GenLSDD: A Deep Generative Approach to Ligand and Structure-Based Drug Design

Yoon, Elliott (School: Tenafly High School)

Deep learning, particularly generative models, have shown promise to expedite the drug discovery process by generating novel molecules with desirable properties/bioactivity. Current research on generative drug design follows one of two main approaches: ligand-based or structure-based. The ligand-based approach trains a model on structures of known small-molecule ligands of a target protein, while the structure-based approach trains a model directly on the structure of the target protein, both with their own inherent strengths and weaknesses. This research proposes Generative Ligand and Structure-based Drug Design (GenLSDD), a deep learning model for generative drug design that combines the strengths while counteracting the weaknesses of both approaches. GenLSDD consists of two modules. The first module is a ligand-based recurrent neural network trained to generate chemically viable molecular structures. The second module is a structure-based graph neural network trained to predict an input molecule's potency in inhibiting a target protein, measured in half-maximal inhibitory concentration (IC50). The two modules were then trained in tandem via reinforcement learning, with the structure-based predictive module providing reward values to optimize the ligand-based generative module towards generating molecules with greater potency. As a case study, GenLSDD was applied to generate inhibitors targeting MET, a tyrosine kinase receptor linked with various forms of cancer. Out of a sample of 1000 molecules generated, 23 were predicted to be highly potent (IC50 < 10 nM). Additionally, GenLSDD, achieved similar or better performance than state-of-the-art models on all evaluated metrics. Overall, GenLSDD represents a powerful deep learning tool to enhance the drug design process.