A Survey of Using Machine Learning Algorithms During the COVID-19 Pandemic

3 authors, including:

Mostafa Al-Emran
British University in Dubai
118 PUBLICATIONS   **6,630** CITATIONS

SEE PROFILE

Mohammed N. Al-Kabi
Zarqa University
134 PUBLICATIONS   **2,415** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

The International Conference on Information Systems and Intelligent Applications (ICISIA 2022) View project

COVID-19 View project

A Survey of Using Machine Learning Algorithms During the COVID-19 Pandemic

Mostafa Al-Emran , **Mohammed N. Al-Kabi,**
and Gonçalo Marques

Abstract The emergence of novel coronavirus (COVID-19) is considered a worldwide pandemic. In response to this pandemic and following the recent developments in artificial intelligence (AI) techniques, the literature witnessed an abundant amount of machine learning applications on COVID-19. To understand these applications, this study aims to provide an early review of the articles published on the employment of machine learning algorithms in predicting the COVID-19 infections, survival rates of patients, vaccine development, and drug discovery. While machine learning has had a more significant impact on healthcare, the analysis of the current review suggests that the use of machine learning is still in its early stages in fighting the COVID-19. Its practical application is hindered by the unavailability of large amounts of data. Other challenges, constraints, and future directions are also discussed.

Keywords Coronavirus · COVID-19 · Artificial intelligence · Machine learning · Infections · Vaccine development · Drug discovery

M. Al-Emran (✉)
Faculty of Engineering & IT, The British University in Dubai, Dubai, UAE
e-mail: mustafa.n.alemran@gmail.com

M. N. Al-Kabi
Department of Information Technology, Al Buraimi University College, Al Buraimi, Oman
e-mail: mohammed@buc.edu.om

G. Marques
Polytechnic of Coimbra, ESTGOH, Rua General Santos Costa,
3400-124 Oliveira do Hospital, Portugal
e-mail: goncalosantosmarques@gmail.com

1   Introduction

A novel coronavirus (COVID-19) has appeared in Wuhan, China, since December 2019 and has spread to many countries in a short period [1]. In March 2020, the World Health Organization (WHO) has declared the outbreak of COVID-19 as a worldwide pandemic. Therefore, a global response is required to provide healthcare systems to work against this crisis [2]. The delivery of such systems requires the availability of new technologies, such as artificial intelligence (AI), machine learning, and the Internet of Things (IoT), to combat the new epidemic [3]. AI techniques have gradually made a paradigm shift in the healthcare sector [4]. With the recent developments of AI applications, these applications have been employed in fields that are formerly believed to be the fields of only human experts [5]. The emergence of these innovative techniques will be the key in enhancing the identification, prevention, and prediction of COVID-19 cases [6]. In this regard, various expectations have been raised by the scientific community concerning the role of AI techniques in improving the treatment and diagnosis of COVID-19 infections [7].

Healthcare organizations are in an urgent need for decision support systems that could help them in handling the pandemic and suggesting proper solutions to avoid the outbreak [3]. Machine learning, as one of the well-known applications of AI, has been extensively applied on several COVID-19 datasets [8]. Since its emergence in December 2019, the literature witnessed an abundant amount of machine learning applications on COVID-19. This has been perceived through the analysis and prediction of the current and potential future patients [3]. Tracking the confirmed, recovered, and death cases is another application of machine learning. These applications can also be extended to the development of vaccines [3] and drug discovery [9]. To understand these applications, the main contribution of this study is to provide an early review of the articles published on the employment of machine learning algorithms in predicting the COVID-19 infections, survival rates of patients, vaccine development, and drug discovery. The current study also attempts to provide the main limitations and challenges of the existing techniques.

2   COVID-19 Infection Prediction

A vast amount of research studies were conducted for applying the machine learning algorithms to predict the infections of COVID-19. Through the use of a mobile-based web survey, Rao and Vazquez [10] proposed a machine learning algorithm to accelerate the prediction and identification of COVID-19 infected cases. Besides, Maghdid et al. [11] have applied machine learning algorithms for identifying and predicting the preliminary stage of some COVID-19 symptoms based on smartphone sensors (microphone, camera, temperature, and inertial). Further, Metsky et al. [12] developed machine learning algorithms to design nucleic acid detection assays. This approach has detected 67 viral species and subspecies

through which the COVID-19 virus is one of them. In addition, Qi et al. [13] designed a machine learning-based CT radiomics model for predicting the period of residence in the hospital for those who are suffering from pneumonia associated with COVID-19 based on a dataset of 52 patients with laboratory-confirmed cases. The designed model is based on two algorithms, including Random Forest and Linear Regression. Moreover, Yu et al. [14] developed a supervised decision-tree classifier based on several features, including CT images, Anal-swab or Urine specimens, measurements of Throat-swab, and chest radiography. The study aimed to predict the COVID-19 pediatric cases using a dataset of 105 infected children.

Chest computed tomography (CT) scans are considered one of the essential techniques to evaluate the severity of COVID-19 [15]. Therefore, several machine learning-based studies were conducted by applying deep learning techniques on CT scan images to predict the infection of COVID-19. For example, Tang et al. [15] applied the Random Forest model on chest CT images of 176 patients by classifying them into severe/non-severe COVID-19 cases. Besides, Zheng et al. [16] applied a supervised deep learning model on a dataset of 540 patients by analyzing their CT scans to predict the COVID-19 infections. Further, Gozes et al. [17] applied a deep learning approach on a dataset of 157 patients by adopting their CT scans and classifying them into COVID-19 infected cases and non-infected cases. Additionally, Li et al. [18] applied a deep learning approach on the CT scans of 3322 patients for identifying the infected COVID-19 cases. Moreover, Xu et al. [19] employed two CNN three-dimensional classification models using a deep learning approach for screening 618 CT images to measure the probability of COVID-19 infections. In addition, Narin et al. [20] applied three convolutional neural network models on 100 Chest X-ray images to detect the infected patients of COVID-19 pneumonia.

For predicting the development of COVID-19 infections to make better decision-making, several studies were also carried out. In that, Fong et al. [21] used GROOMS methodology in conjunction with the Composite Monte-Carlo simulator (CMC) to improve the deep learning network and fuzzy rule induction to predict the development of COVID-19 infections. In the same vein, Jia et al. [22] applied three mathematical models (i.e., Logistic, Bertalanffy, and Gompertz) to predict the evolution of COVID-19 infected cases. Besides, Qiang et al. [23] employed three encoding algorithms for screening the spike protein features to predict the infection risk and monitor the evolution of COVID-19. In addition, Poole [24] suggested the employment of machine learning for building models that use big data to predict and monitor the spread of COVID-19 and seasonal flu. It has been suggested that there is a correlation between climatological temperatures, latitude, and the spread of COVID-19. Further, Bai et al. [25] used two methods, namely deep learning and multivariate logistic regression to compare between the patients' data at the admission stage and hospital residence and predict the progression of the COVID-19 disease.

3   Survival Prediction of COVID-19 Patients

In addition to employing the machine learning algorithms in identifying and predicting the infection of COVID-19 cases, these algorithms have also been used in predicting the survival rates of COVID-19 patients. For instance, Yan et al. [26] proposed a prognostic prediction model based on the XGBoost machine learning algorithm to determine the crucial predictive biomarkers of disease severity that could be used to predict the survival of COVID-19 infected patients using a dataset of 2799 patients. In the same vein, Yan et al. [27] adopted the XGBoost supervised classifier to predict the survival of severe COVID-19 patients using a larger dataset of 3000 patients. Besides, Yan et al. [28] employed the XGBoost machine learning-based prognostic model to predict the survival rates of COVID-19 cases using a dataset of blood samples of 404 infected patients.

4   Vaccine Development

Several data centers and research labs reported that they are employing AI techniques to look for a vaccine against COVID-19 [29]. In that, some studies reported the use of machine learning in vaccine development. For instance, Ong et al. [30] have reviewed the current status of coronavirus vaccine development and applied the "Vaxign RV" and "Vaxign-ML" techniques to predict the COVID-19 protein candidates for developing the vaccine. In addition, Prachar et al. [31] tested 19 epitope-HLA-binding prediction techniques and used them for the purpose of vaccine development for COVID-19. Despite the efforts made in employing machine learning algorithms for vaccine development, it is suggested that it is not very likely that the vaccine would be available soon [29].

5   Drug Discovery

Several studies were conducted with the potential of using AI techniques in screening the existing drugs and expediting the process of antiviral development to assist in treating the COVID-19. Magar et al. [32] developed a machine learning model to discover the antibodies that potentially inhibit the COVID-19. Besides, Patankar [33] trained the Long Short-Term Memory (LSTM) model to screen 310,000 drug-like compounds from the ZINC database to inhibit the RNA Dependent RNA Polymerase for COVID-19. Further, Tang et al. [34] developed an advanced deep Q-learning network with the fragment-based drug design (ADQN-FBDD) to produce effective drugs against COVID-19.

6 Critical Reflections

It is apparent that while several machine learning studies were carried out to combat the COVID-19, the use of machine learning is still limited in providing more efficient results. This is a surprising outcome as machine learning has had a more significant impact on Medicine, which has led to a paradigm shift in the healthcare sector [35]. Machine learning algorithms could be used to provide attribute prediction (e.g., infection prediction, survival prediction), which although possible, has not been perceived in the reviewed literature. Unlike the other domains, the emergence of COVID-19 and its impact on humanity requires that the patients' data be available to the public. This would allow researchers to analyze these data and generate efficient results that serve the healthcare sector. To this end, a large number of COVID-19 patients' attributes need to be available in order to determine the interrelationships among these features and how this affects the infection or survival rate of patients.

It is argued that AI still did not prove its efficiency concerning COVID-19 [29]. This argument is also supported by the current investigation with regard to machine learning. This stems from several reasons. First, machine learning requires a massive amount of data to train and test the prediction models. Unlike other diseases, there is still inadequate data that can be used to predict and track its outbreak. Second, it has been observed from the existing literature that most of the machine learning studies tend to determine and predict COVID-19 infections with small datasets. The use of small datasets might lead, in some scenarios, into possible biased or unreliable results. Therefore, generalizing the results should be treated with caution. Third, most of the published studies on using machine learning in the COVID-19 pandemic have not been peer-reviewed and tend to use the Chinese datasets [29]. This, in turn, raises some concerns, such as accuracy and reliability.

AI techniques were praised for their potential in contributing to the discovery of new drugs [29]. However, there was a minimal number of studies that contributed to the discovery of drugs and the development of vaccines concerning the use of machine learning in COVID-19. Even the conducted studies did not evaluate the efficiency of the drugs or vaccines on the patients' clinical features. It was hoped that machine learning would provide more interesting patterns. In that, machine learning could be used to determine the relationship between the developed drugs and their impact on patients' treatment using their clinical features. Machine learning could also be used to test the efficiency of the proposed vaccines and their effect on future infections. These issues open up new opportunities for future studies and encourage scholars to uncover new areas of research.

7   Conclusion and Future Insights

The widespread use of machine learning in relevance to COVID-19 is evident through the reviewed studies. However, the effectiveness of machine learning was not perceived concerning the prediction of COVID-19 infections and patients' survival rates. Most of the conducted studies tended to use simple classification algorithms with minimal attempts to take advantage of more recent algorithms. These outcomes open the door for future research and encourage scholars to work on more advanced techniques, such as developing new algorithms or improving the existing ones. In addition, the limited access to patients' data hinders the use of machine learning in an appropriate manner. Hence, the availability of patients' clinical features is another challenge to the application of machine learning algorithms. In order to fight the pandemic, it is essential that authorities take special care of handling the COVID-19 data and communicating them to the public [29].

In summary, AI techniques in general and machine learning, in particular, have the potential to combat the COVID-19 and relevant crisis. However, it can be seen from the rapid review of the existing studies that the use of machine learning in fighting the pandemic is still in its early stages. This outcome is also supported by the conclusions drawn in previous studies [29, 36], which reported that AI techniques are still at their preliminary stages in working against the COVID-19. While the use of machine learning is still limited, the current pandemic may expedite its application for more advanced issues related to the prediction of COVID-19 infections, survival rates of patients, and vaccine and drug developments.

References

1. Zhu, N., et al.: A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. (2020). https://doi.org/10.1056/NEJMoa2001017
2. Belfiore, M.P., et al.: Artificial intelligence to codify lung CT in Covid-19 patients. Radiol. Medica (2020). https://doi.org/10.1007/s11547-020-01195-x
3. Vaishya, R., Javaid, M., Khan, I.H., Haleem, A.: Artificial intelligence (AI) applications for COVID-19 pandemic. Diabetes Metab. Syndr. Clin. Res. Rev. (2020). https://doi.org/10.1016/j.dsx.2020.04.012
4. Albahri, A.S., et al.: Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): a systematic review. J. Med. Syst. **44**(7). (2020). https://doi.org/10.1007/s10916-020-01582-x
5. Yu, K.H., Beam, A.L., Kohane, I.S.: Artificial intelligence in healthcare. Nat. Biomed. Eng. **2**(10), 719–731 (2018). https://doi.org/10.1038/s41551-018-0305-z
6. Long, J.B., Ehrenfeld, J.M.: The role of augmented intelligence (AI) in detecting and preventing the spread of novel coronavirus. J. Med. Syst. (2020). https://doi.org/10.1007/s10916-020-1536-6
7. Neri, E., Miele, V., Coppola, F., Grassi, R.: Use of CT and artificial intelligence in suspected or COVID-19 positive patients: statement of the Italian Society of Medical and Interventional Radiology. Radiol. Medica (2020). https://doi.org/10.1007/s11547-020-01197-9

8. Arpaci, I., et al.: Analysis of twitter data using evolutionary clustering during the COVID-19 pandemic. Comput. Mater. Contin. **65**(1), 193–203 (2020). https://doi.org/10.32604/cmc.2020.011489

9. Ahuja, A.S., Reddy, V.P., Marques, O.: Artificial intelligence and COVID-19: a multidisciplinary approach. Integr. Med. Res. (2020)

10. Rao, A.S.R.S., Vazquez, J.A.: Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when cities/towns are under quarantine. Infect. Control Hosp. Epidemiol. (2020). https://doi.org/10.1017/ice.2020.61

11. Maghdid, H.S., Ghafoor, K.Z., Sadiq, A.S., Curran, K., Rabie, K.: A novel AI-enabled framework to diagnose coronavirus COVID 19 using smartphone embedded sensors: design study (2020)

12. Metsky, H.C., Freije, C.A., Kosoko-Thoroddsen, T.-S.F., Sabeti, P.C., Myhrvold, C.: CRISPR-based COVID-19 surveillance using a genomically-comprehensive machine learning approach. bioRxiv (2020). https://doi.org/10.1101/2020.02.26.967026

13. Qi, X., et al.: Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. medRxiv (2020)

14. Yu, H., et al.: Data-driven discovery of clinical routes for severity detection of COVID-19 pediatric cases. medRxiv (2020)

15. Tang, Z., et al.: Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images (2020)

16. Zheng, C., et al.: Deep learning-based detection for COVID-19 from chest CT using weak label. medRxiv (2020)

17. Gozes, O., et al.: Rapid AI development cycle for the coronavirus (COVID-19) pandemic: initial results for automated detection & patient monitoring using deep learning CT image analysis (2020)

18. Li, L., et al.: Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology, 200905 (2020). https://doi.org/10.1148/radiol.2020200905

19. Xu, X., et al.: Deep learning system to screen coronavirus disease 2019 pneumonia (2020)

20. Narin, A., Kaya, C., Pamuk, Z.: Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks (2020)

21. Fong, S.J., Li, G., Dey, N., Crespo, R.G., Herrera-Viedma, E.: Composite monte carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. Appl. Soft Comput. **93**, 1–14 (2020)

22. Jia, L., Li, K., Jiang, Y., Guo, X., Zhao, T.: Prediction and analysis of coronavirus disease 2019 (2020)

23. Qiang, X.-L., Xu, P., Fang, G., Liu, W.-B., Kou, Z.: Using the spike protein feature to predict infection risk and monitor the evolutionary dynamic of coronavirus. Infect. Dis. Poverty **9**(1), 1–8 (2020). https://doi.org/10.1186/s40249-020-00649-8

24. Poole, L.: Seasonal influences on the spread of SARS-CoV-2 (COVID19), causality, and forecastabililty (3-15-2020). SSRN Electron. J. (2020). https://doi.org/10.2139/ssrn.3554746

25. Bai, X., et al.: Predicting COVID-19 malignant progression with AI techniques. medRxiv (2020)

26. Yan, L., et al.: Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medrxiv.org (2020)

27. Yan, L., et al.: Prediction of survival for severe Covid-19 patients with three clinical features: development of a machine learning-based prognostic model with clinical data in Wuhan. medRxiv (2020). https://doi.org/10.1101/2020.02.27.20028027

28. Yan, L., et al.: A machine learning-based model for survival prediction in patients with severe COVID-19 infection (2020). https://doi.org/10.1101/2020.02.27.20028027

29. Naudé, W.: Artificial intelligence versus COVID-19: limitations, constraints and pitfalls. Ai Soc. 1–5 (2020)
30. Ong, E., Wong, M.U., Huffman, A., He, Y.: COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. bioRxiv (2020)
31. Prachar, M., et al.: COVID-19 vaccine candidates: prediction and validation of 174 SARS-CoV-2 epitopes. bioRxiv (2020). https://doi.org/10.1101/2020.03.20.000794
32. Magar, R., Yadav, P., Farimani, A.B.: Potential neutralizing antibodies discovered for novel corona virus using machine learning (2020)
33. Patankar, S.: Deep learning-based computational drug discovery to inhibit the RNA dependent RNA polymerase: application to SARS-CoV and COVID-19 (2020). doi: https://doi.org/10.31219/osf.io/6kpbg
34. Tang, B., He, F., Liu, D., Fang, M., Wu, Z., Xu, D.: AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. bioRxiv (2020)
35. Rajkomar, A., Dean, J., Kohane, I.: Machine learning in medicine. N. Engl. J. Med. (2019). https://doi.org/10.1056/NEJMra1814259
36. Bullock, J., Pham, K.H., Lam, C.S.N., Luengo-Oroz, M.: Mapping the landscape of artificial intelligence applications against COVID-19 (2020). arXiv preprint: arXiv:2003.11336