Artificial Intelligence and Disruptive Innovation

AI in the Medical Response to COVID-19: A Gap Between the Hype and the Reality

Ira Hofer, MD Maxime Cannesson, MD, PhD

Over the past several years, the lay and scientific press have forecasted that artificial intelligence (AI) will revolutionize medicine. Futurists went so far as to hypothesize that physicians would be replaced by computers in fields such as radiology and pathology, while other fields would be disrupted by novel diagnostics and therapeutics. In 2019, investors poured $4 billion into startups focused on AI in health care, topping the $2.7 billion invested in 2018 (asamonitor.pub/2WTaUPt).

In 2020, COVID-19 upended the world. Along with the myriad challenges, COVID-19 also accelerated many changes that were already under way. In health care, an industry notoriously slow at adopting change, COVID-19 has led to an increase in the adoption of telemedicine, remote monitoring in the ICUs, rapid development and deployment for new testing modalities (such as rapid PCR), and authorization for new types of vaccines less than a year after viral genome sequencing (asamonitor.pub/3n2fRA7). However, despite this rapid adoption of new technologies and the hype surrounding AI in medicine, AI has been mostly absent from the COVID-19 response. The sole exception, public health models of disease spread, has been widely criticized as inaccurate at predicting actual events.

The gap between the promise of AI in medicine and lack of AI as part of the COVID-19 response offers a useful lens with which to understand the current gaps in leveraging AI in health care and the challenges that must be overcome to make the promise into a reality. Specifically, the gap has demonstrated issues in two areas: development of the models themselves and implementation of the models into clinical workflows.

To illustrate these issues, it is useful to use a specific use case – for example, a model predicting which patients will suffer COVID disease requiring a ventilator (severe COVID disease). A model that can successfully predict which patients would suffer severe disease will have allowed better allocation of scare resources during a surge and also facilitate decisions on discharge planning.

Challenges in model development

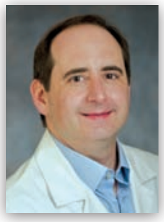
AI models use a set of well-labeled data to create a model to predict something of interest – in this example, severe COVID disease – and then apply this model to data that the model has never seen before. The "magic" of AI is that the algorithms are sophisticated enough to perform well on the unseen sets of data. However, successful models require access to large, well-labeled datasets for training and validation.

Unfortunately, these data are exceedingly hard to come by. A well-labeled data set of severe COVID disease requires detailed information of both patient COVID test results and hospital data. But hospital electronic health records (EHRs) from one hospital are not linked to those from other hospitals or to public databases with COVID test results, creating highly fragmented data. While interoperability standards have been published, their implementation remains fragmented (*IEEE Pulse* 2019;10:25-27). Simultaneously, identifiers for common data, such as lab results or vital signs, differ from one hospital to another, causing data amalgamation to be a herculean task.

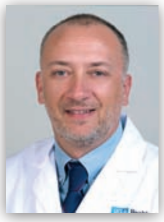
This fragmentation makes the key ingredient of AI models – the data – hard to amass. As of mid-February 2021, the U.S. has had 1.7 million COVID hospitalizations among 27 million cases (6%); enough to power robust models. Each individual hospital, though, has no more than a few thousand hospitalizations. Thus, without interoperability, researchers struggle to amalgamate the number of records to create powerful enough models. Further, the models they do create lack validation outside their own hospitals – it is hard to understand if a model trained on data from one hospital will perform similarly at another hospital due to hospital-specific factors.

> "The gap between the promise of AI in medicine and lack of AI as part of the COVID-19 response offers a useful lens with which to understand the current gaps in leveraging AI in health care and the challenges that must be overcome to make the promise into a reality."

These problems with model creation are even more acute in traditionally underserved areas, such as Black and Brown communities, where patients are more likely to get care from multiple different providers. This exacerbates issues of bias in AI models. For example, doctors might order a coagulation panel in patients they are particularly concerned about – thus the mere presence of these results in a model may increase the predicted risk. However, if the data are missing due to fragmented care, the models may confuse absence of data due to fragmentation with failure to order a test, thereby underestimating the patient's risk and preventing them from getting the right treatment.



Ira Hofer, MD
Associate Professor, University of California, Los Angeles.



Maxime Cannesson, MD, PhD
Professor and Chair, University of California, Los Angeles.

Model implementation

Even when models are successfully developed, there remain significant challenges with implementation into clinical care, both pragmatically and philosophically.

Pragmatically, most current EHRs were not designed to run custom AI models. Unlike an iPhone, implementing standardized ways to call data or interact with the other software (APIs) can be challenging and time-consuming; if they are implemented, they are not necessarily compatible with the APIs from another EHR. Thus, outside developers cannot rapidly create and deploy machine learning models involved in patient care.

Once technically implemented, the models must be integrated into workflows in a way that facilitates patient care. However, many AI models are "black boxes," which means it can be difficult (or even impossible) to understand what is driving the results for a specific patient. If a model says a patient has an 80% chance of needing a ventilator, what should clinicians do with that information? An 80% chance of needing a ventilator is a 20% chance of not needing a ventilator; blindly following the model would lead to the "wrong" choice in one out of five patients. With no way to interpret the model, we cannot abandon clinical judgement.

Then there is the challenge of ethics. As the previous example shows, models reflect the data they are trained with – they cannot say what someone should do. Who should get the ventilator – a 40-year-old with a 50% chance of survival without it, or a 70-year-old with a 5% chance of survival? Does it matter if they have children? Their occupation? What if the model is wrong or biased? These questions are exceedingly complex and as clinicians we should not delegate them to a computer.

In contrast to other areas like telemedicine, or vaccines, implementation and integration of models into clinical care has almost never been done. Thus, it is much harder than scaling an already implemented technology.

Looking to the future

AI is a promising yet immature technology when it comes to medicine. An apt analogy is the internet boom of the late 1990s. The promise of the current internet and smartphones was there, but the ecosystem and infrastructure around it was too underdeveloped. After the hype (the bubble burst), companies were able to do the hard work of laying the foundation for the rapid expansion of today. It is time to lay this foundation for AI in health care.

Maturation takes time. Solutions to each of the challenges noted above will be found, but mistakes will be made along the way. Nonetheless, there are ways to expedite the process. Incentives and regulation can be used to address issues of data ownership and create technical standards for interoperability. Much as the Affordable Care Act created strong incentives for EHR adoption, and HIPPA clarified data privacy rights, updated legislation can require EHRs to adopt open and standardized APIs, providing a stable and clear environment for researchers, entrepreneurs, and industry to develop the models.

Second, funding and incentives can be leveraged to focus researchers on the right issues. Using funding already approved, organizations like the National Institutes of Health and National Science Foundation can issue requests for proposals and be otherwise encouraged to support studies that address issues like model ethics, framing of results, scalable featurization techniques, and model interpretability. By specifically focusing attention on the key friction points to model adoption, these organizations can help accelerate the path to finding solutions to these issues.

Lastly, it is imperative that the ultimate stakeholders – front line doctors and nurses as well as patients – be involved in this process. On the local level, this means creation of diverse committees inside hospitals to evaluate solutions and provide feedback. In industry, this means expanding medical advisory panels and providing ways for patients to give input on these apps. Lastly, in government it means creating advisory groups to address the big-picture issues.

COVID has changed our world. In some cases – such as telemedicine, drug development, and regulation – these changes have shown us the opportunities for some underutilized technologies. In others – like AI – they have helped expose the reality behind the hype. But we often learn more from our failures than our successes. In this case, understanding why AI did not play a major role in the COVID response helps us understand the gaps that must be overcome to make this promising future a reality. ■

Disclosures:
Dr. Hofer is the founder and President of Extrico Health, a company that helps hospitals leverage data from their electronic health record for decision-making purposes. Dr. Hofer receives research support and serves as a consultant for Merck. This work was funded by NIH grant 1K01HL150318.
Dr. Cannesson has an ownership interest in Sironis, a company developing closed-loop systems, and serves as a consultant and receives research support from Edwards Lifesciences (Irvine, California) and Masimo Corp. (Irvine, California). This work is supported by NIH R01 HL144692 and R01 EB029751.