Treating COVID-19 With Machine Learning

Zhang, Evan (School: Dover-Sherborn High School)

The past three years have seen SARS-CoV-2 wreak havoc on society with limited effective treatments available due to the amount of trial and error involved. However, with recent advancements in artificial intelligence, it is ready now more than ever to fight SARS-CoV-2. A method to discover novel antiviral drugs can be obtained using Chemprop, a machine-learning backbone for predicting molecular properties. This is done by passing hundreds of thousands of data points into a classifier model. The dataset includes molecular information represented by SMILES strings and the observed efficacy in inhibiting SARS-CoV-2 in laboratory tests. The resulting model can predict the effectiveness of untested molecules, which then can be manually tested, minimizing the time-consuming hunting traditionally done by human scientists. With promising performance, the proposed method pushes the boundary of what machine learning can do in drug research. The trained model achieved a high accuracy in predicting the effectiveness of drugs against SARS-CoV-2 with an AUC score of 0.8455. However, the model loses accuracy when predicting the effectiveness of drugs against SARS-CoV, a different strand of coronavirus, with an AUC of 0.7302. Finally, the model was run on an unseen dataset to find the molecule most likely effective against COVID-19. The result was a molecule with SMILES string CN1CCN(CC1)C(=O)COC=2C=CC(C)=CC2 also called 1-(4-Methyl-piperazin-1-yl)-2-p-tolyloxy-ethanone. The model's ability to find likely drugs can hasten drug research drastically, potentially saving countless lives during future pandemics.