ELSEVIER

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Artificial intelligence and COVID-19: Present state and future vision☆

Anthony C. Chang, MD, MBA, MPH, MS

The Sharon Disney Lund, Medical Intelligence and Innovation Institute (MI3), Children's Hospital of Orange County, USA

ARTICLE INFO

ABSTRACT

The COVID-19 pandemic has lead to catastrophic number of deaths and revealed that much work still remains with data and artificial intelligence. To fully comprehend the dynamics of a pandemic with relevance to artificial intelligence, a primer on global health concepts is first presented. Following this, various aspects of diagnosis and therapy and the relationship to artificial intelligence are presented along with a future projection of an ideal deployment of artificial intelligence in a pandemic. Final thoughts are made about lessons learned and what lies ahead.

Introduction

Camus was prescient as the world has seen the ravages of the novel coronavirus and COVID-19 through most of this year with no immediate relief in sight. The Canadian company BlueDot was purported to be the first organization to disclose the possibility of an outbreak back in December by using machine learning and natural language processing. Artificial intelligence since then, however, has been somewhat under-leveraged as a valuable resource to mitigate the stress and burden of frontline healthcare workers in the tenacious fight against the virus ([1]), and it is important to understand why artificial intelligence has not been more effective. First and foremost, it is essential to have a full under-standing of this virus and the pandemic in the context of global health and epidemiological tenets. This foundation can then be applied to our appreciation for how artificial intelligence has helped to combat this virus currently and how it will be more impactful in the future.

Global health primer with relevance to artificial intelligence

SARS, MERS, and Ebola are familiar recent pandemics that strike fear even amongst seasoned global healthcare workers even though the mortality (774, 38, and 11,325 deaths respectively) was magnitudes less than the number of people who have already succumbed to the COVID-19 pandemic (close to 30 million cases including 1 million deaths worldwide as of early September). As a comparison, the recent pandemic of the novel influenza A virus H1N1 in 2009–2010 (the so-called "swine flu") lead to about 60 million cases with about 151,700 to 575,400 deaths worldwide.

The ongoing coronavirus disease 2019 (COVID-19) is a serious multi system disease (albeit respiratory distress appears to be the most serious) as a result of infection from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). COVID-19 as a disease is manifested by fever, fatigue, cough, chest pain, and shortness of breath with the most lethal pathology being ground-glass lesions in the sub-pleural areas of the lungs with progression to consolidation but many other organs are also affected. The coronavirus SARS-CoV-2, a very large RNA virus, is similar to the SARS-CoV that was responsible for the SARS pandemic. SARS-CoV-2 is covered in a lipid bilayer with protein spikes, which bind to the host cell membrane via the ACE-2 surface receptor for entry and replication. TMPRSS2 is an enzyme that aids the virion to enter the host cell.

There are several important epidemiological terms and concepts in the context of a pandemic, which is large epidemic that has spread to involve an entire country or several countries or regions; there is no set number of countries or regions for this global aspect of a pandemic. The testing of the virus (discussion here is focused on the RT-PCR test for viral material in the nasopharynx) is an essential part of early management of an epidemic or pandemic, remains a continual discussion and contro-versy in the news. A broad testing protocol at an early stage (especially if some individuals are relatively asymptomatic as in this pandemic) coupled with contact tracing and surveillance is essential for discovering the true number of new cases that is the underpinning of a successful containment strategy. This is probably the singular reason for high level of success seen in countries like Singapore, New Zealand, and Taiwan. Lack of broad scale testing significantly impacts on any mechanism of disease prediction in a pandemic, from the traditional epidemiological susceptible-exposed-infectious-recovered-susceptible (SIR or SEIRS) model to the IHME (Institute of Health Metrics and Evaluation at the University of Washington) model that does not take into account the

---

virus characteristics (transmission and disease).

There is much confusion and consternation with the number of new cases: this is often more of a reflection of number of people who had testing that turned out to be positive (in the past 24 h) rather than the true number of new cases (which includes a much larger number of people with infection who are not yet tested). Often the number of new cases is increasing fast but is in actuality due to more people getting access to the testing (as in the case in the US). Therefore, the total cases is the cumulative number of cases to date (including those who have recovered from the infection so this is prevalence as opposed to incidence of new cases). In short, if testing is not widely available, the number of true new cases and total cases are usually much higher then the reported number of new and total cases (due to number of people who are infected and not yet tested and diagnosed).

Another key concept to appreciate is the incubation period (or delay in diagnosis) that is about 2–14 days between time of actual SARS-CoV-2 infection to time of symptoms (which then can lead to a positive test). Even though the lockdown in Wuhan had an immediate impact (in retrospect back tracking the status of all the true new cases), the number of new cases in the news at that time did not reflect this downward trend until 12 days later. In addition, the contagiousness of an infectious agent can be estimated by R_0, which is the estimated number of people that any infected person can transmit the infectious agent. While the typical influenza has an R_0 of about 1 and measles has an R_0 of about 16 (the highest of any infectious disease), R_0 for COVID-19 is estimated to be about 2.0–3.0 (so more contagious that the average flu but less contagious than SARS, MERS, or Ebola, all with R_0 of 4 or greater). Contagiousness, however, needs to be in the context of time of manifestation of the disease, as a major challenge of COVID-19 has been the lack of obvious symptoms for many especially during early phases of the infection. In addition, contagiousness can be mitigated with measures such as aggressive testing, hand washing and sanitizing, contact tracing, temperature checkpoints, travel restrictions, and bans of gatherings above a certain size; more stringent measures include: closing sports events and bars and restaurants, closing of schools, and home quarantines except for food and urgent services (which can close as well). The ultimate "worst possible" virus, therefore, would have the following characteristics: high case fatality rate like Ebola, high contagiousness (R_0) like measles, and long incubation time with majority of hosts with little or no early symptoms like SARS-CoV-2. The latter renders data collection for new cases of COVID-19 exceedingly difficult if not impossible in countries with no stringent public health interventions.

The case fatality rate (in%) is the number of people dying from the disease (total deaths from disease) divided by the number of people diagnosed with the disease (total cases with the disease); it is not number of people dying from the disease divided by the number of people in the entire population, as that is the mortality rate. Hence, the number of deaths from COVID-19 disease is much more reliable as an index of disease than the case fatality rate as number of people diagnosed with the disease is dependent upon access to testing. The case fatality rate for pandemics range widely between the seasonal flu of about 0.1% (with about 500,000 deaths per annum worldwide) to 2.5% for the Spanish flu of 1918 (that resulted in 50–100 million deaths worldwide), and is most lethal at about 50% for Ebola (hence the negative publicity). The case fatality rate for COVID-19 has ranged from an astonishingly low 0.05% in Singapore to above 10% in countries like Italy, Belgium, and the United Kingdom. This case fatality rate not only depends on the demographics of the population (as it is more lethal for the senior population) but also how capable any region's health system is in accommodating the relatively large and sudden influx of critically-ill patients. In short, the case fatality rate can be high because of: 1) relatively low level of testing (smaller denominator of the case fatality rate so the final number is bigger); 2) relatively high number of deaths from the disease mainly from an overwhelmed health system (larger numerator so the final number is bigger) or 3) both (as in the case of Italy). Again, accurate data collection of new cases is key for calculating case fatality rate. Of note, the number

of deaths per 100,000 population has ranged from 0.03 in Taiwan to the 60s (2000x higher) in countries like the United Kingdom and United States.

The present state of artificial intelligence and COVID-19

Artificial intelligence and its panoply of methodologies were rendered only partly effective during the pandemic ([2]). The following not an exhaustive review but rather a summary of some of the activities in using artificial intelligence in the 1) epidemiology, 2) diagnosis, and 3) therapy during the pandemic, with each section followed by a short personal observation:

1) Epidemiology of COVID-19

Prediction Models. The two types of models are statistical and mechanistic: while the former uses machine learning for short-term forecasting, the latter takes into account future transmission scenarios. Although the IHME model, a statistical model, won early praise for its more accurate predictions, it appears that it is simply very difficult for any model to be the continual preferred oracle for accurately predicting deaths and spread of the disease. This is in part due to the differences in assumptions amongst models: while the IHME model currently utilizes anonymized mobile phone data and social distancing policies to estimate degree of contact (changed since early May), other models assume other elements (such as stay-at-home orders) for various decreases in contact between people. The aforementioned traditional SIR or SEIRD models are mechanistic in nature and are favored by epidemiologists as these models have a feedback mechanism from ongoing transmission information ([3]). Machine learning to help model social networks can add an important dimension to pandemic prediction models by calculating the likelihood of different people interacting, but these contact rates are difficult at best to predict especially given complex geopolitical landscape of the U.S.

In short, the exceedingly complex nature of pandemics as a result of biological and human forces has been understandably challenging due to relatively inaccurate number of confirmed cases, and these prediction models needed to be more dynamic and flexible as well as more real-time with more nonlinear approaches to decrease uncertainty.

2) Diagnosis of COVID-19

Medical Image Interpretation. There are many reports of artificial intelligence used in COVID-19 in radiologic imaging of COVID-19 with the deployment of deep learning and convolutional neural network (CNN) of chest X-rays and CT or MRI lung images ([4]) ([5]). Deep learning architectures included visual geometry group (VGG)-16 or VGG-19, Resnet 50, Xception, DenseNet201, Inception_ResNet_V2 and Inception_V3 with data sources that ranged from Kaggle, GitHub, and various hospitals, especially from cities in China. Impressive variations of CNN for medical imaging included: combined CNN-LSTM network ([6]), faster regions with CNN ([7]), and a hybrid VGG-based neural network and data augmentation and spatial transformer network (STN) with CNN (VDSNet) ([8]). There were also reports of using synthetic data from generative adversarial networks (GANs) ([9]). Despite the large volume of papers that reported very high area under the curve for receiver operating characteristics (AUROC), this capability was perhaps limited in utility for the front line clinicians as the clinical manifestations of COVID-19 is usually not a diagnostic enigma and radiologic findings were often normal in early infection ([10]). Furthermore, there was a paucity of coupling the CNN usage for medical images with short and longer term clinical outcomes to stratify disease and risk with therapy alternatives.

The collaboration amongst scientists and clinicians and the focus of artificial intelligence in the form of deep learning on medical imaging of the chest and lungs of COVID-19 have been laudable, but perhaps the interest was

exuberant in this area and insufficient in other areas such as decision support and risk stratification.

Viral and Antibody Testing. There are mainly three tests: 1) the molecular reverse transcriptase-polymerase chain reaction (RT-PCR) test for detection of viral RNA during active infection; 2) the antigen test for detection of specific proteins on the viral surface (also for active infection); and the COVID-19 antibody serology test for past infection. In these tests, particularly the former two which test for an active infection, an important aspect of testing would be the prior probability of likelihood of infection in accordance to Bayes' Theorem. There is also ongoing strategy using machine learning as a tool to support the concept of test pooling. In test pooling, samples from many people are combined into one for testing; if such a combined sample is negative, then the assumption is that everyone in the cohort is negative. Estimating the risk for each of the cohort can be a task that machine learning take on by using data from a myriad of sources to find the optimal grouping strategy. This combined test pooling-machine learning strategy can render this approach very efficient and cost-effective, not to mention more real-time from a timing perspective ([11]).

One of the most underappreciated aspects of testing for COVID-19 is the significance of pretesting probability of having the disease in the test interpretation of the disease, and this focus can be particularly useful if testing is adopting the test pooling strategy that will be efficient and effective in testing the population.

3) Therapy of COVID-19

Clinical Knowledge. As of early September, there are close to 50,000 papers searchable in PubMed alone and scientific manuscripts are currently doubling about every 20 days. Yet it was relatively late during the pandemic that intensive care clinicians were sharing the important information that some COVID-19 patients fared better with the proning position compared with being mechanically ventilated. The capability of artificial intelligence to gain valuable and usable insight from thousands of publications has been challenging, at least in the early phases of the pandemic; to complicate matters even more, several initial reports of drugs and their effects in prestigious journals were retracted due to problems with data integrity ([12]). This profound schism between information and knowledge was partly reconciled with the inception of COVID-19 Open Research Dataset (CORD-19), a collaboration between the Allen Institute for AI and leading research groups to form a free resource of over 130,000 scholarly articles on COVID-19. This effort has been made even more meaningful with machine learning algorithms with a human curation overlay for certain topics in COVID-19. A better curated effort is the 2019 Novel Coronavirus Research Compendium (NCRC) from Johns Hopkins School of Public Health to select out papers that lack adequate peer review. In addition, there were studies that examined the impact of artificial intelligence in the prediction of coronavirus clinical severity ([13]), but there was simply not enough sharing of data, especially from the thousands of ICUs, on an international scale. This dismal failure to have a robust data-sharing system for large-scale real-time analysis in health care is now costing thousands of lives and mandates change ([14]).

It is more vital than ever before to have open access to all scientific papers that will be curated and organized using the latest natural language processing tools for clinicians to have actionable insights, and to have an international data repository of patient data for all to gain knowledge by agile clinical research.

Drug Discovery. There were reports on the use of artificial intelligence and deep learning in the AlphaFold system from DeepMind that involved quaternary protein folding prediction to allow scientists to find drugs that can neutralize the viral proteins. The new methodology called free modeling uses deep learning to predict protein structure without relying on prior templates ([15]). Of note, some of this work and update are published without a peer-reviewed process to expedite the potential benefit of this work for the scientific community. In addition, artificial

intelligence can be coupled with drug repurposing by analyzing various sources such as Scopus, Google Scholar, PubMed, and IEEE Xplore databases ([16]). One of the companies that is focused on such a strategy is BenevolentAI based in the United Kingdom, which has discovered the drug baricitinib to be a candidate for treatment of COVID-19. Lastly, machine learning was applied to the design and development of a vaccine that is a hybrid containing structural proteins and a non-structural protein using reverse vaccinology tools for an effective complementary immune response ([17]).

The potential for an artificial intelligence-empowered strategy for protein structure prediction in free modeling form as well as drug repurposing and even vaccine design and development using combinations of viral proteins has been promising but will need to deliver what it is capable of accomplishing.

The future AI-Enabled strategy for epidemics

The quote from the venerable Dr. Anthony Fauci: "You don't make the timeline, the virus makes the timeline" should be challenged this coming decade by mankind taking control of the human vs virus eternal struggle. Let's imagine our strategy against a future "COVID-29" in the near future and how artificial intelligence can be a *tour de force* in the future management of pandemics:

A novel coronavirus outbreak is detected in southern France with clinical manifestation of bleeding and seizures with an R_0 of 7.5 and a case fatality of greater than 50%. The AI-enabled MRI scans of the brain revealed an unusual pattern of brain inflammation and convolutional neural network and natural language processing using GPT-7 as well as unsupervised learning (cluster analysis) are used to collect data on these patients using their ICU data. Zero-shot learning with transfer learning are deployed for ICUs around the world as an alert for these cases. In pursuit of an effective anticipation and containment strategy of the novel virus, mandatory daily testing at home (15 s for results) with wirelessly automated data entry is immediately started for all of France and its surrounding countries. Edge computing is used to complement the digital technology for relevant data entry into the data repository.

A real-time epidemiological map is made publicly available with proactive approach for case identification and tracing of these individuals using devices for temperature monitoring (including infrared scans now required in all public areas and transportation hubs) and travel history with internet of things and everything (IoT and IoE). Public health measures are immediately implemented in the surrounding countries in a precise format using deep reinforcement learning and Monto Carlo tree search (MCTS): some areas are in containment with individuals followed via their smart phones while other areas are in surveillance mode so businesses and schools are not disrupted in most surrounding regions. A multi-agent modeling process is available for scientists to follow the virus. Drones with food and medical supplies are dispatched to people who reside in the containment areas, and telehealth visits with telemonitoring are available for anyone who is in need.

Simulations of disease models using emulators (deep emulator network search, or DENSE) and AI are deployed to speed up simulations many times over of this small outbreak. There is now a global health system central data repository for sharing of data for federated learning using agile clinical research methodology that obviates the absolute dependency on randomized controlled trials during this critical time. Using crowd-sourced AI (including high school and college AI student championship teams as well as startups and NIH), and providing genomic sequencing and protein folding with structure predictions, the novel coronavirus and its complex quaternary biomolecular structure is successful delineated within 2 h by this collective swarm intelligence. A list of the top 10 anti-viral agents with highest benefit-risk ratios via in silico trials is collected within 6 h for use in critically-ill ICU patients. There are also repurposed drugs that are immediately approved by the FDA, which had representatives as part of this process. The patients and their pharmacogenomic profiles are delineated for therapy based on precision medicine and AI. In addition, a new design vaccine is made available in

24 h as there was already ongoing work on a universal coronavirus vaccine (following the success of the universal flu vaccine in 2025).

After 2 months of this small outbreak, a total of 47 patients were infected with 2 deaths and AI with training on synthetic data generated from generative models that is widely utilized in the management of these patients from a global database. The workers in the hospitals had access to AI-enabled 3D-printed equipment including masks and gowns and intelligent robots attended the COVID-29 patients while they were infectious on mechanical ventilation with weaning protocols utilizing fuzzy logic. For some hospitals, a digital twin concept is in place to forecast resource needs and determine allocation. A group review of COVID-29 at the international Biomedical Research and Intelligence Center (iBRAIN) and its Global Pandemic Prevention Task Force (collaborative international center formed after COVID-19 that claimed over 25 million lives, with WHO and CDC as well as representatives from 109 countries with a rotating directorship) include a discussion of the last pandemic of the current era, COVID-19, as a case history; no mitigation or suppression measures are necessary as surveillance and immediate containment with good individualized precision therapy obviated the need for such historic and draconian strategies.

Conclusion

Several important takeaways in reflecting about the role of artificial intelligence in global health crises such as the COVID-19 pandemic:

1) This pandemic exposed the myriad of inadequacies of healthcare data as well as the lack of sharing of this data;
2) The dire need for an agile clinical science approach with real-time data science to providing caretakers the information they so badly need to care for those afflicted with the infection;
3) The mismatch between data science approach to epidemiology and the complex nature of pandemics with its high degree of human behavior and biomedical uncertainty.

To eradicate a pandemic, we need a proactive case identification and tracing strategy by serial mass screening coupled with sophisticated real-time data science-driven modeling as well as an innovative AI-centric therapeutic program. Digital technologies have been and will continue to be an essential part of public health response to COVID-19 with focus on population surveillance, case identification, contact tracing, and containment measures ([18]); these sources of data will be valuable for prediction models. This overall philosophy will separate the infected individuals from the rest of the population while preserving both the hospital capacity to care for the sickest and the economy to continue in the midst of a pandemic. We need artificial intelligence to help guide us to execute an intervention that is effective and to devise novel therapies with a much shorter timeline. This AI-inspired strategy-outcome coupling using deep reinforcement learning and fuzzy thinking as well as human swarm intelligence will minimize mortality while concomitantly preserving economy (akin to how ICU doctors titrate blood pressure and cardiac output with varying doses of combinations of inotropic medications).

Viruses are the near perfect complex adaptive system (CAS) as these machine-like automata self-organize, pursue a common goal (finding a live host to replicate), and do this without a central leader. Albert Camus described a pandemic s as "a shrewd, unflagging adversary; a skilled organizer, doing his work thoroughly and well." Future pandemics (including a second wave of COVID-19 later this year and early next year) may very well be even more dangerous adversaries as these become even more contagious and lethal. We can, however, surpass their capabilities with passion, inspiration, and creativity but we humans also have greed, selfishness, and independence.

Just as we work towards synergy between clinical medicine and artificial intelligence, there also needs to be such a union between global health and data science. COVID-19, the biggest pandemic since the 1918 Spanish flu, is the current generations' world war. Going into battle with viruses without a sound public health strategy is like going to battle without armor, and going into war with viruses without artificial intelligence is akin to going to war without weapons; in both cases, the human toll is unacceptable. As Alan Turing so presciently stated: "One must design machines to fight machines". Our viral overlords are definitely more like machines.

Funding source

Declaration of competing interest

I have no conflict of interest.

References

[1] McCall B. COVID-19 and artificial intelligence: protecting healthcare workers and curbing the spread. Lancet Digit Health 2020;2(4):e166–167.
[2] Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for COVID-19 (SARS-CoV-2) pandemic: a review. Chaos, Solit Fractals 2020;139:110059.
[3] Holmdahl I, Buckee C. Wrong but useful- what COVID-19 epidemiological models can and cannot tell us. N Engl J Med 2020;383:303–5.
[4] Albahri OS, Zaidan AA, Albahri AS, et al. Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: taxonomy analysis, challenges, future solutions and methodological aspects. J Infect and Public Health 2020;42(11):204–12. S1876-0341(20030558-X.
[5] Salehi AW, Baglat P, Gupta G. Review on machine and deep learning models for the detection and prediction of Coronavirus. Mater Today: Proceedings 22 June 2020.
[6] Islam MZ, Islam MM, Asraf A. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. Inform Med Unlocked 2020;20:100412.
[7] Shibly KH, Dey SK, Islam MTU, et al. COVID faster R-CNN: a novel framework to diagnosis novel coronavirus disease (COVID-19) in X-ray images. Inform Med Unlocked 2020;20:100405.
[8] Bharati S, Podder P, Hossain Mondal MR. Hybrid deep learning for detecting lung diseases from X-rays images. Inform Med Unlocked 2020;20:100391.
[9] Albahli S. Efficient GAN-based chest radiograph (CXR) augmentation to diagnose coronavirus disease pneumonia. Int J Med Sci 2020;1 7(10):1439–48.
[10] Laghi A. Cautions about radiologic diagnosis of COVID-19 infection driven by artificial intelligence. Lancet Dig Health 2020;2(5):e225.
[11] Augenblick N, Kolstad J, Obermeyer Z, et al. Group testing in a pandemic: the role of frequent testing, correlated risk, and machine learning. NBER working paper series 27457. July 2020.
[12] Mehra MR, Desai Ss, Kuh S, et al. Cardiovascular disease, drug therapy, and mortality in COVID-19. N Engl J Med 2020;382(25):e102.
[13] Jiang XG, Coffee M, Bari A, et al. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Comput Mater Continua (CMC) 2020;63(1):537–51.
[14] Cosgriff CV, Ebner DK, Celi LA. Data sharing in the era of COVID-19. Lancet Dig Health 2020;2(5):e224.
[15] Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. Nature 2020;577:706–20.
[16] Mohanty S, Al Rashid MH, Mridul M, et al. Application of artificial intelligence in COVID-19 drug repurposing. Diabetol Metab Syndrome: Clin Res Rev 2020;14: 1027–31.
[17] Ong E, Wong MU, Huffman A, et al. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. Front Immunol 2020;11:1581.
[18] Budd J, Miller BS, Manning EM, et al. Digital technologies in the public health response to COVID-19. Nat Med 2020;26(8):1183–92.