

# Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts

Quoc-Viet Pham, Dinh C. Nguyen, Thien Huynh-The, Won-Joo Hwang, and Pubudu N. Pathirana

**Abstract**—The very first infected novel coronavirus case (COVID-19) was found in Hubei, China in Dec. 2019. The COVID-19 pandemic has spread over 215 countries and areas in the world, and has significantly affected every aspect of our daily lives. At the time of writing this article, the numbers of infected cases and deaths still increase significantly and have no sign of a well-controlled situation, e.g., as of 14 April 2020, a cumulative total of 1,853,265 (118,854) infected (dead) COVID-19 cases were reported in the world. Motivated by recent advances and applications of artificial intelligence (AI) and big data in various areas, this paper aims at emphasizing their importance in responding to the COVID-19 outbreak and preventing the severe effects of the COVID-19 pandemic. We firstly present an overview of AI and big data, then identify their applications in fighting against COVID-19, next highlight challenges and issues associated with state-of-the-art solutions, and finally come up with recommendations for the communications to effectively control the COVID-19 situation. It is expected that this paper provides researchers and communities with new insights into the ways AI and big data improve the COVID-19 situation, and drives further studies in stopping the COVID-19 outbreak.

**Index Terms**—COVID-19, coronavirus pandemic, big data, epidemic outbreak, artificial intelligence (AI), deep learning.

## I. INTRODUCTION

Coronavirus disease-19 (COVID-19), caused by a novel coronavirus, has changed the world significantly, not only healthcare system, but also economics, education, transportation, politics, etc. Infected COVID-19 people normally experience respiratory illness and can recover with effective and appropriate treatment methods. What makes COVID-19 much more dangerous and easily spread than other Coronavirus families is that the COVID-19 coronavirus has become highly efficient in human-to-human transmissions. As the writing of this paper, the COVID-19 virus has spread rapidly in 215

Quoc-Viet Pham is with the Research Institute of Computer, Information and Communication, Pusan National University, 2 Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan, 46241 Republic of Korea (e-mail: vietpq@pusan.ac.kr).

Won-Joo Hwang is with the Department of Biomedical Convergence Engineering, Pusan National University, 2 Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan, 46241 Republic of Korea (e-mail: wjh-wang@pusan.ac.kr).

Dinh C. Nguyen and Pubudu N. Pathirana are with the School of Engineering, Deakin University, Waurn Ponds, VIC 3216, Australia (e-mail: cdnguyen@deakin.edu.au and pubudu.pathirana@deakin.edu.au).

Thien Huynh-The is with the ICT Convergence Research Center, Kumoh National Institute of Technology, Gyeongsangbuk-do 39177, Republic of Korea (e-mail: thienht@kumoh.ac.kr).

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MSIT) under Grants NRF-2019R1C1C1006143 and NRF-2019R1I1A3A01060518.

countries, causing infected 1,853,265 people, and 118,854 dead cases, as illustrated in Fig. 1 [1]. The United State of America (USA) is currently the country with the highest COVID-19 cases, where a new day has recently seen around 30,000 confirmed cases and nearly 2,000 deaths. Some other countries like Italy, Iran, Germany, France, and the United Kingdom are also enormously influenced. However, there are no clinical vaccines to prevent the COVID-19 virus and specific drugs/therapeutic protocols to combat this communicable disease.

As the leaders in the war against the novel coronavirus, the World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC) have released a set of public advices and technical guidelines [2], [3]. The cooperation between and efforts from national governments and large corporations are expected to significantly reduce risks from the spread of COVID-19 outbreak. For example, as a search engine giant, Google launched a COVID-19 portal ([www.google.com/covid19](http://www.google.com/covid19)), where we can find useful information, such as coronavirus map, latest statistics, and most common questions on COVID-19. Another example is that IBM, Amazon, Google, and Microsoft with White House developed a supercomputing system for researches relevant to the coronavirus [4]. In response to the pandemic, some publishers now offer free access to the articles, technical standards, and other documents related to the COVID-19-like virus, while web archival services like arXiv, medRxiv, and bioRxiv create a fast link to collected all preprints related to COVID-19 [5]. On the other hand, artificial intelligence (AI) and big data have found in a lot of applications in various fields, e.g., AI in computer science, AI in banking, AI in agriculture, and AI in healthcare. These technologies are expected and may play important roles in the global battle against the COVID-19 pandemic.

To better understand and alleviate the COVID-19 pandemic, many papers and preprints have been published online in the last few months. Our main purpose is to show the effectiveness of AI and big data to fight against the COVID-19 pandemic and review state-of-the-art solutions using these technologies. Moreover, two representative examples of the AI and big data-based solutions are presented for a better understanding. Besides, we highlight challenges and issues associated with existing AI and big data-based approaches, which motivate us to produce a set of recommendations for the research communities, governments, and societies. We note that the papers to be reviewed in this work are collected from various sources

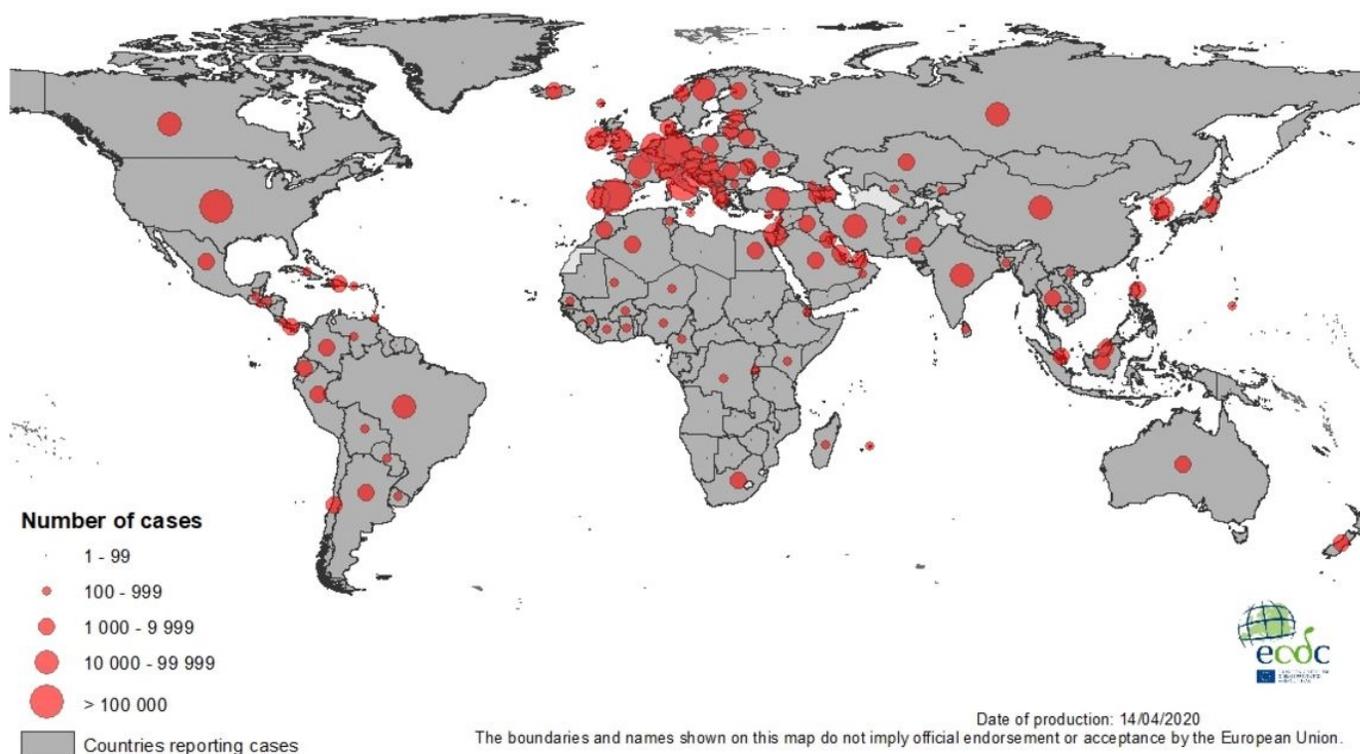


Fig. 1: Illustration of the geographical distribution of COVID-19 cases - worldwide [1]. Data accurate as of April 12, 2020.

like IEEE Xplore and ScienceDirect, as well as preprints from archived websites like arXiv, medRxiv, and bioRxiv.

This work is organized as follows. Section II presents fundamental knowledge of COVID-19, AI, big data, and shows primary motivations behind the use of AI and big data. Next, applications of AI and big data in fighting the COVID-19 pandemic, e.g., identifying the infected patients, tracking the COVID-19 outbreak, developing drug researches, and improving the medical treatment, are reviewed and summarized in Sections III and IV, respectively. In Section V, two examples of AI and big data-based frameworks for combating the COVID-19 disease are described: a phone-based framework for the COVID-19 detection/surveillance, and a data-driven approach using AI for finding antibody sequences. Section VI highlights lessons and recommendations learned from this paper, and discusses the challenges needed to be solved in the future. Finally, Section VII concludes this paper.

## II. BACKGROUND AND MOTIVATIONS

This section presents the fundamentals of COVID-19, AI, big data. This section also shows motivations behind the use of AI and big data in response to the COVID-19 pandemic.

### A. COVID-19 pandemic

COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a betacoronavirus [6]. The first COVID-19 infected case was reported in Wuhan City, Hubei Province of China on 31 December 2019, and has quickly spread out almost all countries in the world, 215 countries and areas at the time of writing this article. There is no sign

that the numbers of infected and dead cases will decrease and the situation will be under control. According to the situation report from the European Centre for Disease Prevention and Control [1], there have been 1,853,265 confirmed cases, and 118,854 dead cases, and the European countries account for around 46.2% and 66.7% percent of the cumulative cases (accurate as of April 14 2020). More seriously, the numbers of new cases are still very large, +83,039 and +6,295 infected and dead cases per day, as report by the CoronaBoard<sup>1</sup>, whereas the fatality rate is 6.34%. The global COVID-19 trend can be found in Fig. 2. Due to the serious situation of COVID-19, WHO has increased the COVID-19 risk assessment to the highest level and declared it as a global pandemic. For a better understanding of the COVID-19 virus about its structure, etiology and pathogenesis, characteristics, and recent treatment progress, we refer the interested readers to refer the paper [7].

As the dramatic impact of COVID-19 on the globe, lots of attempts have been paid for solutions to fight against the COVID-19 outbreak. Government's efforts are mainly responsible to stop the pandemic, e.g., lock down the (partial) area to limit the spread of infection, ensure that the healthcare system is able to cope with the outbreak and provide crisis package to alleviate impacts on the national economics and people, and adopt adaptive policies according to the COVID-19 situation. At the same time, individuals are encouraged to stay healthy and protect others by following some advices like wearing the mask at public locations, washing the hands frequently, maintaining the social distancing policy, and reporting the

<sup>1</sup><https://coronaboard.kr/>

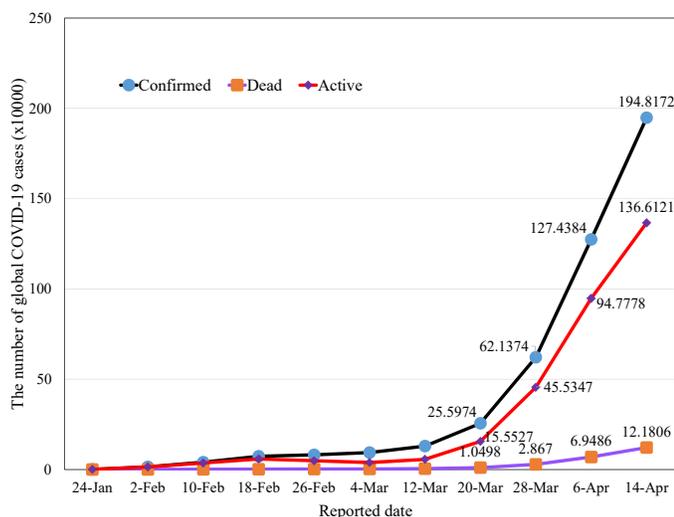


Fig. 2: The global COVID-19 trend (Source: CoronaBoard). Data accurate as of April 14, 2020.

latest symptom information to the regional health center<sup>2</sup>. On the other hand, research and development relevant to COVID-19 are now prioritized, and have been received great efforts from various stakeholders like governments, industries, and academia. Besides the global attempt to develop an effective vaccine and medical treatment for the COVID-19 coronavirus, computer science researchers made initial efforts for the fight against COVID-19. Motivated by the tremendous success of AI and big data in various areas, we present state-of-the-art solutions and approaches based on AI and big data for tackling the COVID-19 coronavirus disease.

### B. Artificial Intelligence

AI is a thriving technology for many intelligent applications in various fields. Some high-profile examples of AI are autonomous vehicles (e.g., self-driving car and drones) in automotive, medical diagnosis and telehealth in healthcare, cybersecurity systems (e.g., malware and botnet detection), AI banking in finance, image processing and natural language processing in computer vision, and modulation classification in wireless communications [8]. Among many branches of AI, machine learning (ML) and deep learning (DL) are two important approaches. Generally, ML refers to the ability to learn and extract meaningful patterns from the data, and the performance of ML-based algorithms and systems are heavily dependent on the representative features [8]. In the meanwhile, DL is able to solve complex systems by learning from simple representations. According to [9], DL has two main features: 1) the ability to learn the right representations shows one feature of DL and 2) DL allows the system to learn the data from a deep manner, multiple layers are used sequentially to learn increasingly meaningful representations.

AI offers a powerful tool to fight against the COVID-19 pandemic. For example, the scientists in [10] developed a DL model to identify existing and commercial drugs for

the “drug-repurposing” (also known as drug repositioning), i.e., finding a rapid drug strategy using existing drugs that can be immediately applied to the infected patients. This study is motivated by the fact that newly developed drugs usually take years to be successfully tested before coming to the market. Although findings in this study are currently not clinically approved, they still open new ways to fight the COVID-19 disease. Insilico Medicine in [11] proposed using the deep generative model for drug discovery (which is defined as the process of identifying new medicines). The COVID-19 protease structures generated from the DL model in this work would be further used for computer modeling and simulations for the purpose of obtaining new molecular entity compounds against the COVID-19 coronavirus. Utilizing DL for computed tomography (CT) image processing, the authors in [12] showed that their proposed DL model when training with 499 CT volumes and testing on 131 CT volumes, can achieve an accuracy of 0.901 with a positive predictive value of 0.840 and a negative predictive value of 0.982. This study offers a fast approach to identify the infected COVID-19 patient, which may provide great helps in timely quarantine and medical treatment. The last example is the use of AI in real-time predicting the infected COVID-19 cases across China [13]. The high accuracy of the AI-based approach proposed in [13] is helpful in monitoring the COVID-19 outbreak and improving the health and policy strategies. Along with the applications mentioned above, the involvement of giant techs is needed because researchers, doctors and scientists can be effectively supported to expedite the research and development of COVID-19 virus. Recently, IBM announced that they are now providing a cloud-based research resource that has been trained on the COVID-19 dataset [14]. Moreover, IBM has adopted their proposed AI technology for drug discovery, from which 3000 novel COVID-19 molecules have been obtained, officially reported in [15]. Another support is from the White House Office of Science and Technology Policy, the U.S. Department of Energy and IBM with the development of COVID-19 HPC Consortium (<https://covid19-hpc-consortium.org/>), which is open for research proposals concerning COVID-19.

### C. Big Data

1) *Definition and Characteristics*: The rapid development of the Internet of Things (IoT) results in a massive explosion of data generated from ubiquitous wearable devices and sensors. The unprecedented increase of data volumes associated with advances of analytic techniques empowered from AI has led to the emergence of a big data era [16]. Big data has been employed in a wide range of industrial application domains, including healthcare where electronic healthcare records (EHRs) are exploited by using intelligent analytics for facilitating medical services. For example, health big data potentially supports patient health analysis, diagnosis assistance, and drug manufacturing [17]. Big data can be generated from a number of sources which may include online social graphs, mobile devices (i.e. smartphones), IoT devices (i.e. sensors), and public data [18] in various formats such as text or video. In the context of COVID-19, big data refers to the patient

<sup>2</sup><https://www.who.int/health-topics/coronavirus>

care data such as physician notes, X-Ray reports, case history, list of doctors and nurses, and information of outbreak areas. In general, big data is the information asset characterized by such a high volume, velocity and variety to acquire specific technology and analytical methods for its transformation into useful information to serve the end users.

- *Volume*: This feature shows the huge amount of data that can range from terabytes to exabytes. According to a Cisco's forecast, the data traffic is expected to reach 930 exabytes by 2020, a seven-fold growth from 2017 [19].
- *Variety*: It refers to the diversity and heterogeneity of big data. For example, big data in healthcare can be produced from healthcare users (i.e. doctors, patients), medical IoT devices, and healthcare organizations. Data can be formatted in text, images, videos with structured or un-structured dataset types [20].
- *Velocity*: It expresses the data generation rate that can be calculated in time or frequency domain. In fact, in industrial applications like healthcare, data generated from devices is always updated in real-time, which is of significant importance for time-sensitive applications such as health monitoring or diagnosis [21].

2) *Big data for COVID-19 fighting*: Big data has been proved its capability to support fighting infectious diseases like COVID-19 [22], [23]. Big data potentially provide a number of promising solutions to help combat COVID-19 epidemic. By combining with AI analytics, big data helps us to understand the COVID-19 in terms of outbreak tracking, virus structure, disease treatment, and vaccine manufacturing [24]. For example, big data associated with intelligent AI-based tools can build complex simulation models using coronavirus data streams for outbreak estimation. This would aid health agencies in monitoring the coronavirus spread and preparing better preventive measurements [25]. Models from big data also supports future prediction of COVID-19 epidemic by its data aggregation ability to leverage large amounts of data for early detection. Moreover, the big data analytics from a variety of real-world sources including infected patients can help implement large-scale COVID-19 investigations to develop comprehensive treatment solutions with high reliability [26], [27]. This would also help healthcare providers to understand the virus development for better response to the various treatment and diagnoses.

### III. APPLICATIONS OF AI FOR FIGHTING COVID-19

This section presents representative applications of AI in fighting the Covid-19 outbreak.

#### A. AI for COVID-19 detection and diagnosis

As one of the most effective solutions to combat the COVID-19 pandemic, early treatment and prediction are of important. Currently, the standard method for classifying respiratory viruses is the reverse transcription polymerase chain reaction (RT-PCR) detection technique. In response to the COVID-19 virus, some efforts have been dedicated to improve this technique [28] and for other alternatives [29]. These techniques are, however, usually costly and time-consuming,

have low true positive rate, and require specific materials, equipment and instruments. Moreover, most countries are suffering from a lack of testing kits due to the limitation on budget and techniques. Thus, the standard method is not suitable to meet the requirements of fast detection and tracking during the COVID-19 pandemic. A simple and low-cost solution for COVID-19 identification is using smart devices together with AI frameworks [30], [31]. This is referred to as *mobile health* or *mHealth* in the literature [32]. These works are advantageous since smart devices are daily used for multi-purposes. Moreover, the emergence of cloud and edge computing can effectively overcome the limitation of battery, storage, and computing capabilities [33], [34].

Another directive for COVID-19 detection is to use AI techniques for medical image processing, which recently appeared in many preprints on coronaviruses [35]–[44]. As we limit this paper to the COVID-19 virus, the interested readers are invited to read the surveys in [45], [46] for other applications of DL in medical image analysis. It is noted from these works that X-ray images and computed tomography (CT) scans are widely used as the input of DL models so as to automatically detect the infected COVID-19 case. Motivated by an important finding that infected COVID-19 patients normally reveal abnormalities in chest radiography images, the authors in [47] tries to design a deep convolutional neural network (CNN) model for the detection of COVID-19 cases. The three-class classification (normal, COVID-19 infected, and non-COVID-19 infected) in this work is helpful if the medical staff needs to decide which cases should be tested with standard methods (between normal and COVID-19 infected cases), and which treatment strategies should be taken (between non- and COVID-19 infected cases). By training on an open source dataset with 16,756 images of 13,645 patients, the proposed CNN model can achieve an accuracy of 92.4%. The use of ML and DL techniques with chest CT scans for COVID-19 detection were considered in [35], [36], [48]–[51], respectively. These works show a high performance as they can achieve a high classification accuracy, e.g., 99.68% in [35], an area under curve (AUC) score of 0.994 in [36], AUC of 0.996 in [48], and accuracy of 82.9% (98.27%) with specificity of 80.5% (97.60%) and sensitivity of 84% (98.93%) in [49] and [50]. As the cost of X-ray scans is usually cheaper than CT images, a large portion of research works for COVID-19 detection utilize DL models with CT images. For example, the work in [40] leverages a deep CNN model, called Decompose, Transfer, and Compose (DeTraC), to process chest X-ray images for the classification of COVID-19. The main purpose of the decomposition layer is to reduce the feature space, thus yielding more sub-classes but improving the training efficiency, whereas the composition layer is to combine sub-classes from the previous layer so as to produce the final classification result. Besides the decomposition and composition layers, a transfer layer is positioned in the middle to speed up the training time, reduce the computation costs, and make the DL model trainable with small datasets. The importance of transfer learning made it a wide consideration in many works, e.g., [38], [41], [42]. To summarize this paragraph, a general representation of DL-based models for the COVID-19 detection and diagnosis can

be showed in Fig. 3, where we present the output as a binary classification: COVID-19 infected and normal.

In the fight against the COVID-19 pandemic, developing efficient diagnostic and treatment methods plays an important role in mitigating the impact of COVID-19 virus. The work in [52] introduces a method based on DL and deep reinforcement learning (DL) to quantify abnormalities in the COVID-19 disease. The input into the proposed learning model is non-contrasted chest CT images, while the output is severity scores, including percentage of opacity (POO), lung severity score (LSS), percentage of high opacity (POHO), and lung high opacity score (LHOS). The proposed learning model, when trained and tested on a dataset of 568 CT images and 100 samples respectively, shows promising results as the Pearson's correlation coefficient between the ground truth and predicted output is 0.97 for POO, 0.98 for POHO, 0.96 for LSS, and 0.96 for LHOS. Another work leveraging DL with CT images is in [53], where two measurement metrics: volume of infection and percentage of infection are quantified. In contrast with other studies, the work in [44] exploits a pre-trained ResNet50V2 model [54] with the Bayesian DL classifier to estimate the diagnostic uncertainty. A random forest (RF) model was investigated in [55] to assess the COVID-19 severity, where 63 features, e.g., infection ratio of the lung and the ground class opacity lesions, are extracted from chest CT images. The result is very promising as it can achieve sensitivity of 0.933, selectivity of 0.745, accuracy of 0.875, and AUC score of 0.91. The key finding from this study is that the severity level is more dependent on the features extracted from the right lung. As an effort to develop a scalable and cost-effective solution for COVID-19 diagnosis, the authors in [56] propose an AI-based framework called AI4COVID-19, which takes into account the deep domain knowledge of medical experts, and uses smart phones to record cough/sound signals as the input data. In particular, the method relies on the fact that COVID-19 disease is likely to have idiosyncrasies from the underlying pathomorphological alterations, e.g., ground-glass opacity (91% for COVID-19 versus 68% for non-COVID-19), vascular thickening (59% for COVID-19 versus 22% for non-COVID-19). Despite interesting results (e.g., average accuracy of 97.74% and response time within 60 seconds), the implementation of AI4COVID-19 at a larger scale is currently limited by several issues, such as quantity and quality of the datasets, and lack of clinical validation.

### B. Identifying, tracking, and predicting the outbreak

A traditional compartmental model to predict the infectious disease is the SIR (Susceptible-Infected-Removed) model. This SIR model can be expressed by the following non-linear differential equations [57]:

$$dS/dt = -\beta SI, \quad dI/dt = \beta SI - \gamma I, \quad dR/dt = -\gamma I, \quad (1)$$

where  $S$ ,  $I$ , and  $R$  denote susceptible, infected, and removed individuals, respectively,  $\beta$  is the transmission rate, and  $\gamma$  is the recovering rate. However, this model is not suitable for the COVID-19 pandemic because of its assumptions, which are 1) the recovered cases will not get infected again and 2) the

model simply ignores the time-varying nature of two parameters  $\beta$  and  $\gamma$ . To overcome issues associated with the traditional SIR model, some efforts have been paid for properly adapting the COVID-19 pandemic. For example, a time-dependent SIR model was proposed in [58], which can adapt to the change of infectious disease control and prevention laws such as city lock downs and traffic halt, as the control parameters  $\beta$  and  $\gamma$  are modeled as time-variant variables. This model can be further improved by having two types of infected patients: detectable and undetectable, which the former has a lower transmission rate than that of the latter. Numerical results in [58] show the proposed time-dependent model can accurately predict the number of confirmed cases in China, and the outbreak time of some countries like Republic of Korea, Italy, and Iran. Several variants of the SIR model other studies, e.g., space-time dependence of the COVID-19 pandemic in [59], SIQR model in [60], where an additional compartment, namely Quarantined, is considered in the traditional SIR model, and stochastic SIR model in [61]. As a summary, these models can predict the outbreak of COVID-19 and send alarm messages to the governments and policy makers so that some actions should be taken in advance prior to the outbreak.

To model the outbreak size of COVID-19, over the last few months, a number of research works have been done using ML and DL models. For example, the authors in [62] argued that the traditional SIR model is not able to capture the effects of more granular interactions such as social distancing and quarantine policies. The authors proposed encoding the quarantine policy as a strength function, which is then included in a neural network for predicting the outbreak size in Wuhan, China. The experimental results using the publicly available data from the Chinese National Health Commission demonstrate that the quarantine policy plays an important role in controlling the COVID-19 outbreak, and the number of infected cases can increase exponentially without the right quarantine policy. Applying the proposed model with quarantine for the United State of America (USA), the infection of COVID-19 may show a stagnation by 20 April if the current COVID-19 policies are executed without any changes. The authors also applied their proposed learning model in [63] to estimate the global outbreak size and obtained similar results. Very recently, some scholars proposed working directly with the empirical data to estimate the outbreak size, for example, logistic model in [64], Bayesian non-linear model in [65], Prophet forecasting procedure in [66], and autoregressive integrated moving average (ARIMA) model in [67]. The study in [68] presents a data-driven approach to learn the phenomenology of COVID-19. The proposed model enables us to read asymptomatic information, e.g., a lag (also known as incubation period) of about 10 days and virulence of 0.14%.

Another interesting work in [69] proposed a modified auto-encoder (MAE) for modeling the transmission dynamics of COVID-19 in the world. Using the data collected from the WHO situation reports, it is shown that the proposed MAE model can predict the outbreak site accurately with an average error less than 2.5%. The experimental results also show that the COVID-19 outbreak can be prevented effectively with fast public health intervention, e.g., compared with one-month in-

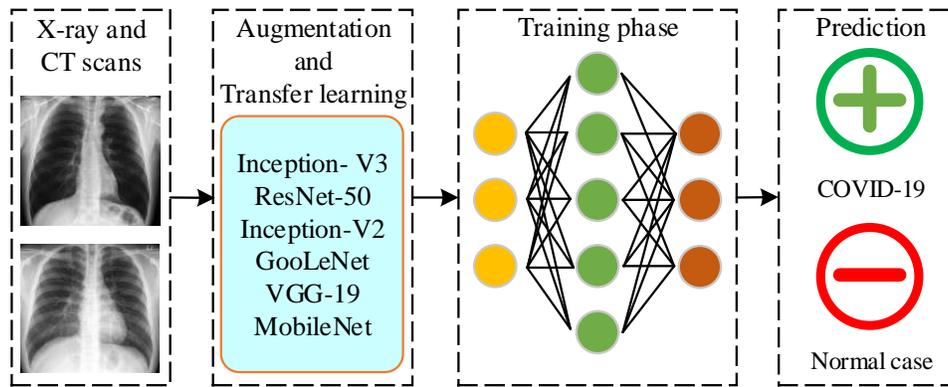


Fig. 3: An illustration of DL-based frameworks for the COVID-19 detection and diagnosis.

tervention, one-week case can reduce the maximum number of cumulative and dead cases by around 166.89 times. Applying the MAE model for the surveillance data of the confirmed Covid-19 cases in China [13], the authors also showed high forecasting capabilities of the proposed MAE model for the transmission dynamics and plateau of COVID-19 in China. The authors in [70] proposed combining medical information (e.g., susceptible and dead cases) and local weather data to predict the risk level of the country. Specifically, a shallow long short-term memory (LSTM) neural network is used to overcome the challenges of small dataset, and the risk level (high, medium, and recovering) of a country is classified by using the fuzzy rule. The experiment results show that the proposed learning model can achieve an average accuracy of 78% over 170 countries.

### C. AI for Infodemiology and Inforeveillance

To date, most reliable information on the COVID-19 pandemic is disseminated through the official websites and social channels of the health organizations like the WHO, and the ministry of health and welfare in each country. However, electronic medium and online platforms (e.g., Facebook, Twitter, YouTube, and Instagram) have showed their importance in distributing information related to COVID-19 like pandemic. Regardless of the quality and source, information from the media platform and the Internet is highly accessible and timely, so further analyses can be performed if the data can be collected and processed properly. As a powerful tool to deal with a vast amount of data, AI have been utilized to have a better understanding of the social network dynamics and to improve the COVID-19 situation. To illustrate the applications of AI during the disease pandemic, the work in [71] presented some real-life examples: 1) using Twitter data to track the public behavior, 2) examining the health-seeking behavior of the Ebola outbreak, and 3) public reaction towards the Chikungunya outbreak.

Similar to the previous pandemics, the recent emergence of COVID-19 has called for several studies from the infodemiology and inforeveillance perspectives. The authors in [72] analyzed the data collected from three popular social platforms in China: Sina Weibo, Baidu search engine, and Ali e-commerce 29 marketplace for assessing public concerns,

risk perception, and tracking public behaviors in response to the COVID-19 outbreak. More specifically, public emotion is evaluated by using a text analysis program, called Linguistic Inquiry and Word Count (LIWC), whereas public attention and awareness, and misinformation are assessed by constructing a Weibo daily index, which is the number of posts with keywords related to the COVID-19 pandemic. Moreover, the Baidu and Ali daily indices are used to evaluate the intention and behaviors to follow the recommended protection measures and/or rumors about ineffective treatments during the COVID-19 outbreak. The results show that quickly classifying rumors and misinformation can greatly alleviate the impact of irrational behavior. Applications computer audition (CA), i.e., speech and sound analysis with AI, to contribute to the COVID-19 pandemic, were considered in [73]. Similar to textual cues in [72], it is possible to collect spoken conversations from news, advertising videos, and social medias for further analyses. Several potential use-cases are also presented, including risk assessment, diagnosis, monitoring the spread and social distancing effects, checking the treatment and recovery, generation of speech and sound. Along the potentials of CA in the fight against COVID-19, there are some challenges needed to be addressed, for example, how to collected COVID-19 patient data, how to process the audio and speech data effectively and in real-time, and how to explain the results obtained from the CA-based solutions.

Another application of AI can be found in [74], where the COVID-19 related data is collected from heterogeneous sources at multiple levels, including official public health organizations (e.g., WHO and county government websites), demographic data, mobility data (e.g., traffic density from Google map), and user generated data from social media platforms. By utilizing the conditional generative adversarial nets to enrich the limited data and a novel heterogeneous graph auto-encoder to estimate the risk in an hierarchical fashion. The proposed  $\alpha$ -Satellite system allows us to be aware of the COVID-19 risk in a specific location, thus enabling the selection and appropriate actions to minimize the impact of COVID-19. Recently, the authors in [75] proposed a novel method, namely Augmented ARGONet, to estimate the number confirmed COVID-19 cases in two days ahead. The data is collected from multiple sources, including official health

reports from China CDC, COVID-19 Internet search activities from Baidu, news media activities from Media Cloud, and daily forecasts achieved by the model proposed in [76]. To enrich the dataset, each data point is also augmented by adding a random Gaussian noise with mean 0 and standard deviation 1. The Lasso regression model is then applied to predict the number of confirmed cases for 32 provinces in China. The results show that the proposed Augmented ARGONet is able to outperform the baseline models for most testing scenarios.

#### D. AI for biomedicine and pharmacotherapy

The world has seen a race for finding effective vaccines and medical treatments in order to combat the COVID-19 virus, which demands for substantial efforts from not only health science but also computer science, with the help of AI and emerging technologies. The authors in [77] have answered the question “Why AI may benefit biomedical research”. First, the vast amount of biomedical data has triggered the use of AI in various areas of biomedicine and pharmaceutical industry. DL with deep neural networks is able to handle high-dimensional and non-structured data with nonlinear relationships, which are in line with biomedical data like transcriptomics and proteomics. Lastly, high capabilities in extracting higher-level features makes DL a potential candidate to analyze, bind, and interpret heterogeneous medical data such as DNA microarray data. The past few years have seen a wide range of AI applications to biomedicine researches, for example, medical image classification, genomic sequence analysis, classification and prediction, biomarkers, structural biology and chemistry, multiplatform data processing, drug discovery and repurposing [78], [79]. Due to the severity of COVID-19 pandemic, AI has recently found medicine applications to curb the spread of the coronavirus, which mainly focus on protein structure prediction, drug discovery, and drug repositioning.

Using AI to discover new medicines and/or existing drugs to be immediately used to treat the COVID-19 virus is beneficial from both the economic and scientific perspectives, especially when clinically verified medicines and vaccines are not available. The work in [80] leverages a COVID-19 virus-specific dataset to train a DL model. Then, this trained model is used to check 4895 commercially available drugs and find potential inhibitors with high affinities. Overall, ten existing drugs are listed as potential inhibitors, for example, the HIV inhibitor *abacavir*, and respiratory stimulants *almitrine mesylate* and *roflumilast*. The work in [81] proposed a data-driven model for drug repurposing through the combination of ML and statistical analysis methods. Initially, the authors selected a list of 6225 candidate drugs, which are then narrowed through consecutive phases, including a network-based knowledge mining algorithm, a DL based relation extraction method, and a connectivity map analysis approach. After the *in silico* and *in vitro* studies, a novel inhibitor, namely CVL218, has been found to be a drug candidate to treat the COVID-19 virus. The potential of this finding is that the inhibitor CVL218 is verified to have a safety profile in monkeys and rats. Another interesting work for drug repurposing is in [82], which leverages a Siamese neural network (SNN) to identify the COVID-19

protein structure versus Ebola and HIV-1 viruses. Advantages this work is that the proposed DL model can be trained without the need for a large dataset and can work directly with available biological dataset instead of public datasets, which are not specific about the COVID-19 virus. Besides drug repurposing, many have been devoted to drug discovery [83]–[86] and protein structure prediction [87]–[89] for combating against the COVID-19 pandemic. For example, the work in [84] considered a DL model named controlled generation of molecules (CogMol) to learn candidate molecules that can bind protein targets of the COVID-19 virus, which are then used to generate candidate drugs for the treatment of the COVID-19 virus. Moreover, a multitask deep neural network is used to predict the toxicity of the generated molecules, thus improving the *in silico* screening process and increasing the success rate of the drug candidates.

#### E. Summary

In this section, we review the applications of AI for COVID-19 detection and diagnosis, tracking and identification of the outbreak, infodemiology and infoveillance, biomedicine and pharmacotherapy. We observe that the AI-based framework is highly suitable for mitigating the impact of the COVID-19 pandemic as a vast amount of COVID-19 data is becoming available thanks to various technologies and efforts. Through the AI studies are not implemented at a large scale and/or tested clinically, they are still helpful as they can provide fast response and hidden meaningful information to medical staffs and policy makers. However, we face many challenges in designing AI algorithms as the quality and quantity of COVID-19 datasets should be further improved, which call for constant effort from the research communities and the help from official organizations with more reliable and high-quality data.

### IV. APPLICATIONS OF BIG DATA FOR FIGHTING COVID-19

This section discusses the potential of big data in fighting the COVID-19 pandemic via four main application domains: outbreak prediction, virus spread tracking, coronavirus diagnosis/treatment, and vaccine/drug discovery, as shown in Fig. 4.

#### A. Outbreak prediction

Big data plays an important role in combating COVID-19 through its capability to predict the outbreak from large-scale data analytics. For example, the work in [90] leveraged real datasets collected from the COVID-19 pandemic in Italy to estimate the outbreak possibility which is of significant importance to plan effective disease control strategies. Instead of using a simple and deterministic model based on human transmission [91], the authors developed more complex models that can accurately formulate the dynamics of the pandemic based on huge data sets from Italian Civil Protection sources. Another data source for outbreak prediction is the public dataset that can be used to visualize the geographical areas with possible outbreak [92]. A first trial is an investigation in Wuhan, aiming to monitor the people migration from and

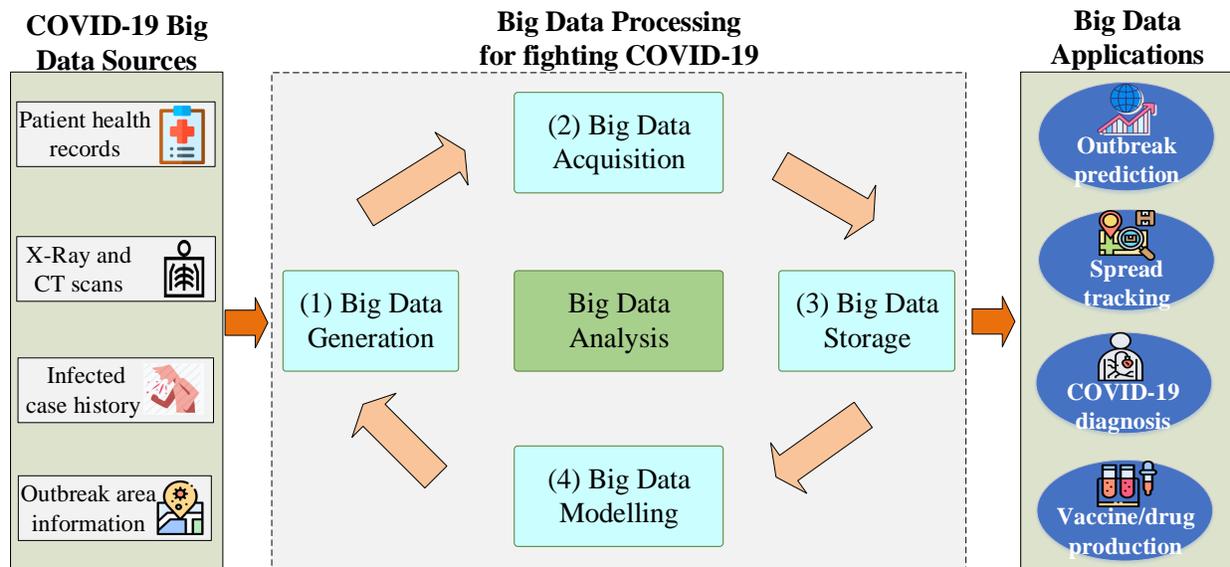


Fig. 4: Big data and its applications for fighting COVID-19 pandemic.

to Wuhan city so that the health agencies can predict the population infected with COVID-19 for quarantine.

Meanwhile, the work in [93] employed the COVID-19 outbreak-related data from authoritative sources such as the National, provincial and municipal Health Commissions of China (<http://www.nhc.gov.cn/>). This big data source can help implement pandemic modeling to interpret the cumulative numbers of infected people, recovered cases in five different regions, i.e., the Mainland, Hubei, Wuhan, Beijing, and Shanghai. More importantly, it allows us to perform simulations to predict the tendency of the COVID-19 outbreak, i.e. identifying the areas at high risks of pandemic and detecting the population with increasing infected cases, all of which contribute to the success of anti-pandemic campaigns.

Big data also enables outbreak prediction on the global scope. From the data analysis point of view, the outbreak is predicted using available data points that put the accuracy of the fitting for reliable estimation under doubt due to the lack of comprehensive investigations. Accuracy may depend on the number of contributed factors, ranging from infected cases, population, living conditions, environments, etc. Motivated by this, a research effort in [94] leveraged a large dataset from various regions and countries such as Korea, China, to estimate the pandemic based on a logistic model that can adjudge the reliability of the predictions. Another trial in [95] used Google Trends as a data engineering tool to collect coronavirus-related information in China, South Korea, Italy, and Iran. The data comes from a textual search with five geographical settings: 1) worldwide, to investigate the global interest in coronaviruses; 2) China, where there is currently the highest number of infected cases; 3) South Korea, where interest has increased since 19 February because hundreds of new infected cases were confirmed; 4) Italy; and 5) Iran, where since 22 February hundreds of new cases have been detected. This data source aggregation would help visualize the outbreak tendency and estimate the possible outbreak. The government reports

collected from the coronavirus pandemics in China, South Korea, Italy, and Iran are also collected and analyzed by [96] with a data optimization model, aiming to generate an accurate prediction of the daily infected cases, which potentially long-term predictions on the future outbreak.

Moreover, the work in [97] employed datasets from John Hopkins University repository, which are obtained from confirmed cases, death cases and recovered cases of all countries, to build prediction models using data learning. The trial is implemented in India, indicating that the proposed data analytic can estimate the outbreak in short-term intervals, i.e. two-week duration, which potentially expands to build larger models for long-term predictions. Meanwhile, another data analytic method in [98] was investigated in the US with the large-scale datasets collected from American cities. The key purpose of the study is learning from recent data to calculate prediction errors to optimize the data modeling model for improving the quality of future estimation, possibly for coronavirus-like pandemics.

### B. Virus spread tracking

Another role of big data is the tracking of COVID-19 spread, which is of paramount importance for healthcare organizations and governments in controlling successfully the coronavirus pandemic [99]. A number of newly emerging solutions using big data have been proposed to support tracking of the COVID-19 spread. For instance, the study in [100] suggested a big data-based methodology for tracking the COVID-19 spread. A large dataset is collected from China National Health Commission with 854,424 air passengers left 55 Wuhan Airport to 49 cities in China during the period of December 2019-January 2020. A multiple linear model is built using local population and air passengers as estimated variables that are useful to quantify the variance of reported cases in China cities. More specifically, the authors employed a Spearman correlation analysis for the daily user traffic

from Wuhan and the total user traffic in this period with the number of 49 confirmed cases. The analytic results show a high correlation between the positive infection cases and the population size. The authors in [101] considered leveraging both big data for spatial analysis methods and Geographic Information Systems (GIS) technology which would facilitate data acquisition and the integration of heterogeneous data from health data resources such as governments, patients, clinical labs and the public.

Another research effort in [102] combined datasets collected from China, Singapore, South Korea, and Italy to build a comprehensive analytic model for virus spread tracking. Based on data learning and modelling, a macroscopic growth law is derived that allows us to estimate the maximum number of infected patients in a certain area. This is important for effective assessment of COVID-19 prevention and monitoring the potential spread of COVID-19 disease, especially the nearby regions of the central epidemic. Different from [102], the study in [103] proposed a temperature-based model that evaluates the relationship between the number of infected cases and the average temperature in different countries necessary for coronavirus tracking. A large dataset is collected from 42 countries which developed the epidemic earlier and built a dataset of 88 countries. The analysis results from COVID-19 tracking show that in northern hemisphere countries, the growth rate should significantly decrease due to the warmer weather and lockdown policies, compared to the southern hemisphere countries.

The previous works have focused on big data solutions, where ground truth data, in the form of historical syndromic surveillance reports, can be utilized to build data models for disease tracking. However, the accuracy and reliability of the modelling models are under questions. The authors in [104] suggested an unsupervised model for COVID-19 spread tracking from online data. They selected a wide range of related symptoms as identified and collected by a survey from the National Health Service (NHS) in the United Kingdom (UK) by incorporating a basic news media coverage metric associated with confirmed COVID-19 cases. Then, a transfer learning method is proposed for mapping supervised COVID-19 models from a country to another country where the COVID-19 epidemic possibly spreads.

Recently, China relies on big data technologies to analyze the people's movement via their mobile phones and mobile apps [105]. The data from the wireless operators is used to build the spreading trajectory of COVID-19 which is highly essential to find out who have contacted with the infected person for quarantine measures. For instance, recently the government agency in Shanghai city suggests big data as a solution for controlling the COVID-19 spread. A big data platform has been designed to gather information from human temperature, travel data and quarantine areas for analytics. Another example is that the China National Health Commission (NHC) [106] has allowed local governments for using big data-based tools such as modelling, estimation combined with AI, to track the epidemic in a real-time fashion.

### C. Coronavirus diagnosis/treatment

In addition to outbreak prediction and spread tracking applications, big data has the potential to support COVID-19 diagnosis and treatment processes. In fact, the potential of big data for diagnosing infectious diseases like COVID-19 has been proved via recent successes, from early diagnosis [107], prediction of treatment outcomes, and building supportive tools for surgery. Regarding COVID-19 diagnosis and treatment, big data has provided various solutions as reported in the literature. A study in [108] introduced a robust, sensitive, specific and highly quantitative solution based on multiplex polymerase chain reactions that are able to diagnose the SARS-CoV-2. The model consists of comprised of 172 pairs of specific primers associated with the genome of SARS-CoV-2 that can be collected from the Chinese National Center for Biological information (<https://bigd.big.ac.cn/ncov>). The proposed Multiplex PCR scheme has been shown to be an efficient and low-cost method to diagnose *Plasmodium falciparum* infections, with high coverage (median 99%) and specificity (99.8%). Another research in [109] implemented a molecular diagnostic method for genomic analyses of SARS-CoV-2 strains, with a focus on studying on Australian returned travelers with COVID-19 disease using genome data available at <https://www.gisaid.org/>. This study may provide important insights into viral diversity and support COVID-19 diagnosis in areas with lacking genomic data.

Meanwhile, the authors in [110] used the proteomics cells that get infected with COVID-19 virus for diagnosis. More specifically, data includes 6381 proteins in human cells with the COVID-19 infection. Based on that, a cooperative analysis of an impact pathway analysis and a network analysis has been derived to analyze the data gathered from the Kyoto Genes storage. Another solution in [111] leveraged publicly available mouse and nonhuman primate for non-neural expression of SARS-CoV-2 entry genes. As a trial, a bulk and single-cell RNA-Seq dataset is tested to detect all types of cells that got infected with COVID-19 virus, which is necessary for diagnosis and/or prognosis in COVID-19.

Due to the limited dataset for COVID-19 diagnosis and treatment, more experimental data has been generated in [112] where all the 6 epitopes (A, B, C, E, F/G and H) can induce the body to produce corresponding antibodies and generate specific humoral immunity. The sequence of S, E, M protein and its proximal sequences have been employed, which produced 420, 334 and 329 sequences in total from NCBI database to build the final dataset after genome annotation. Based on that, the authors focus on a diagnosis procedure by the prediction of the 3D structure of target protein and prediction of conformational B cell epitopes of target protein of SARS-CoV-2 coupled with an analysis of epitope conservation.

Interestingly, the recent work in [113] presented a comprehensive guideline with useful tools to serve the diagnosis and treatment of COVID-19. This guideline consists of the guideline methodology, epidemiological characteristics, population prevention, diagnosis, treatment of COVID-19 disease. As a first trial, a large-scale data of Zhongnan Hospital of Wuhan University has been analyzed from a collection system

where 11,500 persons were screened, and 276 were identified as suspected infectious victims, and 170 were diagnosed. An array of clinical tests have been implemented from the big dataset, from Typical and Atypical CT/X-ray imaging manifestation to hematology examination and detection of pathogens in the respiratory tract.

#### D. Vaccine/drug discovery

Developing a novel vaccine is very crucial to defending the rapid endless of global burden of the COVID-19 pandemic. Big data can gain insights for vaccine/drug discovery against the COVID-19 pandemic. Few attempts have been made to develop a suitable vaccine for COVID-19 using big data within this short period of time. The work in [114] used the GISAID database ([www.gisaid.org/CoV2020/](http://www.gisaid.org/CoV2020/)) that has been used to extract the amino acid residues. The research is expected to seek potent targets for developing future vaccines against the COVID-19 pandemic. Another effort for vaccine development is in [115] that focused on investigating the spike proteins of SARS CoV, MERS CoV and SARS-CoV-2 and four other earlier out-breaking human coronavirus strains. This analysis would enable critical screening of the spike sequence and structure from SARS CoV-2 that may help to develop a suitable vaccine.

The approaches of reverse vaccinology and immune informatics can be useful to develop subunit vaccines towards COVID-19. [116]. The strain of the SARS-CoV-2 has been selected by reviewing numerous entries of the online database of the National Center for Biotechnology Information, while the online epitope prediction server Immune Epitope Database was utilized for T-cell and B-cell epitope prediction. An array of steps may be needed to analyze from the built dataset, including antigenicity, allergenicity and physicochemical property analysis, aiming to identify the possible vaccine constructs. A huge dataset has been also collected from the National Center of Biotechnology Information for facilitating vaccine production [117]. Different peptides were proposed for developing a new vaccine against COVID-19 via two steps. First, the whole genome of COVID-19 was analyzed by a comparative genomic approach to identify the potential antigenic target. Then, an Artemis Comparative Tool has been used to analyze human coronavirus reference sequence that allows us to encode the four major structural proteins in coronavirus including envelope (E) protein, nucleocapsid (N) protein, membrane (M) protein, and spike (S).

Big data also helps provide strategies for drug manufacturing for fighting COVID-19. For example, a solution based on molecular docking was proposed in [118] for drug investigations. More than 2500 small molecules in the FDA-approved drug database were first screened and validated through a molecular docking program called Glide. As a result, fifteen out of twenty-five drugs validated in exhibited significant inhibitory potencies, inhibiting signaling pathway and inflammatory responses and prompting drug repositioning against COVID-19. A big data-driven drug repositioning scheme was also introduced in [119]. The key purpose of this project is to apply ML to combine both knowledge graph and literature for COVID-19 vaccine development.

#### E. Summary

Reviewing from the rapidly emerging literature, we find that big data plays an important role in combating the COVID-19 pandemic through a number of promising applications, including outbreak prediction, virus spread tracking, coronavirus diagnosis/treatment, and vaccine/drug discovery. In fact, big data potentially enables outbreak prediction on the global scope using data analytic tools on huge datasets collected from available sources such as health organizations (e.g. WHO), healthcare institutes (i.e. China National Health Commission) [94], [95]. Big data has also emerged as a promising solution for coronavirus spread tracking by combining with intelligent tools such as ML and DL [102] for building prediction models that is very useful for governments in monitoring the potential COVID-19 outbreak in the future. Moreover, big data has the potential to support COVID-19 diagnosis and treatment processes. The research results from the literature studies prove that big data can help healthcare provide various medical operations from early diagnosis, disease analysis, and prediction of treatment outcomes [108], [109]. Finally, data learning from big datasets also helps finding potential targets for an effective vaccine against SARS-CoV-2 [115], and integrating large-scale knowledge graph, literature and transcriptome data, supporting to discover the potential drug candidates against SARS-CoV-2 [119].

### V. EXAMPLES OF AI AND BIG DATA FRAMEWORKS FOR COVID-19

This section presents two examples of AI- and big data based frameworks of how AI and big data are currently helping in the fight against the COVID-19 pandemic

#### A. Phone-based solution for detection and surveillance

Thanks to recent advancements in hardware capabilities, evolution of 5G wireless communications, and emergence of edge/cloud computing, mobile phone have many opportunities during the outbreak, e.g., outbreak identification, detection and diagnosis, treatment and infected case management, and disease elimination. Adapted from [120], we propose a mobile phone based framework for COVID-19 detection and surveillance, as shown in Fig. 5. Wireless connectivity, e.g., WiFi, 4G, and 5G, has been widely deployed in almost all the countries, so every individual with a mobile phone can connect to the Internet and can do many activities online [121]. The DL model can be trained in the cloud, even at a server collocated at the network edge, which is then pushed to mobile phone for further purpose [33]. For example, by using the embedded cameras and biosensors, a mobile phone can collect personal information, for example, X-ray and CT images, cough sound, and heart rate, which may be encrypted and compressed before sending to the cloud for training [31], [56]. Review on recent advances in biosensors for mobile health platforms can be found in [122]. The massive number of mobile devices connected to the Internet may relax the limited data sent from an individual phone. According to the latest annual Internet report from Cisco [123], by 2023 around 29.3 billion networked devices will be connected to the Internet,

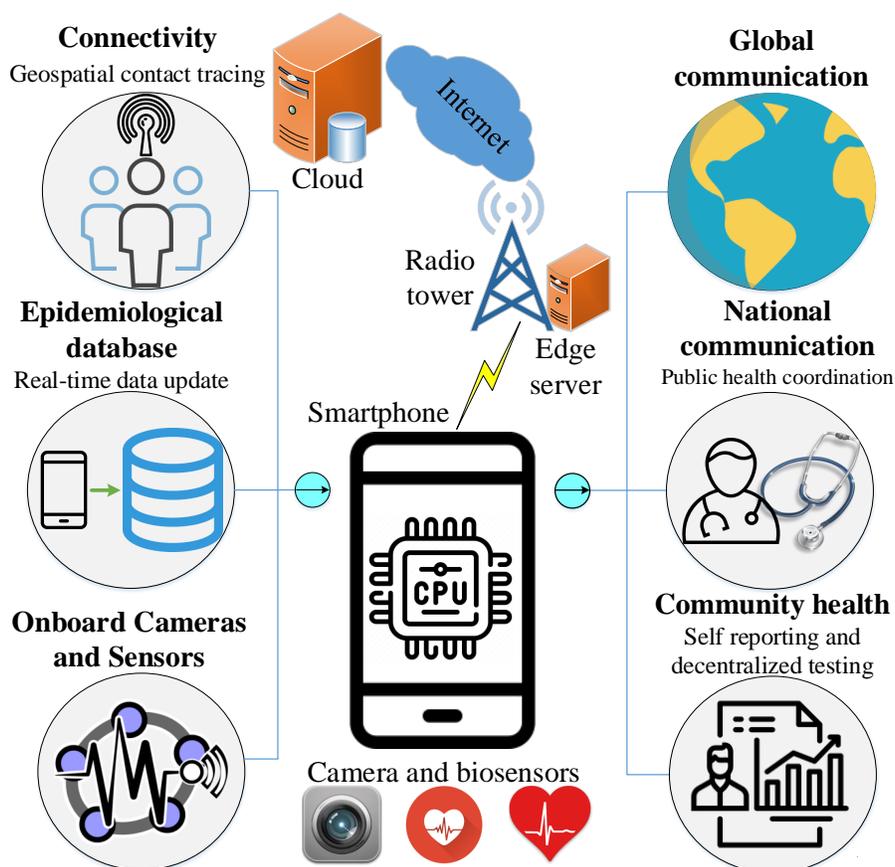


Fig. 5: An AI-based framework using mobile phones for COVID-19 diagnosis and surveillance. Adapted from [120].

where mobile phones and tablets account for 23% and 3% respectively, and over 70% of the world population will have mobile connectivity.

Despite many potentials, a number of challenges need to be addressed for the success of AI-based mHealth [124]. The very first challenge is from the capability and reliability of hardware and software, which are used in mobile devices to the medical purpose. This challenge can be overcome by developing standard mobile devices with specific computing/sensing components and system on chip (SoC). Another challenge arises from the protocol to collect/store the diagnosis results, and the path way to improve the patients' trust, especially when they normally prefer to have a face-to-face talk with doctors and clinicians. For example, depending on local and national policies, the diagnostic results collected from individuals in a region are required to transmit to and store at the local server instead of a cloud server, which would be used to train the DL model for global communication. Lastly, a challenge in utilizing mobile devices for the detection and surveillance is how to evaluate the clinical cost and effectiveness. This is reasonable since the results using phone-based frameworks are usually questionable and not comparable with those of the direct tests.

#### B. AI and big data for neutralizing antibody discovery

Under the current situation that specific medicines and vaccines for COVID-19 virus are not available, it is urgent to find

efficient and fast solution to combat the COVID-19 disease and prevent the virus outbreak. Motivated by this, the authors in [125] introduced a data-driven framework that combines AI, big data, and medical knowledge to identify antibody sequences that can inhibit the growth of COVID-19 virus. The schematic illustration of the framework proposed in [125] is pictorially shown in Fig. 6. The initial dataset is consisted of 1831 antigen and antibody sequences of various viruses, for example, HIV, H1N1, Dengue, SARS, and Ebola from the CATNAP tool [126], and their corresponding half maximal inhibitory concentration (IC<sub>50</sub>) values<sup>3</sup>. Moreover, the authors proposed mining more 102 data samples of from the RCSP protein data bank [128] (URL: <https://www.rcsb.org/>), which is then added to the initial data set, thus creating the final dataset of 1933 samples.

The next phase is extracting the features from the dataset of 1933 samples. In doing so, the authors exploited an open-source cheminformatics software (available at <https://www.rdkit.org/>), namely RDKit, to represent the molecular graph. Moreover, to reduce computational complexity and variance of the training data and, the authors applied an average-pool layer over the extracted features. Five ML models, including random forest, XGBoost, multilayer perceptron, support vector machines, and logistic regression, are then applied to evaluate their performance in terms of five-fold cross validation. The

<sup>3</sup>IC<sub>50</sub> is a quantitative measure that denotes the concentration of a substance required for 50% inhibition [127].

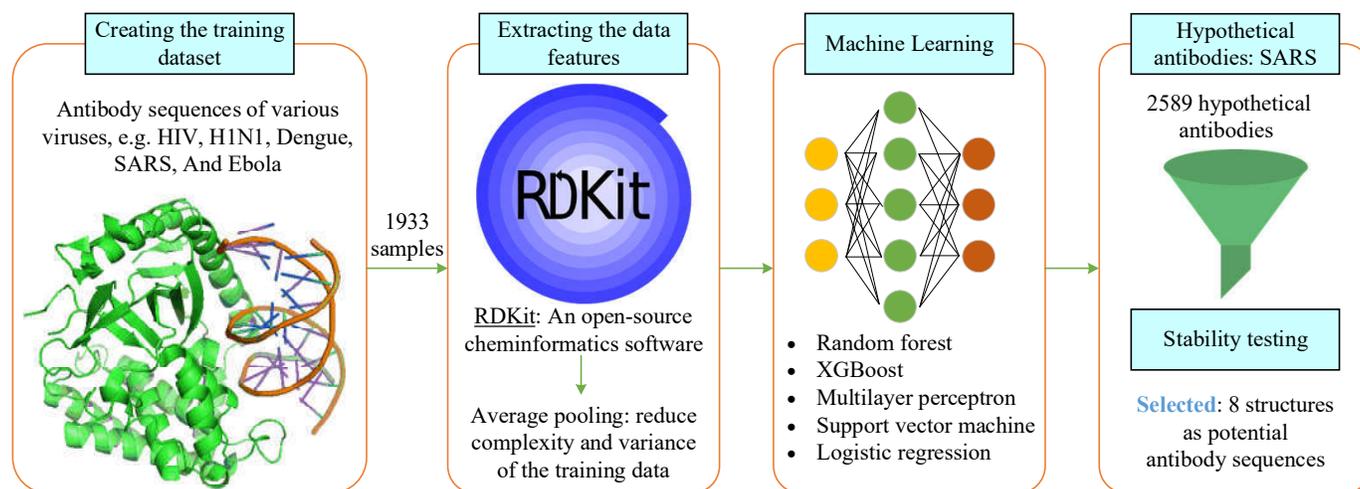


Fig. 6: Illustration of a data-driven framework for discovering antibody sequences to treat the COVID-19 disease [125].

experiment results show that the XGBoost model is the most superior one, which is therefore selected to find the potential antibody candidates for COVID-19 virus. More importantly, the XGBoost model is highly promising as it can achieve an outstanding out-of-class prediction over various virus families, for example, 100% for SARS and Dengue, 84.61% for Influenza, and 75% Ebola and Hepatitis. After that, since COVID-19 virus has similar characteristics (88-89% similarity [129]) with SARS-CoV-2, the authors proposed generating a set of 2589 hypothetical candidates based on neutralizing antibodies and sequence features against SARS. Out of all the candidates, finally 8 structures are selected as potential antibody sequences for neutralizing the COVID-19 disease. The results discovered in this study would be used for further academic researches as well as find effective medicines and vaccines for the treatment of COVID-19.

## VI. CHALLENGES, LESSONS, AND RECOMMENDATIONS

As reviewed in the previous sections, AI and big data have found their great potentials in the globe battle against the COVID-19 pandemic. Apart from the obvious advantages, there are still challenges needed to be discussed and addressed in the future. Additionally, we highlight some lessons stemmed from this paper and give some recommendations for the research communities and authorities.

### A. Challenges and solutions

1) *Regulation:* As the outbreak is booming and the daily number of confirmed cases (infected and dead) increases considerably, various approaches have been taken to control this outbreak, e.g., lockdown, social distancing, screen and testing at a large scale. In this way, regulatory authorities occupy a crucial role in defining policies that can encourage the involvement of residents, scientists and researchers, industry, giant techs and large firms, as well as harmonizing the approaches executed by different entities to avoid any barriers and obstacles in the way of preventing the COVID-19 disease.

Regarding this challenge, many attempts have been made from the first confirmed COVID-19 to current situation. An example is from the quarantine policy in Korea, effective from 1st April 2020. More specifically, all passengers entering Korea are required to be quarantined for 14 days at registered addresses or designated facilities. Moreover, all passengers are required to take daily self-diagnosis twice a day and send the reports using self-diagnosis apps installed in their mobile phones. An “AI monitoring calling system” has been implemented by the Seoul Metropolitan Government to automatically check the health conditions of the people who do not have any mobile phones and/or have not installed the self-diagnosis apps [130]. Another effort is from the collaboration between the Zhejiang Provincial Government and Alibaba DAMO Academy to develop an AI platform for automatic the COVID-19 testing and analysis [131].

2) *Lack of standard datasets:* In order to make AI and big data platforms and applications a trustful solution to fight the COVID-19 virus, a critical challenge arises from the lack of standard datasets. As surveyed in the previous sections, many AI algorithms and big data platforms have been proposed, but they are not tested using the same dataset. For example, the algorithms in [49] and [50] are verified to respectively achieve accuracy of 82.9%/98.27%, specificity of 80.5% and 97.60%, sensitivity of 84% and 98.93%. However, we cannot decide which algorithm is better for the virus detection since two datasets with different numbers of samples are used. Furthermore, most datasets found in the literature have been made thanks to individual efforts, e.g., the authors collect some datasets available in the Internet, then unify them to create their own dataset and evaluate their proposed algorithms.

To overcome this challenge, government, giant firms, and health organizations (e.g., WHO and CDC) play a key role as they can collaboratively work for high-quality and big datasets. A variety of data sources can be provided by these entities, e.g., X-ray and CT scans from the hospitals, satellite data, personal information and reports from self-diagnosis apps. For example, the COVID-19 dataset [14] has been made and led by the Georgetown’s Center for Security and other partners like

Allen Institute for AI, Chan Zuckerberg Initiative, Microsoft Research, and National Institutes of Health. Alternatively, Alibaba DAMO Academy has worked with many Hospitals in China to build AI systems for detecting the COVID-19 infected cases, where Alibaba DAMO Academy is responsible for designing AI algorithms and hospitals are responsible for providing CT scans of more than 5,000 confirmed COVID-19 cases. As reported in [132], this system has been used by more than 20 hospitals in China thanks to its remarkable performance: an accuracy of 96% within 20 seconds only. As individual efforts, an initiative to manage public datasets of COVID-19 medical images is available at [133] and a global collection of open-source projects on COVID-19 is available at <http://open-source-covid-19.weileizeng.com/>.

3) *Privacy and security challenges*: Right now, important things are keeping people healthy and soon controlling the situation; however, how personal data secure and private is still needed and should be investigated. A showcase of this challenge is the scandal of the Zoom videoconferencing app over its security and privacy issues<sup>4</sup>. In the pandemic, authorities may request their people to share their personal information, for example, GPS location, CT scans, diagnosis reports, travel trajectory, and daily activities, which is needed to control the situation, make up-to-date policies, and decide immediate actions. Data is a must to guarantee the success of any AI and big data platforms; however, normally people do not want to share their personal information, if not officially requested. There is a trade-off: privacy/security and performance.

Many technologies are available to solve the privacy and security issues during the COVID-19 pandemic. Here, we consider some potential solutions below, which can serve as research directives in the future.

- *Blockchain*: Basically, blockchain is defined as a decentralized, immutable and public database, where each transaction is verified by all the nodes in the network, which is enabled by consensus algorithms [134], [135]. Blockchain has found its success for various healthcare applications [135], [136], so it is possible to deploy blockchain based solutions to improve the user security and data privacy during the COVID-19 outbreak period. MiPasa is one of the projects that combines two emerging technologies from IBM: blockchain and cloud computing [137]. The purpose of this project is to provide reliable, accessible, and high-quality data to the communities.
- *Federated learning (FL)*: Traditionally, the data should be collected and stored centrally to training DL models. Federated learning offers a new solution in which a majority of personal data is not required to be transmitted to the central server [138]. Applying to the AI-based framework using mobile phones for diagnosing COVID-19 case, each phone can train its own DL model using local data. Mobile phones transmit their trained models to a central server, which is responsible for aggregating to make a global model that is then disseminated to all the mobile phones. We note that a phone is not limited

to mobile phone only, it can be a server at the local health department, whereas the aggregation server can be a global cloud like Microsoft Azure and Amazon Web Services.

- *Incentive mechanisms*: A large and reliable dataset lays a foundation for AI and big data platforms contend with the COVID-19 outbreak. Therefore, there is a need for incentive design to call the participation of more people and entities in contributing their own data. Incentives are needed due to the following reasons: 1) a massive quantity of data are available from people/entities, who may be not requested by the governments to provide their data, and 2) the quality of data should be guaranteed in order to improve the accuracy and performance of learning models. Such incentive mechanisms for health-care, wireless communications, transportation, etc. can be found in [139].

## B. Lessons and Recommendations

Reviewing the state-of-the-art literature, we find that AI and big data technologies play a key role in combating the COVID-19 pandemic through a variety of appealing applications, ranging from outbreak tracking, virus detection to treatment and diagnosis support. On the one hand, AI is able to provide viable solutions for fighting the COVID-19 pandemic in several ways. For example, AI has proved very useful for supporting outbreak prediction, coronavirus detection as well as infodemiology and infoveillance by leveraging learning-based techniques such as ML and DL from COVID-19-centric modeling, classification, and estimation. Moreover, AI has emerged as an attractive tool for facilitating vaccine and drug manufacturing. By using the datasets provided by healthcare organizations, governments, clinical labs and patients, AI leverages intelligent analytic tools for predicting effective and safe vaccine/drug against COVID-19, which would be beneficial from both the economic and scientific perspectives. On the other hand, big data has been proved its capability to tackle the COVID-19 pandemic. Big data potentially provides various promising solutions to help fight the COVID-19 pandemic. By combining with AI analytics, big data helps us to understand the COVID-19 in terms of virus structure and disease development. Big data can help healthcare providers in various medical operations from early diagnosis, disease analysis to prediction of treatment outcomes. With its great potentials, the integration of AI and big data can be the key enabler for governments in fighting the potential COVID-19 outbreak in the future.

Some recommendations can be considered to promote the COVID-19 fighting. First, AI and big data-based algorithms should be optimized further to enhance the accuracy and reliability of the data analytics for better COVID-19 diagnosis and treatment. Second, AI and big data can be incorporated with other emerging technologies to offer newly effective solutions for fighting COVID-19. For example, data analytic tools from Oracle cloud computing have been leveraged to design Vaccine, a new vaccine candidate against the COVID-19 virus [140]. More interestingly, the Chinese government has

<sup>4</sup>Zoom admits user data 'mistakenly' routed through China: <https://www.ft.com/content/2fc518e0-26cd-4d5f-8419-fe71f5c55c98>

recently incorporated drones into their response to the COVID pandemic [141] via a number of applications, such as delivery of testing samples, goods transport, and people's movement monitoring for outbreak tracking. Final, non-technology measures such as social distancing restrictions [142] still play a highly important role in slowing down the virus spread and thus need to be implemented effectively under the management of government agencies, aiming to control the COVID-19 pandemic in the future.

## VII. CONCLUSION

In this paper, we have presented a survey on the state-of-the-art solutions in the battle against the COVID-19 pandemic. Firstly, we have provided an introduction of the COVID-19 virus, the fundamentals and motivations of AI and big data for finding fast and effective approaches that can effectively combat the COVID-19 disease. Then, we have reviewed the applications of AI for detection and diagnosis, tracking and predicting the outbreak, infodemiology and infoveillance, biomedicine and pharmacotherapy. The applications of big data for the COVID-19 disease have been also presented, including outbreak prediction, virus spread tracking, diagnosis and treatment, and vaccine and drug discovery. Furthermore, we have discussed the challenges needed to overcome for the success of AI and big data in fighting the COVID-19 pandemic. Finally, we have highlighted important lessons and recommendations for the authorities and research communities.

## REFERENCES

- [1] "Situation update worldwide, as of 9 April 2020," 2020. [Online]. Available: <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>
- [2] "Coronavirus disease (COVID-19) pandemic," 2020. [Online]. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- [3] "Coronavirus (COVID-19)," 2020. [Online]. Available: <https://www.cdc.gov/coronavirus/2019-nCoV/index.html>
- [4] "White House announces new partnership to unleash U.S. supercomputing resources to fight COVID-19," 2020. [Online]. Available: <https://www.whitehouse.gov/briefings-statements>
- [5] "arXiv announces new COVID-19 quick search," 2020. [Online]. Available: <https://blogs.cornell.edu/arxiv/2020/03/30/new-covid-19-quick-search/>
- [6] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *International Journal of Surgery*, vol. 76, pp. 71–76, 2020.
- [7] H. Li, S.-M. Liu, X.-H. Yu, S.-L. Tang, and C.-K. Tang, "Coronavirus disease 2019 (COVID-19): current status and future perspectives," *International Journal of Antimicrobial Agents*, p. 105951, 2020.
- [8] T. Huynh-The, C. Hua, Q.-V. Pham, and D. Kim, "MCNet: An efficient CNN architecture for robust automatic modulation classification," *IEEE Communications Letters*, vol. 24, no. 4, pp. 811–815, Apr. 2020.
- [9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [10] B. R. Beck, B. Shin, Y. Choi, S. Park, and K. Kang, "Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model," *BioRxiv*, 2020.
- [11] A. Zhavoronkov, V. Aladinskiy, A. Zhebrak, B. Zagribelnyy, V. Terentiev, D. S. Bezrukov, D. Polykovskiy, R. Shayakhmetov, A. Filimonov, P. Orekhov, Y. Yan, O. Popova, G. Vanhaelen, A. Aliper, and Y. Ivanenkov, "Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches," *ChemRxiv*, 2 2020.
- [12] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for COVID-19 from chest CT using weak label," *MedRxiv*, 2020.
- [13] Z. Hu, Q. Ge, L. Jin, and M. Xiong, "Artificial intelligence forecasting of COVID-19 in China," *arXiv preprint arXiv:2002.07112*, 2020.
- [14] "COVID-19 open research dataset challenge (CORD-19): An AI challenge with AI2, CZI, MSR, Georgetown, NIH & The White House," 2020. [Online]. Available: [www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge](http://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge)
- [15] "IBM releases novel AI-powered technologies to help health and research community accelerate the discovery of medical insights and treatments for COVID-19," 2020. [Online]. Available: <https://www.ibm.com/blogs/research/2020/04/ai-powered-technologies-accelerate-discovery-covid-19/>
- [16] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, "Big data analytics: a survey," *Journal of Big data*, vol. 2, no. 1, p. 21, 2015.
- [17] K. Priyanka and N. Kulennavar, "A survey on big data analytics in health care," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5865–5868, 2014.
- [18] M. Cottle, W. Hoover, S. Kanwal, M. Kohn, T. Strome, and N. Treister, "Transforming health care through big data strategies for leveraging big data in the health care industry," *Institute for Health Technology Transformation*, <http://ihealthtran.com/big-data-in-healthcare>, 2013.
- [19] "Almost one zettabyte of mobile data traffic in 2022-cisco," 2020. [Online]. Available: <https://telecoms.com/495666/almost-one-zettabyte-of-mobile-data-traffic-in-2022-cisco/>
- [20] G. Manogaran, D. Lopez, C. Thota, K. M. Abbas, S. Pyne, and R. Sundarasekar, "Big data analytics in healthcare internet of things," in *Innovative healthcare systems for the 21st century*. Springer, 2017, pp. 263–284.
- [21] G. Manogaran, C. Thota, D. Lopez, V. Vijayakumar, K. M. Abbas, and R. Sundarasekar, "Big data knowledge system in healthcare," in *Internet of things and big data technologies for next generation healthcare*. Springer, 2017, pp. 133–157.
- [22] S. Chae, S. Kwon, and D. Lee, "Predicting infectious disease using deep learning and big data," *International journal of environmental research and public health*, vol. 15, no. 8, p. 1596, 2018.
- [23] S. Bansal, G. Chowell, L. Simonsen, A. Vespignani, and C. Viboud, "Big data for infectious disease surveillance and modeling," *The Journal of infectious diseases*, vol. 214, no. suppl\_4, pp. S375–S379, 2016.
- [24] M. Eisenstein, "Infection forecasts powered by big data," *Nature*, vol. 555, no. 7695, 2018.
- [25] "Improving epidemic surveillance and response: big data is dead, long live big data," 2020. [Online]. Available: [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(20\)30059-5/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30059-5/fulltext)
- [26] "Big data in the time of coronavirus (COVID-19)," 2020. [Online]. Available: <https://www.forbes.com/sites/ciocentral/2020/03/30/big-data-in-the-time-of-coronavirus-19/>
- [27] "Understanding the COVID-19 pandemic as a big data analytics issue," 2020. [Online]. Available: <https://healthitanalytics.com/news/understanding-the-19-pandemic-as-a-big-data-analytics-issue>
- [28] V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. Chu, T. Bleicker, S. Brünink, J. Schneider, M. L. Schmidt *et al.*, "Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR," *Eurosurveillance*, vol. 25, no. 3, 2020.
- [29] A. S. Fomsgaard and M. W. Rosenstjerne, "An alternative workflow for molecular detection of SARS-CoV-2-escape from the NA extraction kit-shortage," *medRxiv*, 2020.
- [30] H. S. Maghdid, K. Z. Ghafour, A. S. Sadiq, K. Curran, and K. Rabie, "A novel AI-enabled framework to diagnose coronavirus COVID-19 using smartphone embedded sensors: Design study," *arXiv preprint arXiv:2003.07434*, 2020.
- [31] A. S. S. Rao and J. A. Vazquez, "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," *Infection Control & Hospital Epidemiology*, p. 1–18, 2020.
- [32] B. M. Silva, J. J. Rodrigues, I. [de la Torre Diez], M. López-Coronado, and K. Saleem, "Mobile-health: A review of current state in 2015," *Journal of Biomedical Informatics*, vol. 56, pp. 265–272, 2015.
- [33] Q.-V. Pham, F. Fang, V. N. Ha, M. Le, Z. Ding, L. B. Le, and W.-J. Hwang, "A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art," *CoRR*, 2019. [Online]. Available: [arxiv.org/abs/1906.08452](https://arxiv.org/abs/1906.08452)
- [34] Q.-V. Pham, T. Leanh, N. H. Tran, B. J. Park, and C. S. Hong, "Decentralized computation offloading and resource allocation for mobile-edge computing: A matching game approach," *IEEE Access*, vol. 6, pp. 75 868–75 885, Nov. 2018.

- [35] O. Gozes, M. Frid-Adar, N. Sagie, H. Zhang, W. Ji, and H. Greenspan, "Coronavirus detection and analysis on chest CT with deep learning," *arXiv preprint arXiv:2004.02640*, 2020.
- [36] M. Barstugan, U. Ozkaya, and S. Ozturk, "Coronavirus (COVID-19) classification using CT images by machine learning methods," *arXiv preprint arXiv:2003.09424*, 2020.
- [37] L. O. Hall, R. Paul, D. B. Goldgof, and G. M. Goldgof, "Finding Covid-19 from chest X-rays using deep learning on a small dataset," *arXiv preprint arXiv:2004.02060*, 2020.
- [38] N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien, and S. Elghamrawy, "Detection of coronavirus (COVID-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest X-ray dataset," *arXiv preprint arXiv:2004.01184*, 2020.
- [39] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *arXiv preprint arXiv:2003.13815*, 2020.
- [40] K. E. Asnaoui, Y. Chawki, and A. Idri, "Automated methods for detection and classification pneumonia based on X-ray images using deep learning," *arXiv preprint arXiv:2003.14363*, 2020.
- [41] I. D. Apostolopoulos and T. Bessiana, "Covid-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *arXiv preprint arXiv:2003.11617*, 2020.
- [42] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *arXiv preprint arXiv:2003.10849*, 2020.
- [43] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images," 2020.
- [44] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection," *arXiv preprint arXiv:2003.10769*, 2020.
- [45] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [46] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.
- [47] L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images," *arXiv preprint arXiv:2003.09871*, 2020.
- [48] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. Siegel, "Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," *arXiv preprint arXiv:2003.05037*, 2020.
- [49] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng *et al.*, "A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)," *medRxiv*, 2020.
- [50] U. Ozkaya, S. Ozturk, and M. Barstugan, "Coronavirus (COVID-19) classification using deep features fusion and ranking technique," *arXiv preprint arXiv:2004.03698*, 2020.
- [51] W. cai Dai, H. wen Zhang, J. Yu, H. jian Xu, H. Chen, S. ping Luo, H. Zhang, L. hong Liang, X. liu Wu, Y. Lei, and F. Lin, "CT imaging and differential diagnosis of COVID-19," *Canadian Association of Radiologists Journal*, vol. 71, no. 2, pp. 195–200, 2020.
- [52] S. Chaganti, A. Balachandran, G. Chabin, S. Cohen, T. Flohr, B. Georgescu, P. Grenier, S. Grbic, S. Liu, F. Mellot, N. Murray, S. Nicolaou, W. Parker, T. Re, P. Sanelli, A. W. Sauter, Z. Xu, Y. Yoo, V. Ziebandt, and D. Comaniciu, "Quantification of tomographic patterns associated with COVID-19 from chest CT," *arXiv preprint arXiv:2004.01279*, 2020.
- [53] F. Shan+, Y. Gao+, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, "Lung infection quantification of COVID-19 in CT images with deep learning," *arXiv preprint arXiv:2003.04655*, 2020.
- [54] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.
- [55] Z. Tang, W. Zhao, X. Xie, Z. Zhong, F. Shi, J. Liu, and D. Shen, "Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images," *arXiv preprint arXiv:2003.11988*, 2020.
- [56] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood, S. Riaz, K. Ali, C. N. John, and M. Nabeel, "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app," *arXiv preprint arXiv:2004.01275*, 2020.
- [57] C. Qi, D. Karlsson, K. Sallmen, and R. Wyss, "Model studies on the COVID-19 pandemic in Sweden," *arXiv preprint arXiv:2004.01575*, 2020.
- [58] Y.-C. Chen, P.-E. Lu, and C.-S. Chang, "A time-dependent SIR model for COVID-19," *arXiv preprint arXiv:2003.00122*, 2020.
- [59] K. Biswas and P. Sen, "Space-time dependence of corona virus (COVID-19) outbreak," *arXiv preprint arXiv:2003.03149*, 2020.
- [60] N. Crokidakis, "Data analysis and modeling of the evolution of COVID-19 in Brazil," *arXiv preprint arXiv:2003.12150*, 2020.
- [61] G. Gaeta, "A simple SIR model with a large set of asymptomatic infectives," *arXiv preprint arXiv:2003.08720*, 2020.
- [62] R. Dandekar and G. Barbastathis, "Neural network aided quarantine control model estimation of COVID spread in Wuhan, China," *arXiv preprint arXiv:2003.09403*, 2020.
- [63] —, "Neural network aided quarantine control model estimation of global Covid-19 spread," *arXiv preprint arXiv:2004.02752*, 2020.
- [64] X. Zhou, N. Hong, Y. Ma, J. He, H. Jiang, C. Liu, G. Shan, L. Su, W. Zhu, and Y. Long, "Forecasting the worldwide spread of COVID-19 based on logistic model and SEIR model," *MedRxiv*, 2020.
- [65] C. Bayes, V. S. y Rosas, and L. Valdivieso, "Modelling death rates due to COVID-19: A Bayesian approach," *arXiv preprint arXiv:2004.02386*, 2020.
- [66] B. M. Ndiaye, L. Tendeng, and D. Seck, "Analysis of the COVID-19 pandemic by SIR model and machine learning technics for forecasting," 2020.
- [67] G. Perone, "An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy," 2020.
- [68] M. Magdon-Ismail, "Machine learning the phenomenology of COVID-19 from early infection dynamics," *arXiv preprint arXiv:2003.07602*, 2020.
- [69] Z. Hu, Q. Ge, S. Li, E. Boerwinkle, L. Jin, and M. Xiong, "Forecasting and evaluating intervention of COVID-19 in the World," *arXiv preprint arXiv:2003.09800*, 2020.
- [70] R. Pal, A. A. Sekh, S. Kar, and D. K. Prasad, "Neural network based country wise risk prediction of COVID-19," *arXiv preprint arXiv:2004.00959*, 2020.
- [71] K. Ganasegeran and S. A. Abdulrahman, *Artificial Intelligence Applications in Tracking Health Behaviors During Disease Epidemics*. Cham: Springer International Publishing, 2020, pp. 141–155.
- [72] Z. Hou, F. Du, H. Jiang, X. Zhou, and L. Lin, "Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: social media surveillance in China," *Risk Perception, Emotional and Behavioural Responses to the COVID-19 Outbreak: Social Media Surveillance in China (3/6/2020)*, 2020.
- [73] B. W. Schuller, D. M. Schuller, K. Qian, J. Liu, H. Zheng, and X. Li, "COVID-19 and computer audition: An overview on what speech & sound analysis could contribute in the SARS-CoV-2 corona crisis," *arXiv preprint arXiv:2003.11117*, 2020.
- [74] Y. Ye, S. Hou, Y. Fan, Y. Qian, Y. Zhang, S. Sun, Q. Peng, and K. Laparo, "α-satellite: An AI-driven system and benchmark datasets for hierarchical community-level risk assessment to help combat COVID-19," *arXiv preprint arXiv:2003.12232*, 2020.
- [75] D. Liu, L. Clemente, C. Poirier, X. Ding, M. Chinazzi, J. T. Davis, A. Vespignani, and M. Santillana, "A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using internet searches, news alerts, and estimates from mechanistic models," *arXiv preprint arXiv:2004.04019*, 2020.
- [76] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, 2020.
- [77] P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov, "Applications of deep learning in biomedicine," *Molecular pharmaceuticals*, vol. 13, no. 5, pp. 1445–1454, 2016.
- [78] C. Cao, F. Liu, H. Tan, D. Song, W. Shu, W. Li, Y. Zhou, X. Bo, and Z. Xie, "Deep learning and its applications in biomedicine," *Genomics, Proteomics & Bioinformatics*, vol. 16, no. 1, pp. 17 – 32, 2018.
- [79] S. Ekins, A. C. Puhl, K. M. Zorn, T. R. Lane, D. P. Russo, J. J. Klein, A. J. Hickey, and A. M. Clark, "Exploiting machine learning for end-to-end drug discovery and development," *Nature materials*, vol. 18, no. 5, p. 435, 2019.
- [80] F. Hu, J. Jiang, and P. Yin, "Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model," *arXiv preprint arXiv:2003.00728*, 2020.

- [81] Y. Ge, T. Tian, S. Huang, F. Wan, J. Li, S. Li, H. Yang, L. Hong, N. Wu, E. Yuan *et al.*, "A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19," *BioRxiv*, 2020.
- [82] N. Savioli, "One-shot screening of potential peptide ligands on HR1 domain in COVID-19 glycosylated spike (S) protein with deep siamese network," *arXiv preprint arXiv:2004.02136*, 2020.
- [83] A.-T. Ton, F. Gentile, M. Hsing, F. Ban, and A. Cherkasov, "Rapid identification of potential inhibitors of sars-cov-2 main protease by deep docking of 1.3 billion compounds," *Molecular Informatics*, 2020.
- [84] V. Chenthamarakshan, P. Das, I. Padhi, H. Strobelt, K. W. Lim, B. Hoover, S. C. Hoffman, and A. Mojsilovic, "Target-specific and selective drug design for COVID-19 using deep generative models," *arXiv preprint arXiv:2004.01215*, 2020.
- [85] M. Hofmarcher, A. Mayr, E. Rumetshofer, P. Ruch, P. Renz, J. Schimunek, P. Seidl, A. Vall, M. Widrich, S. Hochreiter *et al.*, "Large-scale ligand-based virtual screening for SARS-CoV-2 inhibitors using deep neural networks," Available at SSRN 3561442, 2020.
- [86] E. Ong, M. U. Wong, A. Huffman, and Y. He, "COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning," *BioRxiv*, 2020.
- [87] J. Jumper, K. Tunyasuvunakool, P. Kohli, D. Hassabis, and A. Team, "Computational predictions of protein structures associated with COVID-19," 2020, DeepMind.
- [88] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Židek, A. W. Nelson, A. Bridgland *et al.*, "Improved protein structure prediction using potentials from deep learning," *Nature*, pp. 1–5, 2020.
- [89] A. Strokach, D. Becerra, C. Corbi-Verge, A. Perez-Riba, and P. M. Kim, "Fast and flexible design of novel proteins using graph neural networks," *BioRxiv*, 2020.
- [90] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri *et al.*, "A SIDARTHE model of COVID-19 epidemic in Italy," *arXiv preprint arXiv:2003.09861*, 2020.
- [91] F. Brauer, C. Castillo-Chavez, and C. Castillo-Chavez, *Mathematical models in population biology and epidemiology*. Springer, 2012, vol. 2.
- [92] B. Chen, M. Shi, X. Ni, L. Ruan, H. Jiang, H. Yao, M. Wang, Z. Song, Q. Zhou, and T. Ge, "Visual data analysis and simulation prediction for COVID-19," *arXiv preprint arXiv:2002.07096*, 2020.
- [93] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of COVID-19 in China by dynamical modeling," *arXiv preprint arXiv:2002.06563*, 2020.
- [94] D. Tátrai and Z. Várallyay, "COVID-19 epidemic outcome predictions based on logistic fitting and estimation of its reliability," *arXiv preprint arXiv:2003.14160*, 2020.
- [95] A. Strzelecki, "The second worldwide wave of interest in coronavirus since the COVID-19 outbreaks in South Korea, Italy and Iran: A Google trends study," *arXiv preprint arXiv:2003.10998*, 2020.
- [96] Y.-S. Long, Z.-M. Zhai, L.-L. Han, J. Kang, Y.-L. Li, Z.-H. Lin, L. Zeng, D.-Y. Wu, C.-Q. Hao, M. Tang *et al.*, "Quantitative assessment of the role of undocumented infection in the 2019 novel coronavirus (COVID-19) pandemic," *arXiv preprint arXiv:2003.12028*, 2020.
- [97] R. Gupta, G. Pandey, P. Chaudhary, and S. K. Pal, "SEIR and regression model based COVID-19 outbreak predictions in India," *medRxiv*, 2020.
- [98] S. Heroy, "Metropolitan-scale COVID-19 outbreaks: how similar are they?" *arXiv preprint arXiv:2004.01248*, 2020.
- [99] "Is big data effective in response to coronavirus outbreak?" 2020. [Online]. Available: <https://www.analyticsinsight.net/big-data-effective-response-coronavirus-outbreak/>
- [100] X. Zhao, X. Liu, and X. Li, "Tracking the spread of novel coronavirus (2019-ncov) based on big data," *medRxiv*, 2020.
- [101] C. Zhou, F. Su, T. Pei, A. Zhang, Y. Du, B. Luo, Z. Cao, J. Wang, W. Yuan, Y. Zhu *et al.*, "COVID-19: Challenges to GIS with big data," *Geography and Sustainability*, 2020.
- [102] P. Castorina, A. Iorio, and D. Lanteri, "Data analysis on coronavirus spreading by macroscopic growth laws," *arXiv preprint arXiv:2003.00507*, 2020.
- [103] A. Notari, "Temperature dependence of COVID-19 transmission," *arXiv preprint arXiv:2003.12417*, 2020.
- [104] V. Lampos, S. Moura, E. Yom-Tov, I. J. Cox, R. McKendry, and M. Edelstein, "Tracking COVID-19 using online search," *arXiv preprint arXiv:2003.08086*, 2020.
- [105] "How China is using AI and big data to fight the coronavirus," 2020. [Online]. Available: <https://www.aljazeera.com/news/2020/03/{China}-ai-big-data-combat-coronavirus-outbreak-200301063901951.html>
- [106] "How China is using big data and artificial intelligence to fight coronavirus," 2020. [Online]. Available: <https://www.internetsearchinc.com/how-{China}-is-using-big-data-and-artificial-intelligence-to-fight-coronavirus/>
- [107] C. Garattini, J. Raffle, D. N. Aisyah, F. Sartain, and Z. Kozlakidis, "Big data analytics, infectious diseases and associated ethical impacts," *Philosophy & technology*, vol. 32, no. 1, pp. 69–85, 2019.
- [108] C. Li, D. N. Debruyne, J. Spencer, Y. Kapoor, L. Y. Liu, B. Zhou, L. Lee, R. Feigelman, G. Burdon, J. Liu *et al.*, "High sensitivity detection of coronavirus SARS-CoV-2 using multiplex PCR and a multiplex-PCR-based metagenomic method," *bioRxiv*, 2020.
- [109] J.-S. Eden, R. Rockett, I. Carter, H. Rahman, J. de Ligt, J. Hadfield, M. Storey, X. Ren, R. Tulloch, K. Basile *et al.*, "An emergent clade of SARS-CoV-2 linked to returned travellers from Iran," *bioRxiv*, 2020.
- [110] I. Ortea and J.-O. Bock, "Re-analysis of SARS-CoV-2 infected host cell proteomics time-course data by impact pathway analysis and network analysis. a potential link with inflammatory response." *BioRxiv*, 2020.
- [111] D. Brann, T. Tsukahara, C. Weinreb, D. W. Logan, and S. R. Datta, "Non-neural expression of SARS-CoV-2 entry genes in the olfactory epithelium suggests mechanisms underlying anosmia in COVID-19 patients," *bioRxiv*, 2020.
- [112] J. R. Lon, Y. Bai, B. Zhong, F. Cai, and H. Du, "Prediction and evolution of B cell epitopes of surface protein in SARS-CoV-2," *bioRxiv*, 2020.
- [113] Y.-H. Jin, L. Cai, Z.-S. Cheng, H. Cheng, T. Deng, Y.-P. Fan, C. Fang, D. Huang, L.-Q. Huang, Q. Huang *et al.*, "A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-ncov) infected pneumonia (standard version)," *Military Medical Research*, vol. 7, no. 1, p. 4, 2020.
- [114] S. F. Ahmed, A. A. Quadeer, and M. R. McKay, "Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies," *Viruses*, vol. 12, no. 3, p. 254, 2020.
- [115] A. Banerjee, D. Santra, and S. Maiti, "Energetics based epitope screening in SARS CoV-2 (COVID 19) spike glycoprotein by immunoinformatic analysis aiming to a suitable vaccine development." *bioRxiv*, 2020.
- [116] B. Sarkar, M. A. Ullah, F. T. Johora, M. A. Taniya, and Y. Araf, "The essential facts of wuhan novel coronavirus outbreak in China and Epitope-based vaccine designing against 2019-nCoV," *BioRxiv*, 2020.
- [117] M. I. Abdelmageed, A. H. Abdelmoneim, M. I. Mustafa, N. M. Elfadol, N. S. Murshed, S. W. Shantier, and A. M. Makhawi, "Design of multi epitope-based peptide vaccine against e protein of human 2019-ncov: An immunoinformatics approach," *BioRxiv*, 2020.
- [118] Z. Li, X. Li, Y.-Y. Huang, Y. Wu, L. Zhou, R. Liu, D. Wu, L. Zhang, H. Liu, X. Xu *et al.*, "FEP-based screening prompts drug repositioning against COVID-19," *bioRxiv*, 2020.
- [119] Y. Ge, T. Tian, S. Huang, F. Wan, J. Li, S. Li, H. Yang, L. Hong, N. Wu, E. Yuan *et al.*, "A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19," *bioRxiv*, 2020.
- [120] B. Udagama, P. Kadhiresan, H. N. Kozlowski, A. Malekjhani, M. Osborne, V. Y. Li, H. Chen, S. Mubareka, J. Gubbay, and W. C. Chan, "Diagnosing COVID-19: The disease and tools for detection," *ACS Nano*, 2020.
- [121] Q.-V. Pham, L. B. Le, S. Chung, and W. Hwang, "Mobile edge computing with wireless backhaul: Joint task offloading and resource allocation," *IEEE Access*, vol. 7, pp. 16444–16459, Jan. 2019.
- [122] Z. Geng, X. Zhang, Z. Fan, X. Lv, Y. Su, and H. Chen, "Recent progress in optical biosensors based on smartphone platforms," *Sensors*, vol. 17, no. 11, p. 2449, 2017.
- [123] "Cisco annual internet report (2018–2023)," 2020. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html>
- [124] C. S. Wood, M. R. Thomas, J. Budd, T. P. Mashamba-Thompson, K. Herbst, D. Pillay, R. W. Peeling, A. M. Johnson, R. A. McKendry, and M. M. Stevens, "Taking connected mobile-health diagnostics of infectious diseases to the field," *Nature*, vol. 566, no. 7745, pp. 467–474, 2019.
- [125] R. Magar, P. Yadav, and A. B. Farimani, "Potential neutralizing antibodies discovered for novel corona virus using machine learning," *arXiv preprint arXiv:2003.08447*, 2020.
- [126] H. Yoon, J. Macke, A. P. West Jr, B. Foley, P. J. Bjorkman, B. Korber, and K. Yusim, "CATNAP: a tool to compile, analyze and tally neutralizing antibody panels," *Nucleic acids research*, vol. 43, no. W1, pp. W213–W219, 2015.
- [127] S. Offermanns and W. Rosenthal, Eds., *IC50 Values*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 611–611. [Online]. Available: [https://doi.org/10.1007/978-3-540-38918-7\\_5943](https://doi.org/10.1007/978-3-540-38918-7_5943)

- [128] H. M. Berman, P. E. Bourne, J. Westbrook, and C. Zardecki, "The protein data bank," in *Protein Structure*. CRC Press, 2003, pp. 394–410.
- [129] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *International Journal of Antimicrobial Agents*, vol. 55, no. 3, p. 105924, 2020.
- [130] "Seoul introduces the COVID-19 AI monitoring call system," 2020. [Online]. Available: <http://english.seoul.go.kr/seoul-introduces-the-covid-19-%E3%80%8Ccai-monitoring-call-system%E3%80%8D/>
- [131] "How next-generation information technologies tackled COVID-19 in China," 2020. [Online]. Available: [www.weforum.org/agenda/2020/04/how-next-generation-information-technologies-tackled-covid-19-in-china/](http://www.weforum.org/agenda/2020/04/how-next-generation-information-technologies-tackled-covid-19-in-china/)
- [132] "How DAMO academy's AI system detects coronavirus cases," 2020. [Online]. Available: <https://www.alizila.com/how-damo-academys-ai-system-detects-coronavirus-cases/>
- [133] R. Kalkreuth and P. Kaufmann, "COVID-19: A survey on public medical imaging data resources," *arXiv preprint arXiv:2004.04569*, 2020.
- [134] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, "Blockchain for 5G and beyond networks: A state of the art survey," *arXiv preprint arXiv:1912.05062*, 2019.
- [135] D. Nguyen, M. Ding, P. N. Pathirana, and A. Seneviratne, "Blockchain and AI-based solutions to combat coronavirus (COVID-19)-like epidemics: A survey," *10.36227/techrxiv.12121962.v1*, 2020.
- [136] T.-T. Kuo, H.-E. Kim, and L. Ohno-Machado, "Blockchain distributed ledger technologies for biomedical and health care applications," *Journal of the American Medical Informatics Association*, vol. 24, no. 6, pp. 1211–1220, 2017.
- [137] "MiPasa project and IBM Blockchain team on open data platform to support covid-19 response," 2020. [Online]. Available: <https://www.ibm.com/blogs/blockchain/2020/03>
- [138] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.
- [139] H. Gao, C. H. Liu, W. Wang, J. Zhao, Z. Song, X. Su, J. Crowcroft, and K. K. Leung, "A survey of incentive mechanisms for participatory sensing," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 918–943, 2015.
- [140] "AI and cloud computing used to develop COVID-19 vaccine," 2020. [Online]. Available: <https://www.drugtargetreview.com/news/59650/ai-and-cloud-computing-used-to-develop-covid-19-vaccine/>
- [141] "3 ways China is using drones to fight coronavirus," 2020. [Online]. Available: <https://www.forbes.com/sites/zakdoffman/2020/03/05/meet-the-coronavirus-spy-drones-that-make-sure-you-stay-home>
- [142] "Social distancing for coronavirus COVID-19," 2020. [Online]. Available: <https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/how-to-protect-yourself-and-others-from-coronavirus-covid-19/social-distancing-for-coronavirus-covid-19>