

AI, Biomedicine, and Pandemics: Directions and Opportunities

IJCAI 2020 Panel Discussion on AI Approaches for COVID-19 and Future Pandemics

Eric Horvitz

January 2021

Numerous advances in biomedicine have provided the methods and tools needed to mount a fast-paced response to the COVID-19 pandemic. With unprecedented speed, understandings of the structure and function of the SARS-CoV-2 virus were developed, followed by the design of promising therapeutics and vaccines. Within weeks of its detection in Wuhan, the sequence of the single-stranded RNA virus was published. Two months later, critical details of the virus' structural and non-structural proteins were identified and the virus' mode of entry into human cells was decoded as an interaction between the SARS-CoV-2 spike protein and the human ACE2 receptor. Over the next several months, promising vaccines and therapeutics were designed and testing was commenced of vaccine safety and efficacy.

The response of the scientific community to the pandemic has been impressive in the context of the history of vaccine design. Still, many hundreds of thousands of people perished and many more suffered from the direct and indirect influences of COVID-19. Now, at the one-year mark, as we celebrate the start of the worldwide distribution of efficacious vaccines, a challenge comes to mind: How might we go from detection to the development of an effective therapy or vaccine in months or even weeks versus a year?

I believe that we can significantly reduce the time from the detection of a rising respiratory viral threat to developing understandings of structure and mechanism to producing and fielding mitigations. While multiple advances will contribute to a speed-up, AI technologies will play a significant role—particularly supervised and unsupervised machine learning with deep neural networks (DNNs).

Advances in DNNs have already been complementing more traditional methods for identifying protein structure. Current mainstays include imaging of protein structure through cryo-electron microscopy, performing biophysical simulations to identify feasible conformations of proteins via considerations of energies and constraints, and molecular dynamics calculations to build insights about conformational dynamics of proteins as they interact. DNN methods have been developed to predict protein structure from amino acid sequences and to learn representations that link protein function to sequence and structure. These capabilities will play an increasingly important role in helping biomedical researchers to gain understandings of the operation of viruses more quickly by identifying sets of hypotheses about protein structure and function.

Beyond understanding natural proteins, machine learning methods will be critical in enabling faster and lower-cost generation and prioritization of sets of protein candidates considered in *de novo* protein design. To date, protein design efforts in synthetic biology must grapple with large combinatorial spaces, relying on cycles of analysis that require a time-consuming intermixture of computation and wet-lab experimentation to screen computationally identified candidates. For example, the design process involved in the *de novo* creation of small, stable minibinder proteins targeting the SARS-CoV-2 Spike protein [1] required (1) generating millions of candidate proteins, (2) identifying synthetic genes for a promising subset of ~100,000 candidates, and then (3) the experimental screening of these candidates for binding activity. The candidates with highest binding activities were then refined via (4) additional computation-driven refinements and (5) additional cycles of experimental testing. More accurate designs can reduce the time and costs and raise the yield of these cycles.

Another promising direction is to harness machine learning and decision-theoretic inference in the guidance of computation and wet-lab experimentation. The allocation of scarce computing and experimentation resources can be directed via considerations of current uncertainties about protein structure and function. There is great promise in harnessing machine learning methods to compute well-calibrated uncertainties in sets of predictions about protein structure and function to guide computation and experimentation. Whether uncertainties are epistemic (due to limitations of the models) or aleatoric (per intrinsic noise or incompleteness in observations), estimates of uncertainty can be harnessed to determine the expected value of additional computational analyses and the expected value of information of specific wet-lab experiments. Such expectations consider how uncertainties can be expected to be reduced via computation or experimentation, with a goal of minimizing time or costs to achieving goals of therapeutics or vaccines.

References

[1] Cao L, *et al.* [De novo design of picomolar SARS-CoV-2 miniprotein inhibitors](https://doi.org/10.1126/science.abd9909). Science. 09 Sep 2020; DOI: 10.1126/science.abd9909