

# Substantial Role of Big Data in Global Outbreak: From Tracking and Diagnosing Virus to Vaccine Discovery

Hina ZAHOOR<sup>1</sup> , Nasir MUSTAFA<sup>2</sup> 

<sup>1</sup>Department of Social Service, İstanbul Gelişim University, Faculty of Health Science, İstanbul, Turkey

<sup>2</sup>Department of Nursing, İstanbul Gelişim University, Faculty of Health Science, İstanbul, Turkey

**Cite this article as:** Zahoor H, Mustafa N. Substantial role of big data in global outbreak: from tracking and diagnosing virus to vaccine discovery. *Arch Health Sci Res.* 2022;9(3):144-148.

144

## ABSTRACT

The recent pandemic is ramification of coronavirus SARS-CoV2. This outbreak has greatly affected every domain of individuals' lives. It has spread over 214 countries and over 392 million positive cases were reported till February 6, 2022. Due to the latest breakthroughs in the sphere of digital divide, Big Data can aid in dealing with the enormous data of COVID-19 derived from state health surveillance, health monitoring, and daily briefing of government bodies. "Big data" is huge amounts of facts that work wonderfully. It has become a subject of particular interest for the last 2 decades due to its unseen significant potential in it. The purpose of the current review is to overview the potential applications of Big Data. Furthermore, issues and challenges associated with the solutions to the pandemic situations were highlighted and last, recommendations were provided for effective control of the pandemic situation. This review is an effort to provide a fresh insight into the way of big data in terms to stop the pandemic outbreak.


**Keywords:** Pandemic outbreak, public health, big data, COVID-19, deep learning

## Recent Pandemic Outbreak

According to the available data, a number of cases were reported in the city of Wuhan, with respiratory symptoms and unknown causes in late December 2019.<sup>1</sup> The recent virus was first coined as "COVID-19" and, afterwards known as "SARS-CoV-2." The novel virus has spread all over the world, 214 countries were affected and more than 4.250.335 deaths reported. The infected individual may present with mild to severe symptoms such as high temperature, pain in throat to respiratory failure and death.<sup>2</sup> The novel virus, is a Communicable single-strand, RNA virus.<sup>3</sup> From a genetic point of view, the coronavirus shares approximately 50% of MERS-CoV and 79% of SARS-CoV. Moreover, SARS-CoV-2 also shares a receptor-binding with SARS-CoV.<sup>4</sup> There is no indication that the numbers of cases will decrease and the condition will be in hand. The global pandemic swing can be visualized in Figure 1.

In the present situation, the combined effort between national authorities and larger enterprises are foreseen to substantially lessen the threats from the spread of novel virus. For instance, Google, as a giant browser, initiated a portal for pandemic COVID-19 ([www.google.com/covid19](http://www.google.com/covid19)), where one can obtain valuable information, like, recent statistics, COVID-19 map and the most frequent questions on corona virus. Google, IBM, and Amazon also developed a system of supercomputing with the White House for the coronavirus-related researches.<sup>5</sup> Moreover, some publishing house offer free access to the COVID-19-related documents and research articles and archival services of web such as arXiv and bioRxiv also developed speedy link to collect COVID-19-related preprint papers. These technological advancements are playing a significant role in the fight against the recent outbreak. Several research papers and preprints have been made available online in the recent months to enhance the understanding of COVID-19 and to lessen the cases. The aim of this review paper is to highlight the role of big data in the battle against recent pandemic of COVID-19. Furthermore, issues and problems associated with existing big data set techniques are also provided in the present review to produce range of recommendations for the research bodies.

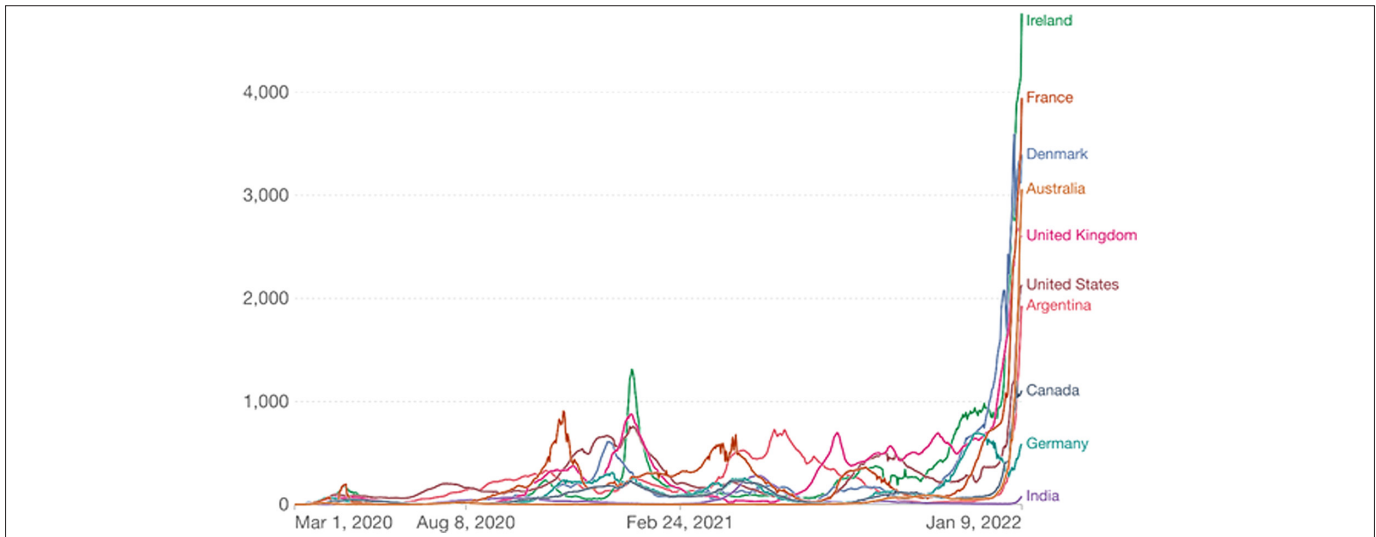
**Corresponding author:** Hina ZAHOOR, e-mail: [hinazahoor79@yahoo.com](mailto:hinazahoor79@yahoo.com)

 Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Received: November 17, 2021

Accepted: April 9, 2022

Available Online Date: May 30, 2022



**Figure 1.** The global COVID-19 trend (Source: Johns Hopkins University CSSE COVID-19 Data). Data accurate as of January 9, 2022.

During the recent era, the term big data has become enormously popular across the globe. Nearly all domains of research, whether it relates to academics' sector or any other industry is producing and analyzing big data for variety of purposes. Big Data is accentuated by 3 V's: (i) velocity (means unprecedented speed of data attainment, management and control); (ii) volume (means huge amount of availability of data); and (iii) variety (various sources and channels that can generate and share huge data).<sup>6-8</sup> Furthermore, on the basis of sources Big Data can be classified as molecular Big Data, imaging-based Big Data, sensor-based Big Data, and digital and computational Big Data.<sup>9-11</sup> So, Big Data assist in managing the massive, unusual amount of information taken from public health care monitoring, real-time outbreak of pandemic surveillance, forecasting, daily situation reporting and updating from public agencies and government bodies.<sup>12</sup>

### Big Data Application Fighting Against COVID-19

Data storage and analysis, in terms of COVID-19 big data analytics for COVID-19, is characterized primarily through some main techniques from the big data view, like multi-domain data set analysis, high-dimensional analysis, deep analysis, and parallel computing. In the following part of the article, we have highlighted the role of big data in rapid and efficient virus alert, tracking, diagnosing COVID-19 cases, and discover a potential medical treatment enabled by these techniques.

#### Rapid and Efficient Virus Alert

Big Data has the ability to monitor the outbreak of the disease in real-time. In regard to past pandemics outbreaks and epidemics, the recent pandemic is unparalleled in the sense that open-access data sets comprising daily new cases in every country or even city are generally available. For instance, G. Giordano and coauthors used real data sets obtained during the recent outbreak in Italy to evaluate the outbreak risk which is considerably important for strategic planning against effective disease surveillance.<sup>13</sup> Authors drafted more complicated models rather than using simple human transmission-based approach,<sup>14</sup> that can precisely devise the progress of the recent epidemic based on massive data sets gathered from Italian Civil Protection sources. one more example of data source for epidemic prediction to figure out the possible outbreak regions.<sup>15</sup> Another study used official source in China for the recent pandemic related data like (<http://www.nhc.gov.cn/>) to find the aggregate numbers of Covid patients and

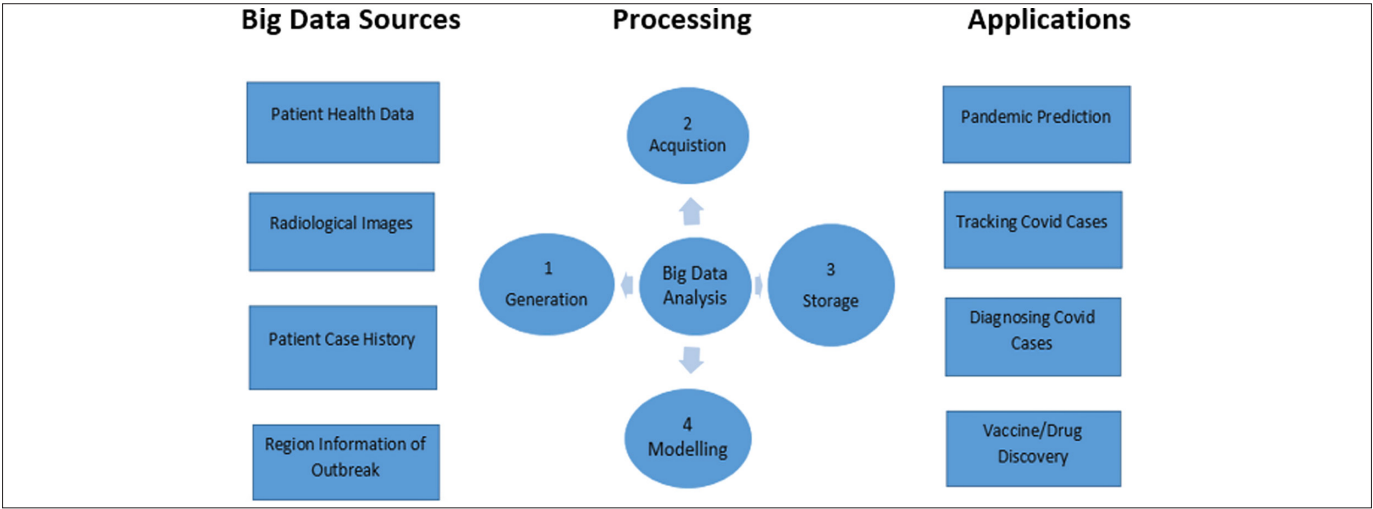
recovered cases in various regions in China.<sup>16</sup> Moreover, Y.-S. Long and co-authors employed government reports collected from China, South Korea, Italy, and Iran with the purpose to estimate the daily infected cases which in long run evaluate the future outbreak risk.<sup>17</sup>

Indeed, these data-processing techniques help in the early prediction of the epidemic as digital application formulated at the John Hopkins University, USA.<sup>18</sup> Furthermore, other unconventional streams of data and social web also have the potential to rebuild early pandemic outbreaks as in one of the study researchers conducted a population-based observational study in China, tracking the health-related social websites and found that unconventional data sets help in understanding the pandemic spread and then designing and implementing the effective health strategies.<sup>19</sup> In the same way, Qin and coworkers<sup>20</sup> employed "Social media search indexes" to find new suspected or confirmed cases of COVID-19 on the basis of symptoms and concluded that new suspected cases of COVID-19 could be detected 6-9 days prior and the confirmed cases 10 days earlier.

#### Big Data Role in Tracking and Diagnosing COVID-19 Cases

One of the major parts of big data is to track the recent pandemic spread, which is of cardinal significance for health care institutions and governmental authorities to monitor successfully the virus outbreak.<sup>21</sup> Several emerging solutions for the recent pandemic by employing big data have been considered to track corona spread. Like in the study,<sup>22</sup> researchers proposed a big data-based approach to track corona spread. Similarly, the comprehensive analytic model and temperature-based model mentioned in the studies<sup>23,24</sup> are extremely helpful for tracking corona spread. Moreover, proposed block-chain technology mentioned in the research studies<sup>25-27</sup> is beneficial in terms of access to healthcare services, enhancing disease control, and manage pandemics by active decision-making.

Big Data sources, their processing, and applications are summarized in Figure 2. Additionally, big data has the ability to support corona diagnosis and treatment besides pandemic prediction and to track the spread. Scientists have successfully introduced an effective molecular kit that is not only helpful to prompt and accurate diagnosis of COVID-19 but also effective in differentiating between coronavirus and SARS-CoV. Moreover, the salivary diagnostic technique is also effective to detect the virus. In addition, other molecular diagnostic tests, such as microarray-based, nucleic acid amplification, and computed tomography (CT)



**Figure 2.** Big data and its applications for fighting COVID-19 pandemic.

scan of the chest are essential to monitor and evaluate the progression of disease in the context of intensity and clinical features. Currently, several researches are being in progress to find radiological indicators on an early basis, which would be extremely valuable in terms of stratifying COVID-19 cases and clinically treat them.<sup>8,19,21</sup>

**Role of Big Data in Developing Effective Treatment Plan and Discovery of Drug/Vaccine**

In terms of COVID-19 treatment, no specific therapeutic plan has approved so far and medical practitioners are providing supportive management to the patients.<sup>28</sup> Recently, researchers have proposed 3 main strategies in this chaotic situation with the aim to fight the corona virus: (i) feasible use of broad-spectrum anti-viral medications, already available; (ii) possible modification /adaptation considering biochemical and biophysical characteristics of the virus; (iii) repositioning of medicine (Western or Chinese).<sup>2</sup>

Big data provides insight for discovery of vaccine/drug against the corona outbreak. For example, S. F. Ahmed and coworkers mentioned

in his study the use of GISAID database ([www.gisaid.org/CoV2020/](http://www.gisaid.org/CoV2020/)) to identify the potential vaccine targets for the novel coronavirus.<sup>29</sup> Similarly, A. Banerjee and coauthors, in their study emphasized on spike proteins with the aim to develop a suitable vaccine.<sup>30</sup> Moreover, M. I. Abdelmageed and co-authors reviewed the online database of the National Center of Biotechnology and use Information immune informatics approach to develop vaccine against corona.<sup>31</sup> Big data also facilitates in drug manufacturing to fight against COVID-19. Scientists proposed molecular docking for drug investigations to fight against corona and for this purpose over 2000 molecules in drug database approved by the United States Food and Drug Administration were screened.<sup>32</sup>

**Challenges and Future Recommendations**

According to Hua and Shaw,<sup>33</sup> the major determinants of success of Chinese authorities in the fight against the recent pandemic of corona is the combination of effective administration, rigorous measures, surveillance, participation of general public and use of big data and advanced methodologies. Big Data has the immense potential to

| Table 1. Summary of Recent Research Articles on Potential Application of Big data for Corona Management |  |              |
|---|--|--------------|
| Potential Application   | Research Studies   | Contribution |
| Pandemic prediction   | [13] This paper proposed a new model that predicts the course of the epidemic to help plan an effective control strategy.  |              |
|   | [16] This paper proposed a generalized SEIR model by including the self-protection and quarantine to analyze this epidemic.  |              |
| Virus tracking  | [22] This paper used the traffic flow to quantify the potential infectious people. An analytical technique using a large record taken from the Chinese national health commission to track the corona spread. The analytical findings of the paper indicated a significant association between the corona-infected individuals and size of the population.   |              |
|   | [23] To evaluate the effectiveness of the containments on the epidemic spreading of the new Coronavirus disease 2019, this paper carried on an analysis of the time evolution of the infection in different Countries, by considering well-known macroscopic growth laws, the Gompertz law, and the logistic law. A proposed analytical method based on a data set taken from 4 countries method to track virus spread tracking. This design can assess the maximum infected cases in a particular region. |              |
|   | [24] This paper suggested temperature-based technique to track COVID cases to assess the association between the number of patients and the mean temperature using a large record taken from several countries. They analyzed a possible bias: poor countries, often located in warm regions, might have less intense testing.   |              |
|   | [25] This paper proposed a low cost Block-chain and AI-linked self-testing model to diagnose Corona cases. The model has the tendency to control the virus spread.   |              |
| Virus diagnosis   |  |              |
|   | [30] In this study the researchers screened few pairs of spike protein epitopic regions and selected their energetic, IC50, MHC II reactivity and found some of those to be very good target for vaccination. A possible role of glycosylation on epitopic region showed profound effects on epitopic recognition. The paper might be helpful for the urgent development of a suitable vaccination regimen against SARS CoV-2.   |              |
|   | [31] This paper study used an immunoinformatics approach to design a multiepitope peptide vaccine against COVID-19 . This method has the potential to sharply screen the sequence and pattern of SARS CoV-2 spike and is useful in the development of vaccines.  |              |
| Vaccine/Drug production   | [32] This paper demonstrated that the FEP-based screening showed remarkable advantages in prompting drug repositioning against COVID-19. A model based on big data was designed using an immune informatics approach to develop a vaccine.   |              |

combat with recent pandemic and other urgencies, and it is expected to increase from now on. It can be useful to track and diagnosis of virus, and in strategic planning against effective disease surveillance. Studies have also revealed the role of big data in developing effective treatment plan and discovery of drug/vaccine.<sup>29-31</sup> Further studies are recommended to probe how to enforce such advanced techniques protecting human rights, meanwhile, simultaneously, assuring great standards.

Summary of recent research articles on potential application of Big data for the ongoing Corona management is mentioned in Table 1.

As mentioned in the preliminary sections, big data has found its significant potential in the fight against novel pandemic. Despite of apparent advantages, there are quiet some challenges required to be considered and addressed in days to come. Furthermore, we point out some results lessons derived from the recent paper and make some recommendations for the research entities.

Among critical challenges, insufficient standard databases are one of the biggest challenges for the purpose to make big data a naive application. As mentioned in studies,<sup>34,35</sup> many platforms of big data have been suggested, but they are not examined utilizing the same database. To tackle this issue, governmental authorities, big companies and healthcare enterprises play a significant role in this regard as they can work together for high standard big databases. They can provide a variety of sources of data from the health units and hospitals in the form of radiological scans of CT and x-ray and various statistics collected from satellite and self-diagnosis.

Furthermore, security issues in terms of sharing personal data is also a great challenge. During the recent pandemic as the officials may ask the patients to share their private information, such as, radiological scans, patients' diagnosis reports, and daily activity to make and implement effective strategies to manage the pandemic situation. Data is essential to ensure the success of big data platform, though, generally people are reluctant to share their private information, unless officially asked. It's a trade-off between personal security and execution. Various possible solutions are accessible, such as, blockchain, federated learning, and incentive techniques to deal with the privacy and security concerns during the recent situation of pandemic that can act as avenues of research in the future.

Moreover, recent literature review explored that big data provides several appropriate solutions to combat the corona situation by helping health care professionals in various health interventions from disease diagnosis, analysis to treatment plan. Few recommendations are deemed in this situation of corona. First, big data-based algorithms should be enhanced more to improve the authenticity and credibility of the data modeling in terms of diagnosis and treatment of corona. Second, big data should be merged with other new advanced technologies such as tools for data analysis, Oracle cloud computing and 5G wireless technologies for the effective measures against pandemic. Last, other non-technological steps like social distancing so far play significant role to curb the disease spread hence must be implemented efficiently under the administration of government bodies with the aim to overcome the recent situation of pandemic in near future.

**Peer-review:** Externally peer-reviewed.

**Author Contributions:** Concept – H.Z.; Design – H.Z.; Supervision – N.M.; Materials – H.Z., N.M.; Data Collection and/or Processing – H.Z., N.M.; Analysis and/or Interpretation – H.Z.; Literature Review – H.Z., N.M.; Writing – H.Z.; Critical Review – H.Z., N.M.

**Declaration of Interests:** The authors have no conflicts of interest to declare.

**Funding:** The authors declared that this study has received no financial support.

## References

- Brüssow H. The novel coronavirus—A snapshot of current knowledge. *Microb Biotechnol.* 2020;13(3):607-612. [CrossRef]
- Perrella A, Carannante N, Berretta M, Rinaldi M, Maturo N, Rinaldi L. Novel coronavirus 2019 (Sars-CoV2): a global emergency that needs new approaches? *Eur Rev Med Pharmacol Sci.* 2020;24(4):2162-2164. [CrossRef]
- Li X, Zai J, Zhao Q, et al. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. *J Med Virol.* 2020;92(6):602-611. [CrossRef]
- Ashour HM, Elkhatib WF, Rahman MM, Elshabrawy HA. Insights into the recent 2019 novel coronavirus (SARS-CoV-2) in light of past human coronavirus outbreaks. *Pathogens.* 2020;9(3):186. [CrossRef]
- White House announces new partnership to unleash U.S. supercomputing resources to fight COVID-19; 2020. Available at: <https://www.whitehouse.gov/briefings-statements/white-house-announces-new-partnership-unleash-u-s-supercomputing-resources-fight-covid-19>. accessed date: March 23, 2020 [online].
- Bragazzi NL, Damiani G, Martini M. From Rheumatology 1.0 to Rheumatology 4.0 and beyond: The contributions of Big Data to the field of rheumatology. *Mediterr J Rheumatol.* 2019;30:3-6.
- Dini G, Bragazzi NL, Montecucco A, Toletone A, Debarbieri N, Durando P. Big Data in occupational medicine: the convergence of -omics sciences, participatory research and e-health. *Med Lav.* 2019;110(2):102-114. [CrossRef]
- Wong ZSY, Zhou J, Zhang Q. Artificial Intelligence for infectious disease Big Data Analytics. *Infect Dis Health.* 2019;24(1):44-48. [CrossRef]
- Gianfredi V, Bragazzi NL, Nucci D, et al. Harnessing big data for communicable tropical and sub-tropical disorders: implications from a systematic review of the literature. *Front Public Health.* 2018;6:90. [CrossRef]
- Bragazzi NL, Gianfredi V, Villarini M, et al. Vaccines meet big data: state-of-the-art and future prospects. From the classical 3Is ("Isolate-Inactivate-Inject") vaccinology 1.0 to vaccinology 3.0, vaccinomics, and beyond: a historical overview. *Front Public Health.* 2018;6:62. [CrossRef]
- Bragazzi NL, Dini G, Toletone A, Brigo F, Durando P. Leveraging big data for exploring occupational diseases-related interest at the level of scientific community, media coverage and novel data streams: the example of silicosis as a pilot study. *PLoS ONE.* 2016;11(11):e0166051. [CrossRef]
- Bragazzi NL, Guglielmi O, Garbarino S. SleepOMICS: how big data can revolutionize sleep science. *Int J Environ Res Public Health.* 2019;16(2):291. [CrossRef]
- Giordano G, Blanchini F, Bruno R, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med.* 2020;26(6):855-860. [CrossRef]
- Brauer F, Castillo-Chavez C, Castillo-Chavez C. *Mathematical Models in Population Biology and Epidemiology.* Berlin: Springer; 2012.
- Chen B, Shi M, Ni X, et al. Visual data analysis and simulation prediction for COVID-19. *Int J Educ Excellence.* 2020;6(1):95-114.
- Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. *medRxiv.* 2020. [CrossRef]
- Long YS, Zhai ZM, Han LL, et al. Quantitative assessment of the role of undocumented infection in the 2019 novel coronavirus (COVID-19) pandemic. *arXiv preprint arXiv:* 2020;2003.12028.
- Pandey G, Chaudhary P, Gupta R, Pal S. SEIR and regression Model based COVID-19 outbreak predictions in India. *arXiv preprint arXiv* 2020;2004.00958.
- Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Digit Health.* 2020;2(4):e201-e208. [CrossRef]
- Qin L, Sun Q, Wang Y, et al. Prediction of number of cases of 2019 Novel coronavirus (COVID-19) using social media search index. *Int J Environ Res Public Health.* 2020;17(7):2365. [CrossRef]
- Ienca M, Vayena E. On the responsible use of digital data to tackle the covid-19 pandemic. *Nat Med.* 2020;26(4):463-464. [CrossRef]
- Zhao X, Liu X, Li X. Tracking the spread of novel coronavirus(2019-ncov) based on big data. *medRxiv.* 2020.
- Castorina P, Iorio A, Lanteri D. Data analysis on coronavirus spreading by macroscopic growth laws. *arXiv Preprint ArXiv:2003.00507.* 2020.
- Notari A. Temperature dependence of COVID-19 transmission. *arXiv preprint arXiv:2003.12417.* 2021;763:144390. [CrossRef]

25. Mashamba-Thompson TP, Crayton ED. Blockchain and artificial intelligence technology for novel, coronavirus disease-19 self-testing. *Diagnostics (Basel)*. 2020;10(4):198. [\[CrossRef\]](#)
26. Bhattacharya S, Singh A, Hossain MM. Strengthening public health surveillance through blockchain technology. *AIMS Public Health*. 2019;6(3):326-333. [\[CrossRef\]](#)
27. Chattu VK, Nanda A, Chattu SK, Kadri SM, Knight AW. The emerging role of Blockchain technology applications in routine disease surveillance systems to strengthen global health security. *Big Data Cogn Comput*. 2019;3(2):25. [\[CrossRef\]](#)
28. Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How big data and artificial intelligence can help better manage the COVID-19 pandemic. *Int J Environ Res Public Health*. 2020;17(9). [\[CrossRef\]](#)
29. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARSCoV-2) based on SARS-CoV immunological studies. *Viruses*. 2020;12(3):254. [\[CrossRef\]](#)
30. Banerjee A, Santra D, Maiti S. Energetics based epitope screening in SARS CoV-2 (COVID 19) spike glycoprotein by immuno-informatic analysis aiming to a suitable vaccine development. *bioRxiv*, 2020.
31. Abdelmageed MI, Abdelmoneim AH, Mustafa MI, et al. Design of a multi-epitope-based peptide vaccine against the E protein of human COVID-19: an Immunoinformatics approach. *BioMed Res Int*. 2020;2020:1-12. [\[CrossRef\]](#)
32. Wu D, Zheng X, Liu R, et al. Free energy perturbation (fep)-guided scaffold hopping. *Acta Pharmaceutica Sinica B*. 2022;12(3):1351-1362.
33. Hua J, Shaw R. Corona virus (COVID-19) “Infodemic” and emerging issues through a data lens: the case of China. *Int J Environ Res Public Health*. 2020;17(7):2309. [\[CrossRef\]](#)
34. Wang B, Kang J, Ma X, et al. A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *medRxiv*, 2020.
35. Ozkaya U, Ozturk S, Barstugan M. Coronavirus (COVID-19) Classification Using Deep Features Fusion and Ranking Technique. *arXivpreprint arXiv:2004.03698*, 2020.