In the latest human vs. machine match, artificial intelligence wins by a hair

2 November 2022

Vikas Nanda has spent more than two decades studying the intricacies of proteins, the highly complex substances present in all living organisms. The Rutgers scientist has long contemplated how the unique patterns of amino acids that comprise proteins determine whether they become anything from hemoglobin to collagen, as well as the subsequent, mysterious step of self-assembly where only certain proteins clump together to form even more complex substances.

So, when scientists wanted to conduct an experiment pitting a human—one with a profound, intuitive understanding of protein design and self-assembly—against the predictive capabilities of an artificially intelligent computer program, Nanda, a researcher at the Center for Advanced Biotechnology and Medicine (CABM) at Rutgers, was one of those at the top of the list.

Now, the results to see who—or what—could do a better job at predicting which protein sequences would combine most successfully are out. Nanda, along with researchers at Argonne National Laboratory in Illinois and colleagues from throughout the nation, reports in *Nature Chemistry* that the battle was close but decisive. The competition matching Nanda and several colleagues against an artificial intelligence (AI) program has been won, ever so slightly, by the computer program.

Scientists are deeply interested in protein self-assembly because they believe understanding it better could help them design a host of revolutionary products for medical and industrial uses, such as artificial human tissue for wounds and catalysts for new chemical products.

"Despite our extensive expertise, the AI did as good or better on several data sets, showing the tremendous potential of machine learning to overcome human bias," said Nanda, a professor in the Department of Biochemistry and Molecular Biology at Rutgers Robert Wood Johnson Medical School.

Proteins are made of large numbers of amino acids joined end to end. The chains fold up to form three-dimensional molecules with complex shapes. The precise shape of each protein, along with the amino acids it contains, determines what it does. Some researchers, such as Nanda, engage in "protein design," creating sequences that produce new proteins. Recently, Nanda and a team of researchers designed a synthetic protein that quickly detects VX, a dangerous nerve agent, and could pave the way for new biosensors and treatments.

For reasons that are largely unknown, proteins will self-assemble with other proteins to form superstructures important in biology. Sometimes, proteins look to be following a design, such as when they self-assemble into a protective outer shell of a virus, known as a capsid. In other cases, they self-assemble when something goes wrong, forming deadly biological structures associated with diseases as varied as Alzheimer’s and sickle cell.
"Understanding protein self-assembly is fundamental to making advances in many fields, including medicine and industry," Nanda said.

In the experiment, Nanda and five other colleagues were given a list of proteins and asked to predict which ones were likely to self-assemble. Their predictions were compared to those made by the computer program.

The human experts, employing rules of thumb based on their observation of protein behavior in experiments, including patterns of electrical charges and degree of aversion to water, chose 11 proteins they predicted would self-assemble. The computer program, based on an advanced machine-learning system, chose nine proteins.

The humans were correct for six out of the 11 proteins they chose. The computer program earned a higher percentage, with six out of the nine proteins it recommended able to self-assemble.

The experiment showed that the human experts "favored" some amino acids over others, sometimes leading them to incorrect choices. Also, the computer program correctly pointed to some proteins with qualities that didn't make them obvious choices for self-assembly, opening the door to further inquiry.

The experience has made Nanda, once a doubter of machine learning for protein assembly investigations, more open to the technique.

"We're working to get a fundamental understanding of the chemical nature of interactions that lead to self-assembly, so I worried that using these programs would prevent important insights," Nanda said. "But what I'm beginning to really understand is that machine learning is just another tool, like any other."
