Preparing for a future pandemic with artificial intelligence

September 21 2021, by Melissae Fellet

A case study explored using counterfactual reasoning algorithms to test how artificial intelligence might be able to predict patient outcomes using biomedical information. Credit: Shannon Colson | Pacific Northwest National Laboratory

When the novel coronavirus led to a global pandemic last year, doctors and researchers rushed to learn as much as possible about the virus and how our bodies respond to it.

They needed a lot of information, and they needed it fast. Doctors studied whether available medicines could effectively treat the symptoms of COVID-19. Virologists, biologists, and chemists scrambled to understand how the virus affects the molecular workings of cells, information key to designing medicine to treat infection and resulting disease.

Medical and biological data flowed fast and furiously. More than four percent of the world's research published in 2020 was [related to COVID](#), according to the Dimensions database produced by Digital Science. Yet each study provided just a piece of insight into the massive biological puzzle that defines this severe respiratory syndrome

Finding meaning in a sea of messy or incomplete data is precisely what data scientists at Pacific Northwest National Laboratory (PNNL) do. With expertise in applying graph-based machine learning, detailed molecular modeling, and explainable AI to questions of national security and basic science, PNNL researchers are now turning their artificial intelligence tools to the study of fundamental questions about treatments for COVID. What they are learning sharpens the tools available in the computational toolbox for responding quickly to a future pandemic.

Imagining individual treatment effects through counterfactual reasoning

Each time COVID-19 cases surge in another place around the world, access to treatments becomes a concern. When there have been more sick patients than treatment supply, doctors have made difficult

decisions about how to use the available medical resources for greatest benefit.

One type of thinking that can be part of those decisions is counterfactual reasoning. This involves comparing the outcomes of patients who received treatment with their imagined outcomes if counter to fact, they had not been treated based on knowing how similar situations with previous patients turned out.

Artificial intelligence algorithms can also use counterfactual reasoning, provided they have enough prior knowledge to draw on. The amount of COVID-related research last year provided computational scientist Jeremy Zucker and his colleagues with a trove of biochemical details about the novel coronavirus and how our immune systems respond to it.

Taken together, those details can be represented by a data science approach called a knowledge graph. The team used that knowledge graph to derive a counterfactual model for answering a specific scientific question about COVID-19 treatment outcomes.

"With data science that leverages biomedical experimental knowledge about COVID disease progression and treatment response, artificial intelligence can learn to more precisely predict the effect of treatments on individual patient outcomes," Zucker said.

The team applied such an artificial intelligence framework to simulate particular biochemical data collected from hypothetical patients who were severely ill with COVID-19. Each patient had different viral loads, was administered a different dose of a drug, and either recovered or died.

In each case, the team wanted to predict whether a patient who survived would have died had they not been treated with the drug, or if they died,

whether they would have survived had they been given a higher dose of the drug.

The analysis provided more precise information about the treatment's potential benefit to individual patients, compared with algorithms that simply predicted average patient outcomes following treatment.
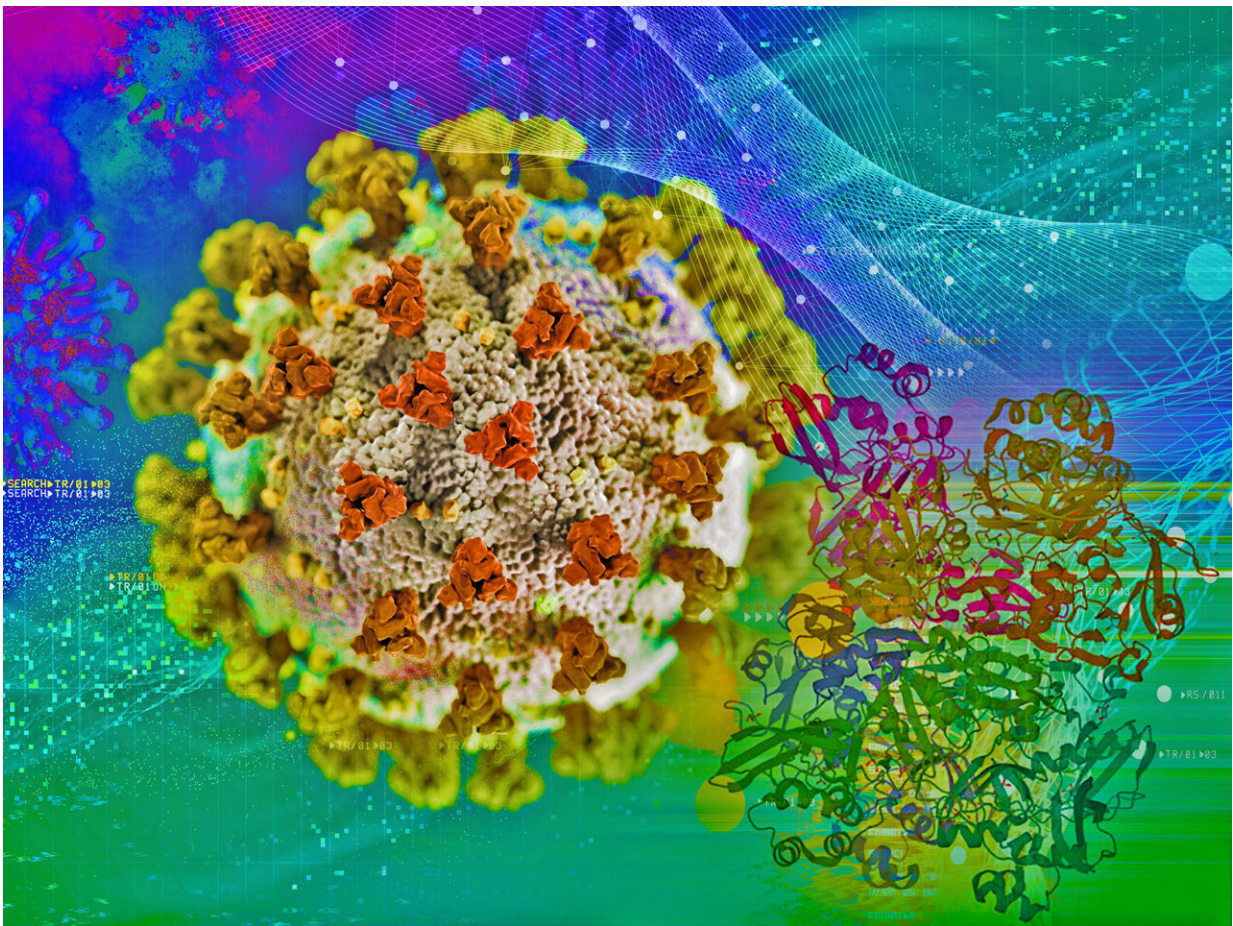
High-throughput biochemical assays targeting a vital viral protein, combined with artificial intelligence-based screening, identified one molecule, out of more than 13,000 tested, with promising antiviral activity against SARS-CoV-2. Credit: Timothy Holland | Pacific Northwest National Laboratory

The scientists reported several case studies of their counterfactual reasoning algorithm in a paper published in a recent special issue of *IEEE Transactions on Big Data* on COVID-19 and artificial intelligence. This work is part of the PNNL-funded Mathematics for Artificial Reasoning in Science (MARS) initiative, and is being applied and evaluated on a DARPA Modeling Adversarial Activity project, which is using causal knowledge graphs at scale to combat COVID-19.

Molecular modeling to assist drug repurposing

Although vaccines for the novel coronavirus are increasingly available around the world, it will take time to slow the spread of the virus and its variants. Therefore, medicines to treat COVID-19 are still needed, and existing approved medicines originally developed for other diseases might be useful.

A team of scientists from PNNL and the University of Washington (UW) School of Medicine screened more than 13,000 compounds from existing drug libraries for the ability to inhibit a vital protein produced by genetic information in the novel coronavirus SARS-CoV-2. Using a series of high-throughput biochemical measurements combined with artificial intelligence-based screening, their work identified one molecule out of that collection with promising antiviral activity against SARS-CoV-2.

Wesley Van Voorhis and his UW team used a cascade of biochemical tests to winnow the thousands of molecules down to three hits that were potent inhibitors in experiments with purified protein.

At PNNL, data scientist Neeraj Kumar and his colleagues used artificial intelligence-based molecular modeling to predict where each hit bound to the viral protein, called nsp15. Chemist Mowei Zhou conducted mass spectrometry measurements of each hit associated with nsp15 in its

natural folded form, using resources at the Environmental Molecular Sciences Laboratory (EMSL), a U.S. Department of Energy Office of Science user facility located at PNNL. These measurements provided information about how tightly each compound bound to nsp15, and confirmed that one of the three compounds, a molecule called Exebryl-1, bound to the protein.

In results published in the journal *PLoS ONE*, the team showed that Exebryl-1 exhibited modest antiviral activity against SARS-CoV-2.

Exebryl-1 was originally designed to treat Alzheimer's disease. In screening tests, it did not have sufficient antiviral activity to be considered an immediate candidate for COVID-19 treatment. However, artificial intelligence may help scientists tweak the structure of Exebryl-1 to improve its antiviral activity against the novel coronavirus.

This work was supported through the National Virtual Biotechnology Laboratory, a consortium of all 17 U.S. Department of Energy national laboratories focused on response to COVID-19, with funding provided by the Coronavirus Aid, Relief, and Economic Security (CARES) Act.

Developing an approach to speed drug discovery during this pandemic could reveal new design steps that might be useful during the next outbreak.

"Drug research and development is a complex, costly, and time-consuming process, particularly considering the majority of molecules advanced from the design phase fail in clinical trials," Kumar said. "Computer-based screening incorporates chemical information during the design process to increase a drug candidate's potential for success in clinical testing."

Graph neural networks could generate tailor-made

therapeutics

Another way to use artificial intelligence for drug design could be to create libraries of possible drug candidates that have never been seen before.

Researchers at Pacific Northwest National Laboratory are exploring different methods of artificial intelligence using graph neural networks to generate libraries of molecular structures for drug discovery. Credit: Shannon Colson | Pacific Northwest National Laboratory

Chemists who develop medicines can identify key features of a molecular structure that make it work. They can also dissect a structure to estimate how challenging it might be to make a molecule.

PNNL computer scientist Sutanay Choudhury, data scientist Neeraj Kumar, data scientist Jenna Pope, and their colleagues at Argonne National Laboratory can recreate the same thought process with artificial intelligence. The team is using graph neural networks to generate structures for molecules that could be candidates for drug development.

Graphs provide a mathematical representation of the connections among items in a network; for example, how atoms in a molecule connect to make a potential drug candidate. Neural networks based on such molecular graphs can learn to find patterns in data that may not otherwise be obvious.

To test their methods for drug design, Choudhury, Kumar, Pope, and their colleagues mapped ways to connect chemical components to make a drug-like molecule, and identified which components contribute to how a molecule behaves as a drug. Finally, they tested two methods of using graph neural networks to piece together molecules using chemically relevant equations.

The scientists presented a workshop paper at the ninth International Conference on Learning Representations, where they compared these methods to design molecules that might inhibit a key SARS-CoV-2 protein called protease.

One method learned structural patterns from more than 7,000 molecules known to inhibit various viral proteases. The team found this analysis tended to generate molecules similar to those in the known database.

In the other method, the algorithm built a molecule atom by atom and

bond by bond, optimizing the desired drug and synthetic properties throughout the virtual construction. The team found this method tended to produce molecules that had not been known before.

Each approach has different benefits for drug development. Repurposing approved drugs could be a fast track to the clinic, and generating entirely new molecular structures injects variation early in the notoriously difficult search for antivirals, Kumar said.

This graph neural network research was part of PNNL's contribution to the Department of Energy's (DOE's) ExaLearn Co-Design Center, a group of eight national laboratories focusing on machine learning technologies.

This center is a product of DOE's Exascale Computing Project, which was launched in 2016 to explore the most intractable supercomputing. Data scientist Draguna Vrabie leads PNNL's participation in the ExaLearn center.

Fundamental research for the future

When an influenza pandemic spread around the world about a century ago, scientists did not know viruses existed; they looked for a bacterial cause for the disease. During the coronavirus pandemic, scientists had sequences of genetic information to track the spread of the virus and its variants, molecular details to develop rapid diagnostic tests, and tools to develop an entirely new class of vaccines. Some of these vaccines were authorized in the U.S. within a year after the novel coronavirus was first discovered.

A hallmark of artificial intelligence is its ability to learn from the past. As PNNL researchers advance and refine AI applications, it could increasingly become part of routine research, too—the type of work that

supported the advances toward tackling this pandemic and can support the response to a future one, too.

This research, which brings together PNNL's strengths in host response to infectious disease, artificial intelligence, and advanced data analytics, is part of a series of PNNL findings about COVID-19. Other PNNL authors on these three papers include Craig Bakker, Jeremy Teuton, Kristie Oxford, Jesse Wilson, Rhema James, and Garry Buchko.

More information: Jeremy Zucker et al, Leveraging Structured Biological Knowledge for Counterfactual Inference: A Case Study of Viral Pathogenesis, *IEEE Transactions on Big Data* (2021). DOI: 10.1109/TBDATA.2021.3050680

Ryan Choi et al, High-throughput screening of the ReFRAME, Pandemic Box, and COVID Box drug repurposing libraries against SARS-CoV-2 nsp15 endoribonuclease to identify small-molecule inhibitors of viral activity, *PLOS ONE* (2021). DOI: 10.1371/journal.pone.0250019