

**Article:**

Christopher R McCudden.

Deus Ex Machina? Predicting SARS-CoV-2 Infection from Lab Tests Using Machine Learning.

Clin Chem 2020; 66:1365-66 <https://doi.org/10.1093/clinchem/hvaa212>

Guest: Dr. Christopher McCudden from the Ottawa Hospital and the Department of Pathology and Laboratory Medicine at the University of Ottawa in Canada.

Bob Barrett:

This is a podcast from *Clinical Chemistry*, sponsored by the Department of Laboratory Medicine at Boston Children's Hospital. I am Bob Barrett.

The COVID-19 pandemic has rapidly spread worldwide and the highly contagious nature of SARS-CoV-2 rapid progression of disease in some infected patients and the subsequent stress on the healthcare system has created an urgent need for rapid and effective diagnostic strategies for the prompt identification and isolation of infected patients. Currently, the diagnosis relies on virus specific real-time reverse transcriptase polymerase chain reaction, or RT-PCR testing. And while such testing can be completed within 48 hours, it can be substantially longer due to many variables including the need for repeat testing or the lack of needed supplies.

In the November 2020 issue of *Clinical Chemistry*, a group from Cornell Medicine in New York City published a paper titled "Routine Laboratory Blood Test Predict SARS-Cov-2 Infection Using Machine Learning." The authors used 27 routine laboratory tests, each with a turnaround time usually within one to two hours to predict an individual's SARS-CoV-2 infection status. That paper was accompanied by an editorial by Dr. Christopher McCudden, a clinical biochemist at the Ottawa Hospital and an Associate Professor and Vice Chair of the Department of Pathology and Laboratory Medicine at the University of Ottawa in Canada. Dr. McCudden is our guest in this podcast. So doctor, we hear a lot about machine learning and artificial intelligence, aren't they both about the same thing?

Christopher McCudden:

That's a good question. Machine learning is a technique where data is supplied to a software tool to train it to make predictions, and it's called machine learning because the software is told not exactly how to do the analysis so much as provided with training data and parameters. Now, machine learning is actually a subset of artificial intelligence, which is broadly defined as a computer mimicking the performance of a human task. Machine learning really is a subset of what's called "narrow artificial intelligence" which means it's just focused on a particular task as opposed to

general artificial intelligence, which is thinking like a person. Anyway, under the umbrella of machine learning are many subsets which include logistic regression, classification and regression trees, support vector machines, and of course, there's now this highly popular deep learning.

Bob Barrett: Your editorial in *Clinical Chemistry* is in regard to a paper in *Clinical Chemistry* that used machine learning to predict SARS-CoV-2 infection using routine laboratory blood tests. What's important about this study?

Christopher McCudden: Well, the report by Yang and colleagues is important because it addresses an urgent health care problem. Our listeners will certainly need no reminder on how disruptive the COVID-19 global pandemic is to life. So the particular area, lab challenge, addressed by that paper is the speed and availability of COVID-19 diagnostic test results. Test availability continues to be impeded by global supply chain shortages and logistic challenges which often cause long turnaround times and delayed results. So, Yang and colleagues aimed to overcome this problem by predicting SARS-CoV-2 infection before the COVID-19 RT-PCR results are readily available, and they did this by using commonly available laboratory data to train a machine learning algorithm.

Bob Barrett: What machine learning techniques did the group at Cornell that authored that report use?

Christopher McCudden: Well, they actually used four different techniques, specifically logistic regression, classification in decision trees, random forests, and gradient boost of decision trees. Now, I won't get into the details of these in this short podcast, but it is worth highlighting that it is common to use more than one machine learning technique and then selecting the best performing one. In general, all algorithms perform better with more data and what's useful both broadly in terms of machine learning in this particular case is that these algorithms can be retrained with new data as it becomes available, and certainly with COVID there's new data all the time.

So really one of the hallmarks of machine learning algorithm is to constantly improve with new information. Typical example, we have internet search engine which is constantly refined with new data. For the report on SARS-CoV-2 infection, that means that the algorithm can be improved in the event that new test information becomes available, such as a new diagnostic test, or new COVID results, or new gold standards around there.

Bob Barrett: What are the challenges with implementing such an algorithm?

Christopher McCudden: Well, implementing algorithms in typical systems used in laboratories and hospitals is quite difficult. And up until recently, really no software had machine learning capabilities built-in or was particularly well integrated as part of the design characteristics. So, from a pure IT standpoint, there's a challenge just in terms of physically getting the data to the algorithm and then the prediction back to the physician. From there, the challenges actually become more difficult. For example, what exactly should be reported to the clinical team, is it a binary answer such as yes or no, they had the COVID infection, or is it a probability?

Should you supply confidence intervals or limitations around the population that were used to train with the algorithm and really exactly how should a physician act on that result and if it was a probability? On the easy to implement side, the patient could be put in isolation and maybe just wait until those results come back. But even then, you would still need to pick a cut-off. Prediction is really like any other lab tests having inherent tradeoffs between sensitivity and specificity. And if the algorithm is going to be used to decide on initial treatment, it becomes much more problematic in terms of picking that exact cut-off and integrating that information in the decision making.

Bob Barrett: Well, finally doctor, given these challenges, is it possible to use machine learning to predict COVID-19 infection status from routine biochemistry and hematology tests here in the real world?

Christopher McCudden: That's another great question. It's clear that machine learning algorithms can be useful in the real world. With COVID-19, there's many existing workflows and routines that have been disrupted and don't allow for typical patient flow. But there is now virtual visits and telemedicine and spacing out of people for visits that have really shown some explosive growth. So I think machine learning and algorithms do have a place in the real world and perhaps it's best to think of these as sort of a new lab test. It will take some time for medicine to really understand how to use the results, but it could be highly beneficial. With effective strategies to mitigate those risks, I think we need to try new things to get through this pandemic.

Bob Barrett: That was Dr. Christopher McCudden from the Ottawa Hospital and the Department of Pathology and Laboratory Medicine at the University of Ottawa in Canada. His editorial "Deus Ex Machina? Predicting SARS-CoV-2 Infection from Lab Tests Using Machine Learning" and the original research paper on that topic appeared in the November 2020 issue of *Clinical Chemistry*. I'm Bob Barrett, thanks for listening.