Applications of Machine learning for pandemics like COVID-19

Shiqi Wu*

Zhejiang University Department of Pharmacy Hangzhou, China

*Corresponding author: Chinashiqi.wu@zju.edu.cn

Keywords: artificial intelligence, machine learning, COVID-19, pandemic.

Abstract: Since COVID-19 was first discovered in December 2019, it has immediately ravaged the world. With its extremely high infectivity and variability, it has become a pandemic that has caused great damage to human life. In the fight against COVID-19, artificial intelligence has helped on multiple angles. This review mainly discusses the impact of machine learning-based artificial intelligence technology on solving the COVID-19 pandemic from the four perspectives — diagnosis, contact tracing, prediction and discovery of drugs. This paper also discusses the limitations and future development of machine learning in practical applications, and has certain suggestions for medical staff and decision makers.

1. Introduction

Coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 was first discovered in December 2019 and spread rapidly around the world, causing more and more infections and deaths and putting huge pressure on limited medical resources especially in low-income or middle-income countries [1]. 69% of the global population aged 60 and over live-in low-income areas, and their risk of contracting COVID-19 and death is greatly increased [2]. Therefore, further optimization of COVID-19 diagnosis, detection and treatment tools is urgently needed.

Digital methods such as artificial intelligence (AI) can effectively achieve this goal. Today's AI mainly relies no longer on symbolic representation and procedural reasoning mechanisms. Modern AI is built on a new foundation, namely machine learning (ML). ML is a scientific discipline that focuses on how computers learn from data. It is the intersection of statistics (learning relationships from data) and computer science (emphasizing efficient computing algorithms) [3].

ML mainly simulates the human decision-making process through two methods — supervised learning and unsupervised learning. Supervised learning provides a certain amount of data input and corresponding results (ie training set) in advance. Through modeling and fitting, the computer predicts the result of unknown data. There are generally two types of supervised learning. The regression problem (linear or multiple) is to provide a series of data and then predict any desired result; the classification problem (eg, decision trees, neural networks) is to predict which category the object belongs to base on the data [4]. The high accuracy of supervised learning has made it more and more like a well-trained doctor who can achieve his goals. For example, in radiology, the automatic detection of lung nodules on chest X-rays is a kind of supervised learning. On the contrary, the training set of unsupervised learning will not have human-labeled results (no feedback), and one will not provide results or know what the results of the training set are. It is simply analyzed by the computer through unsupervised learning algorithms. This kind of output is not predictable, and it is more challenging to find naturally occurring patterns in the data. For example, for heterogeneous diseases like myocarditis, by counting cells such as T lymphocytes, neutrophils, and macrophages in patients with unknown causes, and observing whether there are recurring patterns, the mechanism and treatment of the disease can be further explored. However, if you treat this as a supervised learning at the beginning, and classify patients according to the risk of death, you may not be able to discover new disease mechanisms [5].

In the battle of COVID-19, AI technology represented by ML has played a vital role. This review mainly discusses the various applications of ML technology to solve the challenges during the outbreak, and further explores the current ML technology shortcomings and future development prospects.

2. Applications of ML in COVID-19 pandemic

ML is used to monitor and control the spread of the COVID-19 pandemic. Through diagnosis and contact tracing, high-risk patients are identified for real-time control. The data of individuals is analyzed to predict the risk of death, and it has the potential to be used in vaccination and drug discovery.

2.1 Screening and diagnosis

The diagnosis of COVID-19 faces many problems such as low diagnosis accuracy, slow diagnosis speed, and few experienced clinical diagnosis doctors. The current routine diagnostic technology for COVID-19 is laboratory testing. However, due to its time-consuming, high cost and high equipment requirements, it cannot effectively improve the detection efficiency under high workload conditions. Computed tomography (CT) is an important supplement to reverse-transcription polymerase chain reaction (RT-PCR). Although it is faster in diagnosis, it requires a high level of competence for radiologists. Therefore, some studies have proposed a rapid diagnosis method based on ML technology.

For COVID-19, features on CT images include bilateral ground-glass opacification and consolidation [6]. In order to distinguish between COVID-19, community acquired pneumonia (CAP) and other non-pneumonia CT images, the fully automatic 3D deep learning model (COVNet) is developed and can obtain 2D and 3D images and extract information through CT. After training, the sensitivity of COVNet can reach 90% and the specificity can reach 96% [7], which proves to be effective in diagnosing COVID-19.

Ardakani et al. [8] used 10 commonly used convolutional neural networks (AlexNet, VGG-16, VGG-19, Squeeze Net, Google Net, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, and Xception) and 1020 CT slices to distinguish COVID-19 infection versus other atypical and viral pneumonia diseases. It is found that ResNet-101 and Xception can improve the accuracy of diagnosing COVID-19 by establishing a highly sensitive model, and the performance is very good, serving as an auxiliary tool for radiologists.

Ozturk et al. [9] proposed two-class (COVID vs. No-Findings) and multi-category (COVID vs. No-Findings vs. Pneumonia) classification models based on chest X-ray images, which are completely automated end-to-end structures without the need of extracting features manually. The researchers found that the correct rate of two-class classification reached 98.08%, and the multi-class classification is 87.02%. The model can also be used to diagnose other chest-related diseases such as tuberculosis and pneumonia.

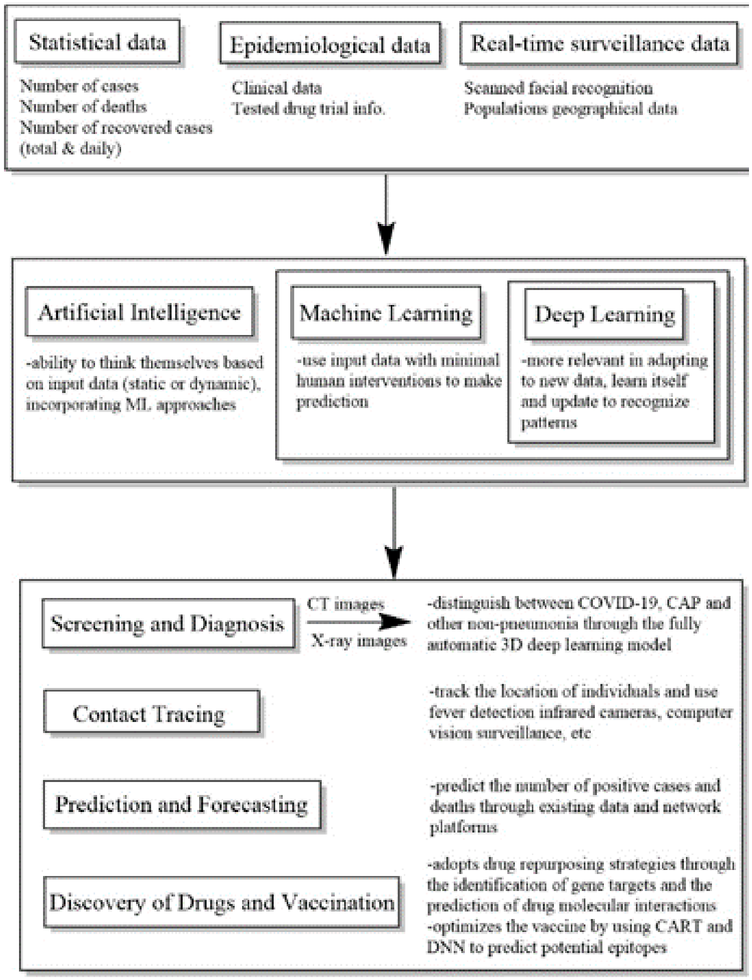

Figure 1. Application of AI for COVID-19

2.2 Contact tracing of the individuals

After being diagnosed with COVID-19, contact tracing of infected individuals is required, which is an important public health measure. Contact tracing identifies individuals who have been in contact with confirmed cases and tracks them to ensure that they can be isolated as soon as possible, so that they can be tested or further treated at the beginning of symptoms. This will help to break the chain of virus transmission and control the expanding progress of the epidemic. Therefore, many countries use different technologies such as Bluetooth, social graph and global positioning system for digital contact tracing.

Smartphones with AI technology can be used to track the location of individuals and detect people who may be potentially affected by using fever detection infrared cameras, computer vision surveillance, and facial recognition systems [10]. One study examined AI-based COVID-19 digital proximity tracking and tracing applications deployed in Qatar and the United Kingdom, and obtained comparative results through diagnostic analysis. Qatar's application EHTERAZ failed to comply with the responsible AI requirement, but provided a lot of help in controlling pandemic. Meanwhile, although the NHS COVID-19 application in the UK has complied with these requirements, its success in combating pandemic has been limited. It is worth noting that this research has made a great contribution to improving the applicability and maturity of AI applications from an ethical point of view [11].

Contact tracing can not only track the epidemic of infectious diseases in the population, but can also be applied to infectious diseases of farmed animals. Rorres et al. [12] recorded the transportation of deer between deer farms in Pennsylvania, the United States for a long time by creating a directed graph, and demonstrated that graph theory can be used to help track contacts of infectious diseases.

However, AI's inability to identify contacts, complex data management, and delays in the steps between identification and isolation are also issues that need to be resolved.

2.3 Prediction and forecasting

ML can track and predict the spread of the virus and the risk of personal infection through existing data and network platforms. It can also predict the number of positive cases and deaths in various regions, especially to help vulnerable countries take corresponding measures.

For example, a study based on Brazil’s cumulative case prediction model is to use the data sets of the country’s 10 high-incidence states (such as Sao Paulo and Rio de Janeiro) through autoregressive integrated moving average (ARIMA), cubist regression (CUBIST), random forest (RF), ridge regression (RIDGE), support vector regression (SVR), and stacking-ensemble learning evaluated time series forecasts. The forecast results show that the model is about one, three and six-days-ahead the COVID-19 cumulative the error ranges of confirmed cases are 0.87%-3.51%, 1.02%-5.63%, and 0.95%-6.90%, which proves that the continuous growth of COVID-19 cases can be predicted and tested, helping the governor and other managers to make better decisions [13].

In addition, ML technology can also predict the health of patients. The fatality rate of critically ill patients will rise sharply with age, which puts tremendous pressure on medical services. A study analyzed blood samples from 485 patients in Wuhan, China, determined 3 indicators (LDH, hs-CRP and lymphocytes) and clinical pathways, and established a model based on the XGBoost machine learning algorithm to identify a prognostic biomarker used to distinguish whether patients need to be treated immediately. The results show that the mortality rate can be predicted at least 10 days in advance with an accuracy rate of over 90%, which maximizes medical resources and is expected to reduce the clinical burden [14].

However, due to lack of data, noisy social media, abnormal data, and algorithm dynamics, AI's prediction of COVID-19 is not completely reliable. Therefore, most of the models used for prediction so far do not use AI technology. Instead, epidemiological models such as SIR models are more commonly used. For example, Maier et al. [15] used the SIR model to capture the isolation of symptomatic infected persons and national isolation measures in response to changes in behavior or policies.

2.4 Discovery of drugs and vaccination

There are two basic strategies for drug development, conventional drug development and drug repurposing [16]. Conventional development is a great burden for scientific research institutions and pharmaceutical companies due to its high time cost and high investment. In recent years, with the continuous development of big data analysis (revealing similar molecular mechanisms among different diseases), computer models (predicting possible drugs for similar pathogenic mechanisms), large-scale screening systems (quick testing of compound functions in different cell lines), drug repurposing has gradually received attention [17]. Therefore, in this regard, the pharmaceutical industry adopts drug repurposing strategies, using AI and ML technology to accelerate and optimize new antiviral drugs. Through the identification of gene targets and the prediction of drug molecular interactions, researchers use Connectivity Map (CMap), Library Integrated Network based Cellular Signatures (LINCS), Genome Wide Association Studies (GWAS), Side Effect Resource (SIDER), and Directionality Map (DMAP) and other tools to greatly improve the chances of drug repurposing [18].

In addition, the ML model optimizes the vaccine by using various techniques such as Classification and Regression Tree (CART) and Deep Neural Network (DNN) to predict potential epitopes. Fast et al [19] recognized the T cell and B cell epitopes of 2019-nCoV based on the antigen presentation and antibody binding characteristics of the viral protein. NetMHCpan4 and MARIA are used to predict the presentation scores of MHC-I and MHC-II respectively, to identify antibodies and develop more effective vaccine. Ong et al [20] used Vaxign and the newly developed machine learning-based Vaxign-ML reverse vaccinology tool to predict new coronary pneumonia vaccine candidates. Through the study of the entire proteome of SARS-CoV-2, it is predicted that 6 proteins including S protein and 5 non-structural proteins are essential for virus adhesion and host invasion.

3. Discussion

3.1 Limitaiton

As mentioned above, ML has been used to solve all aspects of pandemic diseases like COVID-19, but at the same time, its widespread use is accompanied by many limitations and new problems.

1) Data collection

ML usually requires a clean set of annotated data so that the classifier can be well trained. But we don't know what the size of the training sample should be to model an accurate classifier. This is a fatal problem, because limited data will cause the results to deviate from reality, and the analysis of the data will bias. Moreover, it is not easy to collect a large amount of data, and it often takes a long time and a lot of energy [21]. For example, Singapore has been quite successful in data collection by establishing standardized protocols and strict contact tracing across the country [22]. However, even if funds and resources are sufficient, it is impractical to rely on AI in the early stages of a disease outbreak, when predictions or other functions are most needed, because a large amount of information cannot be obtained in the early stages of the outbreak. And even if we can acquire a large amount of data, AI is not foolproof, just like Google Flu Trends (GFT), which was considered a model of big data in 2013. GFT predicts more than twice as many cases of influenza-like illness (ILI) visits as the Centers for Disease Control and Prevention (CDC), so people have discovered the problem of big data hubris and algorithm dynamics [23].

In the process of data collection, quantity is one aspect, and the quality of data is another important point. Informal data sources or news reports will provide heterogeneous data with inherited noise, making the results inaccurate, and the impact of incorrect public health decision-making is worrying and fatal [4].

2) Ethical considerations

A neurosurgery clinic in Toronto once interviewed 30 participants including 18 patients, 7 caregivers and 5 healthcare providers. The results showed that most of the participants supported the use of health data without consent, but required openness and transparency. However, most people believe that the distribution of medical resources should not be done by computers, and almost all participants said that the sale of health data should be prohibited or used only by trusted institutions [24].

These are the ethical concerns faced by the application of AI in health care. For example, digital contact tracing still has limitations in terms of privacy and data security vulnerabilities. Due to the contradictory relationship between AI and privacy issues in the application process, people may have reservations about the use of AI in a pandemic like COVID-19. And this concern about offending data privacy is entirely reasonable.

Like other methods, AI may unconsciously classify ethnic minorities as a group with higher risk of disease when applied. For example, the prevalence of non-Hispanic blacks is about five times that of non-Hispanic whites [25]. Although this may be due to gaps in medical standards, it may make the situation more serious and lead to stigma and discrimination against these groups.

Therefore, in the future, appropriate ethics and legal provisions are needed to regulate the use of AI such as ML. Fortunately, every country is making steps to take privacy, security, transparency, and accountability issues into consideration and work hard to overcome these challenges.

3.2 Prospect

At the moment, the human society's fight against the virus has not yet ended. In the face of this virus's assault on humans, all industries need to cooperate and fight together. The massive data generated in medical and biological research contains very valuable information and knowledge, and AI technology has shown great potential in knowledge extraction and modeling analysis. The widespread use of artificial intelligence and machine learning enables scientists, companies and governments to better manage or even prevent the outbreak of infectious diseases, thereby reducing harm and destruction. Although the entire society is currently focused on COVID-19, AI can pay close attention to other epidemics that may occur around the world, so we will be prepared to respond as needed. Learning from the experience of COVID-19, AI systems like ML can be used for other pandemics in the early detection, internal hospital management and the establishment of patient-specific predictive models based on clinical data. We are constantly moving towards a new era of artificial intelligence in the field of healthcare.

4. Conclusion

Researchers are constantly exploring every possibility of fighting COVID-19 represented by modern technology. This article discusses the application of AI methods based on ML to all aspects of the COVID-19 pandemic, the limitations and future prospects of ML. Due to frequent cross-species infections and accidental spillovers, new viruses similar to SARS-CoV-2 may appear periodically in human society in the future [26]. Therefore, this is also of great significance for the future fight against other epidemic diseases. However, considering the limitations of AI technology in terms of data collection and ethical issues, the safest way for people at present is to let the results of AI guide decision-making, rather than relying on it.

References

[1] Siow, W.T.; Liew, M.F.; Shrestha, B.R.; Muchtar, F.; See, K.C. Managing COVID-19 in resource-limited settings: critical care considerations. Critical care (London, England) 2020, 24, 167, doi:10.1186/s13054-020-02890 - x.

[2] Yang, X.; Yu, Y.; Xu, J.; Shu, H.; Xia, J.; Liu, H.; Wu, Y.; Zhang, L.; Yu, Z.; Fang, M.; et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. The Lancet. Respiratory medicine 2020, 8, 475-481, doi: 10.1016/s2213-2600 (20)30079 - 5.

[3] Abu-Mostafa, Y.S.; Magdon-Ismail, M.; Lin, H.-T. Learning from data; AMLBook New York, NY, USA: 2012; Volume 4.

[4] Chen, J.; See, K.C. Artificial Intelligence for COVID-19: Rapid Review. Journal of medical Internet research 2020, 22, e21476, doi:10.2196/21476.

[5] Deo, R.C. Machine Learning in Medicine. Circulation 2015, 132, 1920-1930, doi:10.1161/circulationaha.115.001593.

[6] Wang, D.; Hu, B.; Hu, C.; Zhu, F.; Liu, X.; Zhang, J.; Wang, B.; Xiang, H.; Cheng, Z.; Xiong, Y.; et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. Jama 2020, 323, 1061-1069, doi:10.1001/jama.2020.1585.

[7] Sun, L.; Song, F.; Shi, N.; Liu, F.; Li, S.; Li, P.; Zhang, W.; Jiang, X.; Zhang, Y.; Sun, L.; et al. Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19. Journal of clinical virology: the official publication of the Pan American Society for Clinical Virology 2020, 128, 104431, doi: 10.1016/j.jcv.2020.104431.

[8] Ardakani, A.A.; Kanafi, A.R.; Acharya, U.R.; Khadem, N.; Mohammadi, A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Computers in biology and medicine 2020, 121, 103795, doi: 10.1016/j.compbiomed.2020.103795.

[9] Ozturk, T.; Talo, M.; Yildirim, E.A.; Baloglu, U.B.; Yildirim, O.; Rajendra Acharya, U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in biology and medicine 2020, 121, 103792, doi: 10.1016/j.compbiomed.2020.103792.

[10] Agbehadji, I.E.; Awuzie, B.O.; Ngowi, A.B.; Millham, R.C. Review of Big Data Analytics, Artificial Intelligence and Nature-Inspired Computing Models towards Accurate Detection of

COVID-19 Pandemic Cases and Contact Tracing. International journal of environmental research and public health 2020, 17, doi:10.3390/ijerph17155330.

[11] El-Haddadeh, R.; Fadlalla, A.; Hindi, N.M. Is There a Place for Responsible Artificial Intelligence in Pandemics A Tale of Two Countries. Information systems frontiers: a journal of research and innovation 2021, 1-17, doi:10.1007/s10796-021-10140-w.

[12] Rorres, C.; Romano, M.; Miller, J.A.; Mossey, J.M.; Grubesic, T.H.; Zellner, D.E.; Smith, G. Contact tracing for the control of infectious disease epidemics: Chronic Wasting Disease in deer farms. Epidemics 2018, 23, 71-75, doi: 10.1016/j.epidem.2017.12.006.

[13] Ribeiro, M.; da Silva, R.G.; Mariani, V.C.; Coelho, L.D.S. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos Solitons Fractals 2020, 135, 109853, doi: 10.1016/j.chaos.2020.109853.

[14] Yan, L.; Zhang, H.-T.; Goncalves, J.; Xiao, Y.; Wang, M.; Guo, Y.; Sun, C.; Tang, X.; Jing, L.; Zhang, M.J.N.m.i. An interpretable mortality prediction model for COVID-19 patients. 2020, 2, 283-288.

[15] Maier, B.F.; Brockmann, D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science (New York, N.Y.) 2020, 368, 742-746, doi: 10.1126/science. abb4557.

[16] Patil, V.M.; Singhal, S.; Masand, N. A systematic review on use of aminoquinolines for the therapeutic management of COVID-19: Efficacy, safety and clinical trials. Life sciences 2020, 254, 117775, doi: 10.1016/j.lfs.2020.117775.

[17] Xue, H.; Li, J.; Xie, H.; Wang, Y. Review of Drug Repositioning Approaches and Resources. International journal of biological sciences 2018, 14, 1232-1244, doi:10.7150/ijbs.24612.

[18] Gns, H.S.; Gr, S.; Murahari, M.; Krishnamurthy, M. An update on Drug Repurposing: Re-written saga of the drug's fate. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie 2019, 110, 700-716, doi: 10.1016/j.biopha.2018.11.127.

[19] Fast, E.; Chen, B.J.B. Potential T-cell and B-cell epitopes of 2019-nCoV. 2020.

[20] Ong, E.; Wong, M.U.; Huffman, A.; He, Y. COVID-19 Coronavirus Vaccine Design Using Reverse Vaccinology and Machine Learning. Frontiers in immunology 2020, 11, 1581, doi:10.3389/fimmu.2020.01581.

[21] Santosh, K.C. AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data. Journal of medical systems 2020, 44, 93, doi: 10.1007/s10916 - 020 - 01562 - 1.

[22] Ng, Y.; Li, Z.; Chua, Y. X.; Chaw, W. L.; Zhao, Z.; Er, B.; Pung, R.; Chiew, C.J.; Lye, D. C.; Heng, D.; et al. Evaluation of the Effectiveness of Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore - January 2-February 29, 2020. MMWR. Morbidity and mortality weekly report 2020, 69, 307 - 311, doi: 10.15585/mmwr.mm6911e1.

[23] Lazer, D.; Kennedy, R.; King, G.; Vespignani, A. Big data. The parable of Google Flu: traps in big data analysis. Science (New York, N.Y.) 2014, 343, 1203-1205, doi:10.1126/science.1248506.

[24] McCradden, M. D.; Baba, A.; Saha, A.; Ahmad, S.; Boparai, K.; Fadaiefard, P.; Cusimano, M.D. Ethical concerns around use of artificial intelligence in health care research from the perspective of patients with meningioma, caregivers and health care providers: a qualitative study. CMAJ open 2020, 8, E90-e95, doi: 10.9778/cmajo.20190151.

[25] COVID-19 in racial and ethnic minority groups. 2020.

[26] Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R. J. N. E. j. o. m. A novel coronavirus from patients with pneumonia in China, 2019. 2020.