COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Hong Qin (University of Tennessee, Chattanooga), July 26, 2023

Title: Develop and Evaluate Computational Frameworks to Predict and Prevent Future Coronavirus Pandemics

Hong Qin CIC Database Profile

NSF Award #: 2200138

YouTube Recording with Slides

Summer 2023 CIC Webinar Information

Transcript Editor: Lauren Close

Transcript

Slide 1

Hi, my name is Hong Qin and I'm faculty at the University of Tennessee, Chattanooga. This talk is about a project I was fortunate enough to work on with a large team of colleagues from North Carolina A&T, Spelman College, Catholic University of America. These are the collaborators on paper, but in reality, there are many many more. In our collaboration we probably had more than a dozen team members that have been working on this.
So this is a National Science Foundation supported project. It's called PIPP Phase One. My goal is to develop an AI based framework to predict and prevent future coronavirus pandemics. Although we say it's coronavirus, the model now has the possibility to be generalizable to presumably many other viral pandemics.

Slide 2
So this is the - ok, it's automatically forwarding - my apologies. Ok, so this is the overview of we proposed in the beginning, although we we have made a lot of modifications right now. The idea is still the same. We propose how do we predict the virus - a new virus pandemic? Our idea is first we generate all the possible SARS-CoV-X from that are predictable based on recombination patterns or habitat change. We then predict those potential SARS-CoV-X reactions. Here's the challenge - it's easy to generate those sequences, but how do we know which one actually will become virulent or transmit in a human population? That's really the key here. So we turn to AI. When I say AI is a black box, well in this case that's probably also a blessing because we really

don't know how a virus jumps from host, from an animal, to a human to become virulent. So AI is a black box but it's probably a great tool we can use even though it may be a black box. That's my argument here.
And my apology, I had to black out certain things because my University is applying for a patent based on the function of this. So the goal is here we have input sequences - this is just primary nucleotide sequences. From this sequence, we will do feature engineering to generate some helpful features. Then we use the sequence feature to feed into our AI model. Based on our AI model we will predict the finish. Given the recent advance of AI, we can also use the [inaudible] dynamic simulation or even experimental methods to verify some candidate genes as the potential drivers for the virulence change. Given the AI model, we also generalize this - basically use transfer learning from coronaviruses, presumably to other viruses such as influenza or HIV, HPV. There's many viruses. I'm not actually really not expert in biology, I'm just computer science.
And so the challenges we are really dealing with - my apologies, it's automatically forwarding again. The challenge is really to predict pathogenicity from sequences. We can also found a potential rule and test the general ability by transforming predict the - future viruses and predict the outcome. It will be AI enabled, the early warning system. We want to use AI to predict vaccine development. Given that we know how the virulence variance changes. So that's the big picture of this.

Slide 3
So here are some results we have developed - that we have accomplished. This is an estimation of the viral fitness based on the information - the data we have. This is actually for the Omicron sub-variant and we can compare the Omicron variant using a method with development pairwise comparison. Then we can estimate the relative difference between those variants.

Slide 4
Based on this data we can generate a so-called viral fitness landscape. For those aware of the theory of evolution - in evolution there is a so-called fitness landscape and there's really a lot of arguments about that, but there's also a lot of theory based on the evolution landscape. With the fitness landscape of a potential virus, we could in theory predict its evolutionary trajectory. This - we can - we have a method to do this. This is the estimated result for the alpha beta delta omicron subvariants in the USA. We are also in the process of applying this to influenza or smallpox. We don't have enough data for smallpox or other viruses, but influenza seem to be promising. So this is a significant finding of this - our current project.

Slide 5
Here's some model results. Without divulging too much detail into the AI model we have developed. There's really a lot of parameters, but the accuracy of current model can achieve quite a high accuracy. Of course, this is based on what we know. The challenge is really about the

unknown. Even though we are predicting the unknown, the data is already known. In machine learning we train it with input training and testing. The challenge about predicting the unknown still remains an open question.

Slide 6

Some people may argue, well why do we have to do deep learning? There is a lot of statistical, genetics, genomics, quantitative treatment - there are so many other methods and why we do we even have to turn into deep learning? Here, we have some evidence that show deep learning, at the minimum, it detects different signatures from the standard genome-wide association studies. On the left, is a result from our deep learning result. This is actually based on the WHO labeled variances Alpha, Beta, Delta, Omicron, and others. We want to see what signature in the SARS-CoV-2 genome are important to contribute to this increase of virulence. Those high signals you can see at the end of the SARS-CoV-2 genome. And around this region I'm highlighting at about 20,000 to 25,000, that's the spike gene which WHO [has identified] and most of the immune response will be reacted to. Based on the conventional knowledge, that should be in fact, that's how WHO classified the SARS-CoV-2 variants. Surprisingly AI models pick up those signals, but those are not the strongest signal, and they pick up many other signals. Some of the stronger signals are not there. By using conventional statistical genomic methods like Genome Wide Association Study - that still picks the spike gene as the stronger signal. There are some others, but not very strong. So in this case, the AI and the conventional statistical method pick, or at minimum, give a different weight to those signals.

This is surprising and also reassuring, in a way. So we picked the conventional wisdom, the spiking, but we also picked some other signals which may or may not be verified by the experimental method. But how do we predict and how do we verify this, right? So we use AI to predict many things and we know AI can generate a lot of false predictions. In this case how can we verify it? That's quite challenging.

Slide 7

We are trying, in general, two different ways. One is to generate a model experimental system to verify the findings. Our team is using a biological model this is the [inaudible] we also have a cell line system to convert those genes into the life cycle and measure their relative activities.

Slide 8

Then we also have a computational person to perform molecular dynamics to simulate how those mutations affect [inaudible] of their activity. In this case, ACE2 and RBD, but we we also have other [inaudible] in which predict in different regions how they react with the human genes - potential targeting the human immune systems.

Slide 9

We also organize a workshop and expert panels to discuss how to design trustworthy AI to promote social trust in the AI - equitable AI. The COVID pandemic has shown there is high disparity in our current system so if we use AI to predict the future pandemics, it can it can easily amplify the hidden disparities in our current system, probably in our current data as well. If we recognize this and we organize a quite diverse panel, and including people from South Africa, Kenya, there's a person from Europe and across the - with different background like lawyers, physicians, policy makers, governmental - NIST - governmental agencies to have all kinds of perspective on how to promote AI how and to promote trust in AI in historically marginalized communities.

Slide 10

We also emphasize a lot of workforce development given our collaborations with Minority Serving Institutions, including Spelman College, North Carolina A&T, and Catholic University of America, which contain a lot of Hispanic students, and my home institution University of Tennessee, Chattanooga. We also collaborate with a hospital system Global South.

Slide 11

Lastly I want to thank NSF, UTC and the AI Tennessee Initiative for the support of this. And lastly, I didn't put in the slides, NSF has released call for the phase 2 application for a National Center, to be at a scale probably 10 times bigger, if not a 100 times bigger so I'm looking for collaborators, especially with community engagement, public policy, I heard someone working policy. So I wish to connect after this.

So I'll just stop here, thank you.