Is AI/ML Assisted Design-to-Synthesis of Drug-like Molecules Feasible?

U. Deva Priyakumar*
Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad 500 032
**deva@iiit.ac.in*

The three paradigms of scientific discovery and problem solving are experimentation, theory, and computation. Recent years are witnessing the emergence of data driven methods as the fourth paradigm not only in technology areas but also in fundamental sciences. Availability of data from previous experiments and computations combined with modern machine learning (ML) methods enable recognizing complex patterns that are helpful in discovering new molecules and materials. Such methods have been successfully applied to number of problems in chemistry and have been shown to provide much faster/accurate solutions compared to the traditional computational methods such as quantum mechanical calculations and molecular dynamics simulations. We will introduce 'inverse problems' in chemistry, discuss how solving this problem will enable efficient solutions and how modern ML methods helps us address this problem. The use of variational autoencoder, reinforcement learning and transformer-decoder model for generation of novel molecules from unexplored chemical space exhibiting properties of interest. Similarly, how machine learning could be used to explore the known chemical space more efficiently than traditional methods will be discussed. This method makes the chemical space sampling ~20x faster. Finally, methods by which molecular structure could be elucidated directly from molecular spectra will be presented. In summary, a pipeline for automating design to synthesis will be discussed followed by the plausible role of modern machine learning methods in solving some of the steps in this pipeline will be presented.

Keywords: artificial intelligence; machine learning; molecular design; property prediction; deep learning

Figure 1: Direct and Inverse molecular design