ISSN 2350-109X
www.indianscholar.co.in

Indian Scholar

An International Multidisciplinary Research e-Journal

LEVERAGING DATA SCIENCE TO CURB COVID-19

Soumya Doshi
Savitridevi Hariram Agarwal International School
doshisoumya1@gmail.com

Abstract

The COVID-19 pandemic has hit the global at a colossal scale.Being a highly infectious disease,with worldwide reported positive cases of 182 million, it has led to agrievousimpact on humanity.As all the countries are struggling to alleviate the losses due to the outbreak,enforcing lockdown has become the primary defence mechanism.With researchers working around the clock to find an advancement in the diagnostics and treatment of the pandemic, it has presented global health services with the most appalling challenge.Inspired by latest advances and applications of data science in these areas, this paper aims at highlighting its importance in responding to the COVID-19 outbreak and preventing the severe effects of the COVID-19 pandemic.It also presents few limitations to big data in handling the epidemic.

Keywords: *Pandemic, Covid-19, Big data,Data analysis*

Introduction

This paper first,presents fundamental knowledge of Covid-19(Section-I) and data science (Section-II) and then it reviews the applications of big data in fighting the covid-19 pandemic(Section-IV).For example: Identifying the covid-19 patients,tracking the covid-19 outbreak,developing drug researches and improving the medical treatment.Lastly, a number of limitations of data science applications are outlined and discussed(Section-IV).

Theory

<u>Section I: Covid-19</u>

The world would remember the year 2020 as a devastating year for humanity on this planet earth. Pneumonia of unknown a etiology (novel coronavirus) identified in the city of Wuhan, China in December 2019 [1] with its first mortality reported on January 10, 2020, has become a pandemic [2] .It is named as COVID-19 (Corona virus disease 2019) by the World Health Organization(WHO) [3] and declared as a pandemic by WHO on 11th March 2020 [4].At the time of writing, globally, as of 6:32pm CEST, 29 June 2021, there have been 181,176,715 confirmed cases of Covid-19,including 3,930.496 deaths, reported to WHO[29]

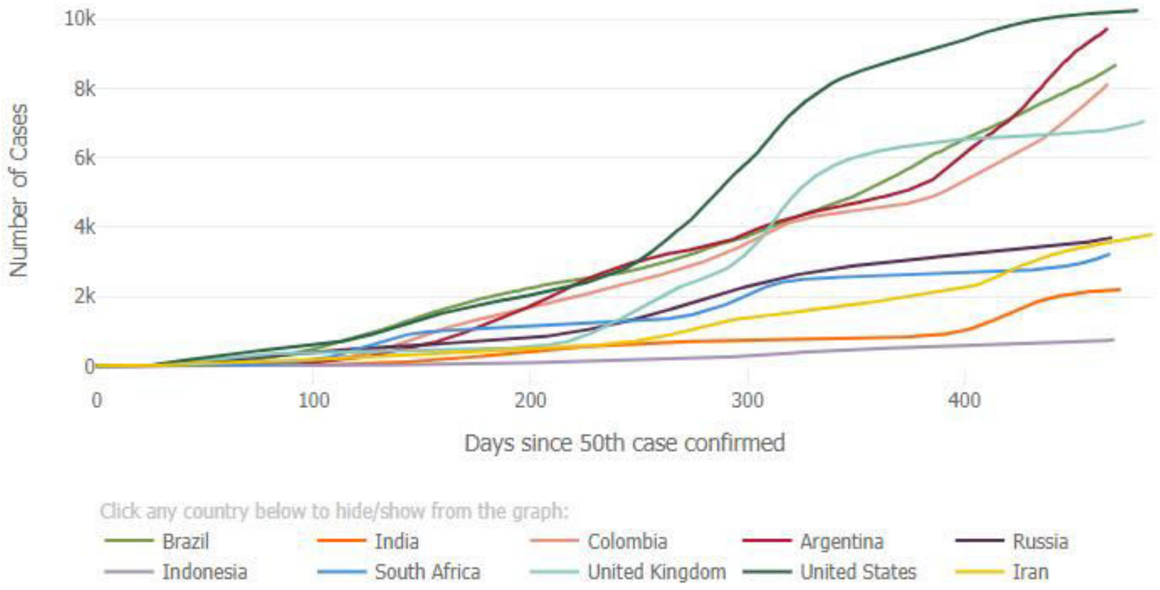

Table 1 -Covid cases [5]

Coronavirus disease-19 (COVID-19), caused by a novel coronavirus, has changed the world notably, not only in the health care space, but also in many aspects of human life such as education, transportation, politics, supply chain, etc. People infected with Corona virus, usually undergo respiratory illness and can recover by taking effective medications. Infected COVID-19 people may undergo respiratory illness but can recover with effective and appropriate treatment methods. What makes COVID-19 much more threateningand easily transmissiblethan other Coronavirus families is that the COVID-19 virus has become highly efficient in human-to-human transmissions[30].Due to the substantial impact of Covid-19 on the globe, abundant efforts are paid as solutions to combat against this 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 handle the outbreak and provide crisis package to decrease effects on the national economics and people, and adapt compatiblepolicies according to the COVID-19 situation. At the same time, individuals are encouraged to stay healthy and protect others by following some advice like wearing the mask at public locations, washing the hands frequently, maintaining the social distancing policy, and reporting the latest symptom information to the regional health center. On the other hand, research and development relevant to COVID-19 are now prioritized, and have received special interest from various stakeholders like governments, industries, and academia. For example, studies in [6], [7] showed tremendous influences of the COVID-19 pandemic on the global supply chain, and took into account different directions of supply chains, including viability, stability, robustness, and resilience.

Over 180 million people across 180+ countries affected and the humungous amount of data requires systematic analysis and trillions of data points to decipher trends to combat COVID-19, which brings forth many opportunities for applying data science techniques [8], hence ameliorating the pressure on health care services.So now let's take a look at how we can use modern technology namely Data science in this regards.

Section II: Data Science

Data science is the discipline that deals with vast volumes of data which use recent tools and techniques to look for camouflaged patterns,extract significant information, and make business decisions. Data science uses complex machine algorithms to build such predictive models.[8].Data science is an umbrella term that encompasses all of the techniques and tools used during the life cycle stages of useful data.

To give a further clarity on understanding this cycle,here is a detailed description of the stages involved in the life cycle of a typical data science project.

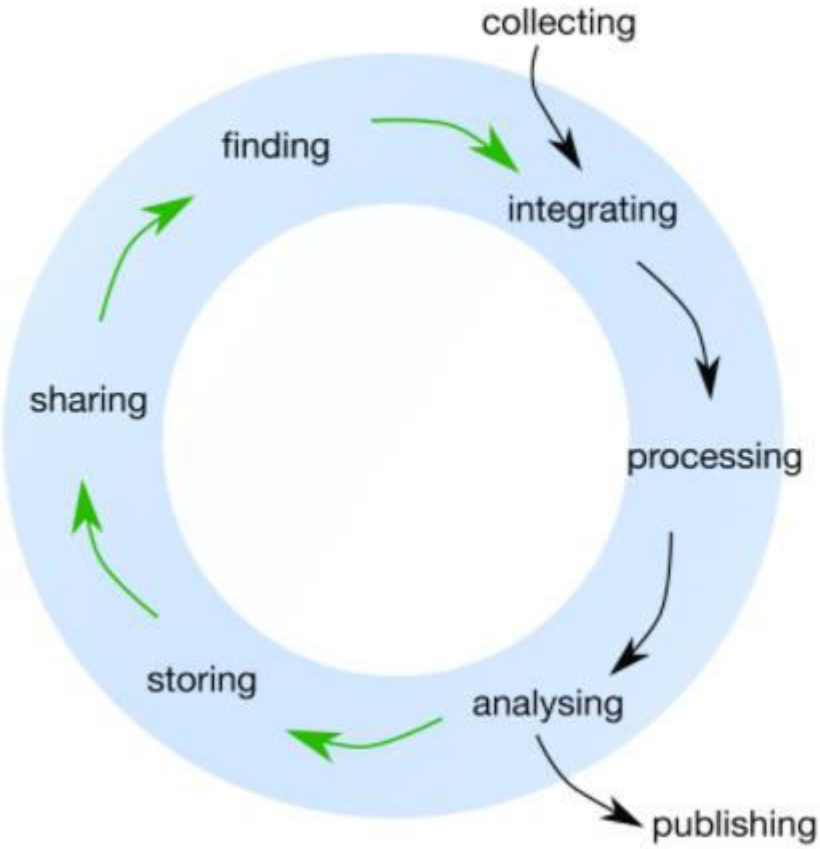

Figure 1:Life cycle of a typical data science project [9]

1) Data generation and collection

First, the cycle starts with collection of data which is generated by people like us.Every search query performed, movie watched, book read, picture taken, message sent, contributes to the massive digital footprint we each generate. After generation comes collection. All the data generated is not collected,maybe out of choice because we do not need or want to, or for practical reasons as the data streams in faster than we can process.

2) Data Processing

After obtaining data,the next immediate thing to do is scrubbing data which includes cleaning,filtering and normalizing.If the data is unfiltered and irrelevant, the analysis result will not be useful.One simple example of normalizing data is reconciling formats of the data.In this process,data is converted from one form to another and everything is consolidated into one standardized formal across the entire data.

3) *Data modeling and interpretation*

Interpreting model and data is final but most crucial step in the cycle.This step helps present the data in a clear and simple way that a human can readily understand and visualize. Modeling data is to reduce the proportion of the data set given.As all values and features are not required for prediction of model,only the relevant ones that contribute to determination of results have to be selected.It is at this stage in the data life cycle when we need to consider, along with functionality, aesthetics, and human visual perception to convey the results of data analysis.

Discussion

In order to combat an epidemic the government has to take decisions such as limiting population movements, allocating scarce resources ,which will play a key role in ensuring the survival and well being of the citizens.One basis for taking these decisions is the availability of the right data..Big data refers to extremely large data sets that require specialized and often innovative technologies and techniques in order to efficiently use the data.

The spread of the global pandemic, COVID-19, has generated tremendous and varied amount of data, which is increasing exponentially. This data can be made useful by leveraging big data analytic techniques in a wide range of areas.

Section III: Data Science applications for Covid-19

Figure 2 shows potential application areas.(next page)

1. Risk assessment and patient prioritization

Healthcare systems around the globe are facing prodigious duress on their resources(e.g availability of intensive care beds,respirators)[31]. This brings about the necessities to immediately access and supervise patient risk,while allocating resources appropriately.In order to recognize the patients at maximum risk for unfavourable outcomes because of care disruptions,health systems can first look for similar care disruption trends in past.Once sufficient data sets have been collected which recognize patients who have experienced health care disruptions in the past, predictive models can then be created.Due to diverse symptoms and disease trajectories, researching technologies for data driven risk-assessment and management in Covid-19 patients would be useful.For example traits like age,gender, or health state can be utilized to provide an estimate of mortality risk.This is particularly important when resources are limited, for example- patient prioritization when intensive care units(ICU) are insufficient.

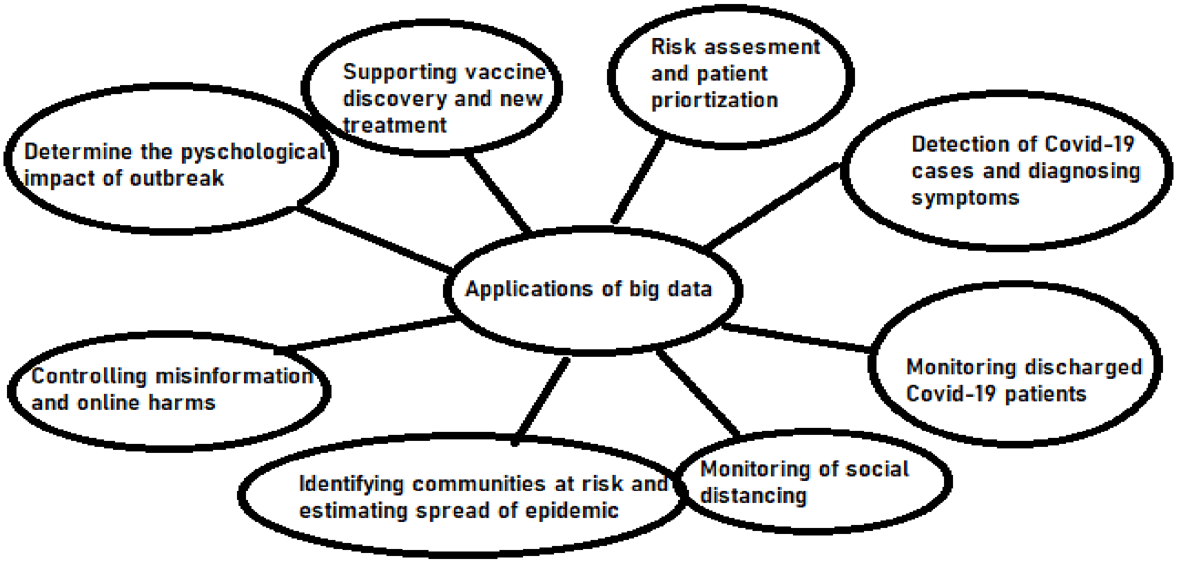
Supporting vaccine discovery and new treatment
Risk assesment and patient priortization
Determine the pyschological impact of outbreak
Detection of Covid-19 cases and diagnosing symptoms
Applications of big data
Controlling misinformation and online harms
Monitoring discharged Covid-19 patients
Identifying communities at risk and estimating spread of epidemic
Monitoring of social distancing

Figure 2

2. Detection of covid -19 cases and diagnosing symptoms

Since majority of infections become evident only upon symptom emergence,ongoing methods for testing are doubtful to identify pre - symptomatic carriers, which is a significant challenge for the implementation of early-stage interventions that reduce transmission.As many as 20% of individuals with COVID-19 are asymptomatic, assisting further viral spread[32]. Some remote computational tools exist which could be expanded, for example- smartwatches can be used for real- time health monitoring and surveillance[33]. An online detection algorithm has been developed which is used to identify early stage of infection by heart rate monitoring [11].Automated tools can further be developed to facilitate screening in larger groups of people(example-airports) by using computer based thermal imaging to detect fever[10].

3. Monitoring discharged covid-19 patients

A study shows deployment of a Remote patient monitoring [RPM] programme with post discharge patients of Covid-19 was associated with a decreased risk of re admission to hospitals, and provided a saleable mechanism to monitor patients in their home environment.The enrolled patients had an app and they self- reported oxygen saturation and temperature daily.And when abnormal symptoms or vital signs were flagged, a pool of nurses assessed the situation.[12]

4. Monitoring of social distancing

This is an no-pharmaceutical intervention adopted by many governments that reduces human contact within the population and hence constraints the spread of Covid-19.Data science can support contact tracing for monitoring of social distancing,for instance by extracting data from social media[13] which can be used for general population tracking to understand compliance with social distancing.This could then be complemented with other datasets(example- cellular trace data or air pollution monitoring)[14] to better understand human mobility patterns in the context of social distancing.However, these solutions present complex trade offs with regards to privacy.

5. *Identifying communities at risk and estimating the spread of epidemic.*

PCCI, which is an organization of data scientists, has launched a vulnerability index as a way to assist community and healthcare leaders to address the factors that cause Covid-19's exponential spread.By analyzing the Geo-spatial distribution of Covid-19 risk factors,PCCI is able to identify communities at risk for Covid-19 allowing for targeted community support and intervention[15]

6. *Controlling the misinformation and online harms*

From suggestions that people can defeat Covid-19 by drinking bleach to deceptive theories that vaccines can alter a person's DNA, the Covid-19 pandemic has made clear the challenges medical misinformation constitutes in this digital age.A study[16] estimates that about 5,800 people were admitted to hospital as a result of false information on social media.Social media platforms are one of the most significant abettors to the spread of misinformation and disinformation, and their algorithms have computed the problem[17]. in order to control this infodemic,classifiers and techniques can be developed to stem this flow.Fondazaine Bruno Kessler(FBC) institute in Italy, uses Twitter data to quantity collective sentiment, social bot pollution, and news reliability and displays this visually[18].

7. *Determine the physiological impact of covid-19 outbreak*

The Covid-19 pandemic and the resulting economic recession have negatively affected many people's mental health and created new barriers for people already suffering from mental illness and substance use disorders.There are a variety of ways pandemic has affected mental health,particularly with widespread social isolation resulting from necessary safety measures[20]

8. *Supporting vaccine discovery and new treatments*

The international effort to discover or re-purpose drug treatments and vaccines can also benefit from extensive data science work predating COVID-19[22].Computational methods can reduce the time spent on examining data, predicting protein structures and genomes[23].It can also assist in identifying eligible patients for clinical trials, which is often a time-consuming and costly part of drug development[24].

Section IV :Key limitations to data science techniques for Covid-19 control

Various challenges may hamper the advantageous outcome from the implementation of big data analysis tools in the health sector that have been confronted while devising solutions in context to the COVID-19 epidemic, which will be discussed in the following subsections.

A. *Data reliability*

The internet and social media are chief sources for circulation of incorrect medical information and rumors,like the effects of virus,impact of vaccine,all of which pose a threat to the government's and health agencies' endeavor to constrict the transmission of virus and maintenance of good health.It is also prone to have unfavourable psychological outcomes on society.Furthermore,absence or inaccuracy of some studies, data may lead to biased study-findings.[28]

B. *Data sharing*

Data sharing plays a crucial role for digital governments and smart cities in tackling such massive public emergencies.In 2015,China's state council issued the action to promote the development of big data[26], which suggested that big data should be used as an essential means to amplify the government's governance range,ameliorate the level of government decision making, and handle risk prevention through well-planned collection and integration of government and social media data.However in practice, data sharing has exhibited a silo effect.Data sharing among Chinese departments and regional governments remains insufficient, and the phenomenon of "block islands" still exists.For example, using private travel and health condition information,a two-dimensional risk assessment code is generated.And when these health codes were used, many regions did not identify the assessment codes generated by other regions.

C. *Data security,privacy and ethics*

Coming up with solutions that showcase reasonable results and at the same time protect privacy stick to sophisticated ethical guidelines is an underrated obstacle.Medical data is confidential and shared only under explicit circumstances and for definite research purposes,as healthcare data security and patient privacy issues are a matter of concern[25].Hence, it is vital to state the working,strategies, and guidelines that govern and ease access to medical data with trade-offs to patients' privacy or without exploiting the data for inappropriate uses,specially when grave situations occur and with the transmission of deadly epidemics that need immediate solutions,such as Covid-19.

Conclusion

The world-wide epidemic of Covid-19 has had an unforeseen impact on the entire human race.This paper not only gives a explains how leveraging data science techniques can ease the pandemic pressure but also outlines the its limitation with regards to this situation and hence how to use technology with it's limitation.With today's relatively advanced data science, if we use big data techniques to advance a rapid, accurate and timely grasp of the development trend of the epidemic or to predict the response of people and take more effective preventive measures,we can expect to reduce the negative impact of epidemics on people's production and life.

References

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, *et al.*
Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China
Lancet, 395 (10223) (2020), pp. 497-506
[2]C. Sohrabi, Z. Alsafi, N. OfiNeill, M. Khan, A. Kerwan, A. Al-Jabir, *et al.*

World health organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19)
Int J Surg (2020)

[3] Organization W.H., et al. Naming the coronavirus disease (COVID-19) and the virus that causes it. 2020a.

[4] WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020

[5] https://coronavirus.jhu.edu/data/cumulative-cases

[6] D. Ivanov and A. Dolgui, "Viability of intertwined supply networks: Extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak", *Int. J. Prod. Res.*, vol. 58, no. 10, pp. 2904-2915, May 2020.

[7] Ivanov, "Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case", *Transp. Res. E Logistics Transp. Rev.*, vol. 136, Apr. 2020.

[8] Wing, J. M. (2019). The Data Life Cycle. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.e26845b4

[9] Philippa C. Griffin, Jyoti Khadake, Kate S. LeMay, Suzanna E. Lewis, Sandra Orchard, Andrew Pask, Bernard Pope, Ute Roessner, Keith Russell, Torsten Seemann, Andrew Treloar, Sonika Tyagi, Jeffrey H. Christiansen, Saravanan Dayalan, Simon Gladman, Sandra B. Hangartner, Helen L. Hayden, William W.H. Ho, Gabriel Keeble-Gagnère, Pasi K. Korhonen, Peter Neish, Priscilla R. Prestes, Mark F. Richardson, Nathan S. Watson-Haigh, Kelly L. Wyres, Neil D. Young, Maria Victoria Schneider
Version 2. F1000Res. 2017; 6: 1618. Published online 2018 Jun 4. doi: 10.12688/f1000research.12344.1 PMCID:PMC6069748

[9] https://www.cnbc.com/2020/04/02/this-smart-thermometer-could-help-detect-covid-19-hot-spots.html

[11] Mishra, T.; Wang, M.; Metwally, A.A.; Bogu, G.K.; Brooks, A.W.; Bahmani, A.; Alavi, A.; Celli, A.; Higgs, E.; Dagan-Rosenfeld, O.; et al. Pre-Symptomatic Detection of COVID-19 from Smartwatch Data. Nat. Biomed. Eng. 2020, 4, 1208–1220. [CrossRef]

[12]Gordon WJ, Henderson D, DeSharone A, Fisher HN, Judge J, Levine DM, MacLean L, Sousa D, Su MY, Boxer R. Remote Patient Monitoring Program for Hospital Discharged COVID-19 Patients. Appl Clin Inform. 2020 Oct;11(5):792-801. doi: 10.1055/s-0040-1721039. Epub 2020 Nov 25. PMID: 33241547; PMCID: PMC7688410.

[13]A. Signorini, A. M. Segre and P. M. Polgreen, "The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic", *PloS One*, vol. 6, no. 5, 2011.

[14]M. Cadotte, "Early evidence that COVID-19 government policies reduce urban air pollution", Mar. 2020.

[15]https://pccinnovation.org/pccis-vulnerability-index-taking-the-fight-to-covid-19/

[16]https://www.ajtmh.org/view/journals/tpmd/103/4/article-p1621.xml

[17]Pennycook G, et al. Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. Psychol Sci. 2020;31(7):770–80.

[18]https://www.fbk.eu/en/press-releases/the-full-version-of-the-of-the-fbk-comune-lab-infodemic-observatory-on-covid19-is-now-released/

[20]Yamada, Y., Ćepulić, DB., Coll-Martín, T. *et al.* COVIDiSTRESS Global Survey dataset on psychological and behavioural consequences of the COVID-19 outbreak. *Sci Data* **8,** 3 (2021). https://doi.org/10.1038/s41597-020-00784-9

[21] Leander, P., Kreienkamp, J., Agostini, M., & PsyCorona Collaboration (2020). Mapping the Moods of COVID-19: Global Study Uses Data Visualization to Track Psychological Responses, Identify Targets for Intervention. *APS Observer*, *2020*(September). https://www.psychologicalscience.org/observer/covid-19-psycorona-global-psychological-response

[22]Mitchell, JBO 2018, 'Artificial intelligence in pharmaceutical research and development', *Future Medicinal Chemistry*, vol. 10, no. 13, pp. 1529-1531. **https://doi.org/10.4155/fmc-2018-0158**

[23]Paul, Debleena et al. "Artificial intelligence in drug discovery and development." *Drug discovery today* vol. 26,1 (2021): 80-93. doi:10.1016/j.drudis.2020.10.010

[24]Lee, J Jack, and Caleb T Chu. "Bayesian clinical trials in action." *Statistics in medicine* vol. 31,25 (2012): 2955-72. doi:10.1002/sim.5404

[25]Karim Abouelmehdi, Abderrahim Beni-Hssane, Hayat Khaloufi, Mostafa Saadi,Big data security and privacy in healthcare: A Review,Procedia,Computer Science,Volume 113,2017,Pages 73-80,ISSN 1877-0509,https://doi.org/10.1016/j.procs.2017.08.292.

[26]The State Council released the "Outline of Action to Promote the Development of Big Data". China Paper NewsLetter. 2015. URL: http://yuxiqbs.cqvip.com/Qikan/Article/detail?id=666542015&from=Qikan_Search_Index

[27]The big data behind the health code is revealed. Today Science and Technology. 2020. URL: http://qikan.cqvip.com/Qikan/Article/Detail?id=7101192711

[28]Richardson, S.; Hirsch, J.S.; Narasimhan, M.; Crawford, J.M.; McGinn, T.; Davidson, K.W.; The Northwell COVID-19 Research Consortium. Presenting Characteristics, Comorbidities, and Outcomes among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020, 323, 2052–2059. [CrossRef]

[29]https://covid19.who.int/ [accessed on 29 June 2021 6:32 pm CEST.]

[30]Elrashdy F, Redwan EM, Uversky VN. Why COVID-19 Transmission Is More Efficient and Aggressive Than Viral Transmission in Previous Coronavirus Epidemics?. *Biomolecules*. 2020;10(9):1312. Published 2020 Sep 11. doi:10.3390/biom10091312

[31]Alexander T Janke, Hao Mei, Craig Rothenberg, Robert D Becher, Zhenqiu Lin, Arjun K Venkatesh. Analysis of Hospital Resource Availability and COVID-19 Mortality Across the United States. *Journal of Hospital Medicine*, Jan. 20, 2021; DOI: 10.12788/jhm.3539

[32]He, Jingjing et al. "Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis." *Journal of medical virology* vol. 93,2 (2021): 820-830. doi:10.1002/jmv.26326

[33] Blaine Reeder, Alexandria David,Health at hand: A systematic review of smart watch uses for health and wellness,Journal of Biomedical Informatics,Volume 63,2016,Pages 269-276,ISSN 1532-0464,