation with leave-one-out cross-validation, DeepVS outperforms the docking programs, with regard to both AUC ROC and enrichment factor. Moreover, using the output of Autodock Vina1.1.2, DeepVS achieves an AUC ROC of 0.81.[3]

Another example within this topic, Zhang et al. suggested two techniques to improve the compound selectivity prediction. They used an improved multitask learning method in Neural Networks (NNs), that involved activity and selectivity for other targets along with a logistic regression. Then, they optimized the compound selectivity prediction by applying the multitask learning method in Deep Belief Networks (DBNs). It helps build a distributed representation model and improve the generalization of the shared tasks. Moreover, they allocated many weights to the supplementary tasks that were connected to the main selectivity prediction task. Compared to other work, those methods significantly increase the accuracy of the compound selectivity prediction, especially, employing the multitask approach in DBNs with adjusted weights gets the best performance.[4]

As last example of using AI/ML in drug discovery and design, Tsubaki *et al.* developed a new method that could predict protein sequences and molecular fingerprints of ligands. In this work, the authors used convolutional neural network and graph neural network, respectively as vector input to predict protein ligand interactions. This method was trained and validated by using some techniques, such as kNN (k-Nearest Neighbor), random forest, logistic regression, and SVM (Support Vector Machine). In conclusion, the authors determined that the vectors achieved from the models can accurately predict the most important amino acid residues at the binding site, which were responsible for drug-target interactions.[5]

To conclude, the use of AI/ML is experimenting a revolution, not only in drug discovery but also in chemistry, nanoscience, biology, medicinal chemistry and so on. This state-of-art technology can reduce the time, experimental source and transform the trial and error into less tedious process. For example, in

this mini-review, all the authors used this tool in order to optimize the interaction between protein-ligand, obtaining high statistical parameters.

REFERENCE

1. Lipinski, C. F.; Maltarollo, V. G.; Oliveira, P. R.; Da Silva, A. B.; Honorio, K. M. Advances and perspectives in applying deep learning for drug design and discovery. *Frontiers in Robotics and AI* **2019,** *6*, 108.
2. Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening—an overview. *Drug discovery today* **1998,** *3* (4), 160-178.
3. Pereira, J. C.; Caffarena, E. R.; Dos Santos, C. N. Boosting docking-based virtual screening with deep learning. *Journal of chemical information and modeling* **2016,** *56* (12), 2495-2506.
4. Zhang, R.; Li, J.; Lu, J.; Hu, R.; Yuan, Y.; Zhao, Z. Using deep learning for compound selectivity prediction. *Current Computer-Aided Drug Design* **2016,** *12* (1), 5-14.
5. Tsubaki, M.; Tomii, K.; Sese, J. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics* **2019,** *35* (2), 309-318.