A Novel Approach to Systemic in silico Drug Discovery Using a Machine Learning-Based Prediction of Binding Affinity

Takahashi, Fuga (School: Novi High School)

Most drugs function by binding to a “target” receptor, activating or inhibiting the “target” signal transduction pathway, which, in ~50% of drug trials, fails to occur, leading to low efficacy. Moreover, serious adverse reactions–killing 1.1 million patients every year–can occur when the drug ligand has strong binding affinities with other, non-target receptors, turning on non-target pathways inadvertently. To discover such issues, a drug must undergo lengthy and expensive clinical trials; recently, machine learning models have been proposed to facilitate this process by predicting the binding affinity between 1 drug and 1 receptor, but are held back to testing 1 at a time. Other than inefficiency, they are plagued by the SAR paradox and overfitting, while being unable to predict binding affinities between a drug and all the other receptors in the body, and therefore unable to predict adverse reactions. In turn, I developed a novel machine learning approach based around a representation of drugs using simplified molecular-input line-entry systems (SMILES), and representation of protein receptors with amino acid sequences, in order to predict the binding affinity between a drug and a wide database of receptors. The model uses a conditional Generative Adversarial Network, a Classification Model, and a Regression Model to produce a numerical value for binding affinity, measured in Kd or Ki, and has final accuracy values of 67%, 91%, and 78-87%, respectively. When used together, the proposed model accurately predicted adverse reactions for multiple drugs with a minimum of 0 false positives, thus exceeding clinical accuracy in certain scenarios, and showing promise for further application to facilitate the drug discovery process.