DeepChem: Using Generative Adversarial Networks to Develop Novel Inhibitors of Carcinomas

Deep, Harsh (School: The Harker School)
Mukhija, Krishay (School: The Harker School)

In our project, we sought to use machine learning to generate new lead molecules and expedite drug discovery. Lead molecules are molecules with an affinity for inhibiting the target protein, and identifying suitable leads is the first stage in drug discovery. Our approach for generating lead molecules differed from current methodologies in 2 areas. First, instead of only training our network on drugs that affected the target protein, we found drugs that targeted proteins that were structurally similar to the target protein. This helped create a more diverse dataset and helped augment the dataset for proteins that currently do not have many drugs that can inhibit them. Secondly, whereas many previous machine learning approaches to drug discovery utilized Recurrent Neural Networks, we used Generative Adversarial Networks(GANs) which provide the benefit of having 2 competing neural networks that continuously learn. Over time, the generator becomes more adept at producing realistic drugs and the discriminator becomes better at discerning which drugs are real versus generated. The effect of this is that the generated molecules more closely resemble the style of input molecules while also containing slight variations. Overall, our generated inhibitors had an average docking score of -7.9 which was better than that of current lead molecules which sat at -7.1, a more negative docking score is preferred; however, our molecules were still below the average docking scores of current carcinoma drugs which had an average docking score of -9.3.

Awards Won: