Machine Learning and COVID-19

Vikas Sood[1], Baljeet Kaur[2*]

Abstract

Human and animal viruses are those formidable pathogens that have the potential to create havoc. We are witness to a grave situation, caused be an animal virus, that acquired the ability to infect humans at an alarming rate. The virus, SARS-CoV-2 has led to an unprecedented loss to the world economy as well as loss of precious human lives. In addition, the virus has led to mental agony for those who survived the onslaught. Scientists from around the world are working hard to develop new drugs and vaccines to tame the virus. In the quest to find a cure against the virus, technology is playing a very crucial role. The machine learning approaches have been instrumental in gaining insights into the COVID-19 pandemic at various stages. In this review, we discuss how machine learning approaches have been used in prevention, diagnosis and drug development against the SARS-CoV-2. We have also discussed how machine learning has identified subtle molecular changes among infected individuals.

Keywords: SARS-CoV-2, COVID-19, Machine Learning, Image Processing

Introduction

Viral infections are one of the major economic health burdens around the world. Currently we are witnessing an unprecedented situation where our world is struggling to fight a pandemic caused by a mighty small organism. This mighty small organism consists of a small stretch of nucleic acid (RNA) and a few proteins and is known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). The disease caused by this virus is known as COVID-19: an acronym for Corona virus Disease of 2019. This disease has made our lives miserable and has led to a complete lockdown of the world, an event never witnessed before. We are forced to change the way we socialize, teach, reach out, and travel. With every passing day, infections around the world are increasing by leaps and bounds. This pandemic has refreshed the Bill Gates TED talk of 2015, where he mentioned that anything killing more than 10 million people in the next decade will be an infectious virus and not a war (https://www.ted.com/talks/bill_gates_the_next_outbreak_we_re_not_ready? language=en). Even as we are struggling with health and economic crisis, it seems that the worst is yet to come. A recent prediction by the WHO, that states that some countries still face an uphill task to combat SARS-CoV-2, has catalyzed the efforts to find a drug against this virus (https://tbsnews.net/coronavirus-chronicle/who-warns-some-nations-still-face-long-hard-battle-covid-19-100660). Scientists from around the world are racing against time to develop vaccines and new antivirals in a hope to tame the pandemic. Several candidate vaccines are in the stage of clinical trial and are expected to hit the market in early 2021 providing us with a hope that just like other pathogens, SARS-CoV-2 will also get controlled.

This pandemic has resulted in a section of the society which we proudly and gratefully refer to as the *Corona Warriors*. Foremost among them are the doctors and the nurses who risk their lives every day to fight this pandemic. Along with them, police, sanitisation workers and delivery agents have played a commendable role during this time of crisis. The

1. Jamia Hamdard, Delhi, India, 110062
2. *Corresponding Author, Hansraj College, University of Delhi, Delhi, India, 110007
vikas1101@gmail.com, baljeetkaur26@hotmail.com

role of yet another community is worth mentioning. This community has aggressively worked from their homes, collaborated with virologists, pathologists, health care workers, administrators, scientists, doctors, radiologists and many more, and charted the patterns of the COVID-19 pandemic. These are the machine learning (ML) warriors. With the access to mountains of COVID-19 related data, this community of developers came forward and developed novel algorithms and machine learning based models to help us gain insights into the patterns of the COVID-19 pandemic.

Since long, machine learning has pervaded the healthcare fabric by extending its help in the early prognosis of many life-threatening conditions, as well as its role in the diagnosis of many diseases. Machine learning has given encouraging results in the field of oncology, radiology, cardiology and cell therapy to name a few. Though the volume of data available is enormous, the relevant information extracted from it needs well trained and tested machine learning algorithms. Since 1960s the process of digitizing the patient data (Baird et al. 1965) came into force which included the storage and retrieval of the patient's data. Based on the health records many artificial intelligence (AI) and ML based systems assist physicians to help make decisions on the prognosis and diagnosis of different diseases. The role of ML techniques for detection of lung cancer (Zhao et al. 2002), skin cancer (Sigurdsson et al. 2004), breast cancer (Fear et al. 2002) and many more has been known for a long time now. The significant role of ML is also realized in the early diagnosis of Parkinson's disease (Pereira et al. 2019), its contribution to detect Alzheimer's disease from circulating non-coding RNAs (Ludwig et al. 2019), and the detection of depression (Swati et al. 2019), to name only a few. Thus, machine learning techniques help the clinicians to handle the patients better,and have made a commendable contribution to enhance the quality of human life. As the world found itself gripped in the relentless clutches of the COVID-19 pandemic, machine learning based approaches were very well suited to understand the various aspects related to the pandemic. In fact, the ML based approaches have been used vastly, right from the prevention of the infection, to the diagnosis, as well as the molecular characterization of the SARS-CoV-2. Some of these areas have been described below.

Machine Learning in COVID-19 Prevention

Prevention is better than Cure. Hence, when the COVID-19was declared a pandemic by the World Health Organization on the 11th March, 2020 (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen), one of the foremost task for every individual was to stay away from the onslaught of this deadly infection. The safest and the most effective way to remain safe from the pathogen is to observe social distancing. Apart from social distancing, contact tracking was one of the approaches that was utilized widely to combat the disease. The *Arogya Setu* application was launched by government of India to sensitize citizens regarding social distancing. This application works on the principle of contact sharing to identify people that might have been in proximity to an infected individual. Although the efficacy of contact sharing applications has yet to be established, a study pointed out that, in order to bring an outbreak under control, around 60% of the population is required to use the application (Fraser et al. 2020). Hence this application could work in cities which have a high smart phone penetration: a prerequisite for contact sharing applications. In fact, the *Arogya Setu* app was successful in predicting around 650 potential COVID-19 clusters which might have been missed out otherwise (https://www.hindustantimes.com/india-news/aarogya-setu-alerted-about-650-clusters/story-1kvGonSkLz77dwH3zMYOQI.html). Similarly, Google used the augmented reality technique in *SODAR* to help people maintain social distancing (https://sodar.withgoogle.com/). Another application worth mentioning is the United Nations Technology Innovation Lab's*1point5,* which builds awareness of the social distancing (https://onepointfive.app).

Machine Learning in COVID-19 Diagnosis

Severe infections of SARS-CoV-2 are frequently associated with pneumonia followed by other complications. Furthermore, the fact that most of the people are asymptomatic or develop disease after a period of time post infection makes things complicated as these people can act as potential spreaders. Hence, identification of disease at the nascent stage becomes extremely important. ML based approaches have been successfully utilized to understand the lung images for possible signs of pneumonia in patients which would otherwise have missed by physicians (https://health.ucsd.edu/news/

releases/Pages/2020-04-07-artificial-intelligence-enables-rapid-covid-19-lung-imaging-analysis.aspx). In this study, authors used more than 2000 chest X-ray images to train their models leading to the identification of pneumonia at a very nascent stage. Similarly, Bayesian Convolutional Neural Network was utilized to predict the cases of COVID-19 from the patient chest X-rays.The method improved the sensitivity from 85.7% to 92.9% leading to better informed decisions (Ghoshal et al. 2020). Similarly, a deep learning method called the VB-NET was utilized on thin section CT images to assess the severity among COVID-19 patients (Shi et al. 2020).

Machine Learning in the COVID-19 Drug Development

With the onset of the COVID-19 pandemic, several groups around the world are trying to identify effective drugs again the virus. Few drugs like Remdesivir and Dexamethasone are being investigatedas potential drugs to cure COVID-19. They have been identified using conventional methods by using the drug and a placebo for clinical trials. Apart from the conventional approaches, ML has been successfully utilized in drug repurposing against SARS-CoV-2. In fact ML based algorithm successfully identified Baricitinib, a drug commonly used in arthritis as the potential candidate against the COVID-19 (https://www.weforum.org/agenda/2020/05/how-ai-and-machine-learning-are-helping-to-fight-covid-19/). Similarly, machine learning approach called Vaxigen-ML was also used to develop a potential vaccine against COVID-19 (Ong et. al. 2020). The approach predicted a combination of structural and non-structural protein to develop as potential vaccine candidate. Though the drugs are now in clinical trials but the speed with which ML identified potential drug againstCOVID-19 from a million of candidates is what make this technique laudable.

Machine Learning in SARS-CoV-2 Characterization

Viral infections inthe host lead to numerous changes at the molecular level. Apart from the host proteome, viral infections lead to rapid changes in the host lipidome and metablome as well. Lipidomic profiling of influenza virus was successfully used to identify novel regulators of inflammation (Tam et. al. 2013). Similar study was performed among the COVID-19 and control patients. However, in this study, all the measurements obtained from the patients were fed to the machine learning model. The model could successfully identify severe COVID-19 patients based on the molecular signatures of serum metabolites with an overall accuracy of 93.5% (Shen et. al. 2020). The approach provides vital clues that can be used to predict whether a patient is prone to more severe COVID-19 manifestation or not, thereby enabling a physician to take timely decision. Another study focused on changes in the upper airway samples of COVID-19 and non-viral acute respiratory illnesses. The researchers successfully utilized machine learning models to differentiate COVID-19 patients based on the host transcriptional response (Mick etal. 2020). Hence, ML based techniques have been successfully utilized to characterize SARS-CoV-2 inducted molecular changes inside the host.

Discussion

COVID-19 has been declared as a pandemic and the disease has resulted in unprecedented loss in terms of mortality and economy. A recent report by United Nations states that the COVID-19 could double the global hunger, further indicates the enormity of the problem. COVID-19 cases around the world are still rising, but the number of deaths due to COVID-19 are decreasing due to better medical interventions. The cure to COVID-19 is urgently required and many groups around the world are using diverse strategies to combat the virus. With the availability of enormous data, machine learning approaches have also been instrumental in gaining insights into the COVID-19 pandemic. The ML techniques have risen to the occasion and helped in containing the alarming situation. The techniques developed in this short period and the insights gained will be instrumental in forming the basis of many future studies. The collaboration between research institutes, technology companies and Pharma industries promise to develop novel, effective and therapeutic approaches based on the vast data that gets generated, thereby expediting the prognosis as well as the diagnosis of the diseases.

References

1. Baird HW, Garfunkel JM. Electronic data processing of medical records. *N Engl J Med* 1965; 272(June (23)):1211–5
2. Fear EC, Elise C., Hagness SC., et al. Confocal microwave imaging for breast cancer detection: Localization of tumors in three dimensions.

IEEE Transactions on biomedical engineering 49.8 2002: 812-822
3. Fraser C., Dörner AL., Ferretti L., et al. Digital contact tracing: comparing the capabilities of centralised and decentralised data architectures to effectively suppress the COVID-19 epidemic whilst maximising freedom of movement and maintaining privacy. https://go.nature.com/2x2czk9 2020.
4. Ghoshal B. and Tucker A. Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection *ArXiv 2020* 2003.10769.
5. Ludwig N., Fehlmann T., Kern F., et al. Machine learning to detect Alzheimer's disease from circulating non-coding RNAs. *Genomics, proteomics & bioinformatics* 2019 17.4: 430-440.
6. Mick E., Kamm J., Oliveira A. et al. Upper airway gene expression differentiates COVID-19 from other acute respiratory illnesses and reveals suppression of innate immune responses by SARS-CoV-2. *Medrxiv* 2020https://doi.org/10.1101/2020.05.18.20105171.
7. Ong, E., Waang H., Wong M,U., et al. Vaxign-ML: Supervised Machine Learning Reverse Vaccinology Model for Improved Prediction of Bacterial Protective Antigens. *Bioinformatics* 20202020 May 1;36(10):3185-3191.
8. Pereira, C. R., Pereira D.R., Weber S.A.T., et al. A survey on computer-assisted Parkinson's disease diagnosis. *Artificial intelligence in medicine* 2019 95: 48-63.
9. Shen B., Yi X., Sun Y., Bi X., et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* 2020 Jul 9;182(1):59-72.e15.
10. Shi F., Xia, L., Shan F., et al. Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification. *arXiv 2020* 2003.09860.
11. Sigurdsson S., Philipsen P.A., Hansen L.K., et al. Detection of skin cancer by classification of Raman spectra. *IEEE transactions on biomedical engineering* 2004 51.10: 1784-1793.
12. Swati R., Kaur B., and Agrawal R.K. Enhanced Depression Detection from Facial Cues Using Univariate Feature Selection Techniques. *International Conference on Pattern Recognition and Machine Intelligence*. Springer, Cham, 2019.
13. Tam V.C., Quehenberger O., Oshansky C.M., et al. Lipidomic Profiling of Influenza Infection Identifies Mediators That Induce and Resolve Inflammation. *Cell* 2013 Jul 3;154(1):213-27.
14. Zhao H., Chung S., Lo B., et al. Enhanced lung cancer detection in temporal subtraction chest radiography using directional edge filtering techniques. Medical Imaging 2002: *Image Processing*. Vol. 4684. International Society for Optics and Photonics, 2002.