

PREDICTIVE ANALYTICS DATA MODEL FOR FORECASTING COVID-19 RISKS AND OUTCOMES

^[1] SrinathVallakirthy

^[1] Healthcare Data IT Futurist, Texas Health Resources, Dallas, TX, United States

Abstract: In a short span, coronavirus has spread to more than 170 countries across the globe. Most of the hospitals and medical supporting equipment get overwhelmed due to the rapid increase in the number of coronavirus patients. In this work, we analyze the recovery rate of coronavirus patients using machine-learning techniques based on COVID-19 data. An exploratory data analysis of the increase in the number of recovery cases in the different geographical region across the globe is conducted. We observe the effect of initial lockdown across the globe in slowing down the spread of coronavirus. However, later there is a sudden spread of coronavirus in all countries, leading to an increased load on medical staff and lack of adequate medical supporting equipment in the absence of any antiviral or vaccine of coronavirus till date. We notice that there is a decrease in the recovery rate of coronavirus patients to 2% (approximately) after a widespread of the virus in June 2020. However, in July 2020, there is a hike in recovery rate of coronavirus patients, about 30% - 40% in comparison to the infected cases. The key findings of this study can help in understanding the trend in coronavirus spread across the globe and its recovery rate. It can help by providing a piece of valuable information to healthcare authorities and workers to design appropriate strategies for reducing the death toll and understanding the recovery rate in different regions of the world.

Index Terms— COVID-19, Machine-learning techniques, coronavirus.

1. INTRODUCTION

COVID-19 has resulted in high mortality rates since its appearance in December 2019. It is considered as an extremely infectious respiratory disease. The number of COVID-19 cases is increasing exponentially, leading to overwhelming treatment centres by patients over the globe. COVID-19 was stated as atypical pneumonia at Wuhan, China in December 2019. Soon, it was detected as novel coronavirus. World Health Organization (WHO) recognized the coronavirus outbreak as a global pandemic on March 11, 2020. COVID-19 has been observed as a highly infectious disease with a high mortality rate and quick community spread in comparison to its predecessor diseases such as MERS [2]. Over a few days, coronavirus is passed to most of the countries in the world [3]. Upto August 5 2020, more than 18,720,558 cases have been confirmed, 704,645 deaths have been reported, and 11,936,545 patients have recovered from this disease [4]. It has been initially noted that mortality rate changes from one country to another due to variation in age, demography and health care resources. Therefore, different countries reported a different mortality rate. For example, China conveyed a total 2.3% mortality rate

among COVID-19 patients. However, a higher rate was observed among patients with age more than 80 years [5]. In Italy, the overall 5% mortality rate has been observed, whereas, among senior patients, it is about 20% mortality rate [6]. An abrupt increase has been observed in significant coronavirus epicenters patients, resulting an overwhelm healthcare points, hospitals and medical workers [7].

A comprehensive analysis of COVID-19 data can provide a piece of valuable information to healthcare authorities and workers to design appropriate strategies for reducing the death toll and understanding recovery rate in different regions of the world.

In this paper, we attempted to track COVID-19 situation and find some statistical and predicting outcomes for understanding and analyzing COVID-19 pandemic. We mainly focus on implementing machine learning techniques for analyzing and predicting recovery cases of COVID-19 from all over the world based on the previous COVID-data available at Johns Hopkins University (JHU) Github [8].

Rest of this paper is organized as follows: Section 2 presents the relevant literature in analyzing coronavirus patient data. Section 3 describes COVID-19 data and

setup for conducting experiments in this work. Section 4 presents and discusses the analysis results of the recovery rate using COVID-19 data. Finally, Section 5 concludes the paper at the end.

2. RELATED WORK

Since the declaration of COVID-19 as a global pandemic by WHO, many governments of different countries are imposing multiple lockdowns for preventing the spread of this virus. Governments have lockdown their nations in different phases except for a few essential services over the globe [9]. Such initiatives of lockdown have resulted in reducing the coronavirus patients. For example, there is a reduction of coronavirus patients from 11.8% to 6.3% on an everyday basis in India [10]. These lockdowns have caused a drastic fall in the economy of all most all countries. Another practical solution to prevent coronavirus is to quarantine patients of critical zones to limit the spread to that particular zone. However, most quarantine centres are also overwhelmed due to the very rapid spread of coronavirus in most of the countries. WHO has also confirmed that there is no antiviral and vaccine readily available to fight against this coronavirus [11]. Researchers are attempting very hard to find anti coronavirus. However, it seems to take time for producing such medicine [12]. Governments and medical staff is making every possible effort to moderate spread of coronavirus and prepare medical equipment for supporting increasing coronavirus patient load. Predicting the number of new coronavirus patients and recovered patients can help to plan medical supporting inventory to some extent. Few papers have been published in recent days on predicting coronavirus patients as below.

Wang et al. [13] suggested a Patient Information Based Algorithm (PIBA) to estimate the death toll due to coronavirus in China. They reported that in Hubei and Wuhan death rate was about predicted 13%. At the same time, the death rate lies in 0.75% to 3% for the rest of China.

Gupta et al. [14] analyzed coronavirus spread in the US and reported a direct relationship between COVID-19 patients and temperature. They expect an abrupt decrease in coronavirus patients in the summer season. But, this does not happen as expected.

Ahmar and Val [15] employed ARIMA and Sutte ARIMA to forecast COVID-19 patients over a short span

of Spanish stock market. The authors reported a prediction with MAPE of 3.6% till April 16, 2020.

Ceylan [16] also employed ARIMA models to predict the number of positive patients in Italy, Spain and France. MAPE of 4% to 6% has been reported. Fanelli and Piazza in [17] predicted COVID-19 cases in Italy, France and China. The authors predicted ventilation units required in Italy in their study.

Most studies cited above attempted to predict the number of coronavirus patients. In this work, we analyzed COVID-19 data over the globe and analyzed the recovery rate of coronavirus patients in different regions at a different time in the world. The findings of this study enable a better understanding of COVID-19 data and plan better strategies for decreasing the increased load of coronavirus patients on medical staff and developing medical supporting equipment required for treating coronavirus patients.

3. Covid-19 dataset and experimental setup

For better understanding the trends of coronavirus patients and their recovery rate in different countries of the world, we collected data from Johns Hopkins University (JHU) Github [8]. The people at Johns Hopkins University (JHU) Github collect data from various international research organizations like WHO, ECDC, various countries government bodies, medical and health care institutes all over the globe.

We collected time series based recovery dataset of COVID-19 recovered cases from January 22, 2020 to June 13, 2020. For performing a comprehensive analysis recovery statistics, we used Python language for developing analysis modules using machine-learning techniques. We used different machine learning libraries, visualization libraries and data processing libraries such as Pandas, NumPy, seaborn, Matplotlib for analyzing COVID-19 data. Experiments have been conducted using Google Colaboratory for this work.

4. Results and discussion

In this work, we analyzed the recovery rate of coronavirus patients in different perspectives for different countries using COVID-19 data. To analyze the space and trace of data, we interpreted the number of recovered patients and exact numbers based on latitude and longitude of different regions, as shown in Figure 1.

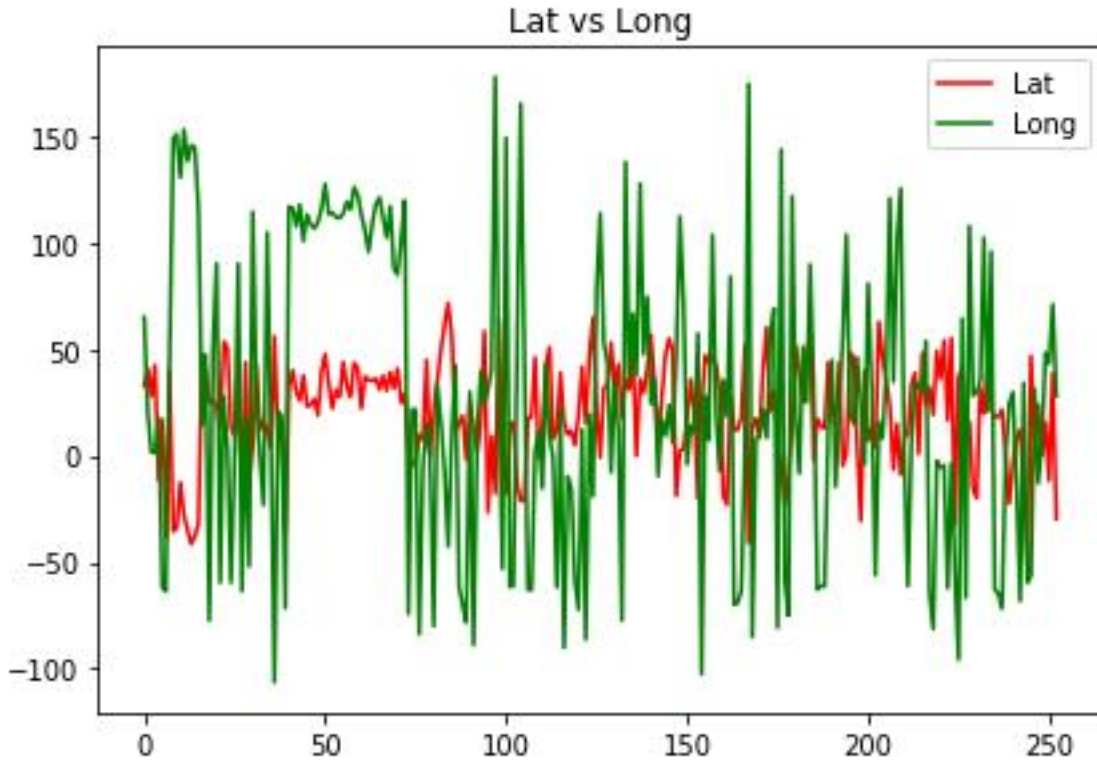


Figure 1 Number of recovered patients on the basis of latitude and longitude

The effect of coronavirus disease recovery rate is latitude and longitude dimension from all over the world. represented in Figure 2, indicating a scatter plot using

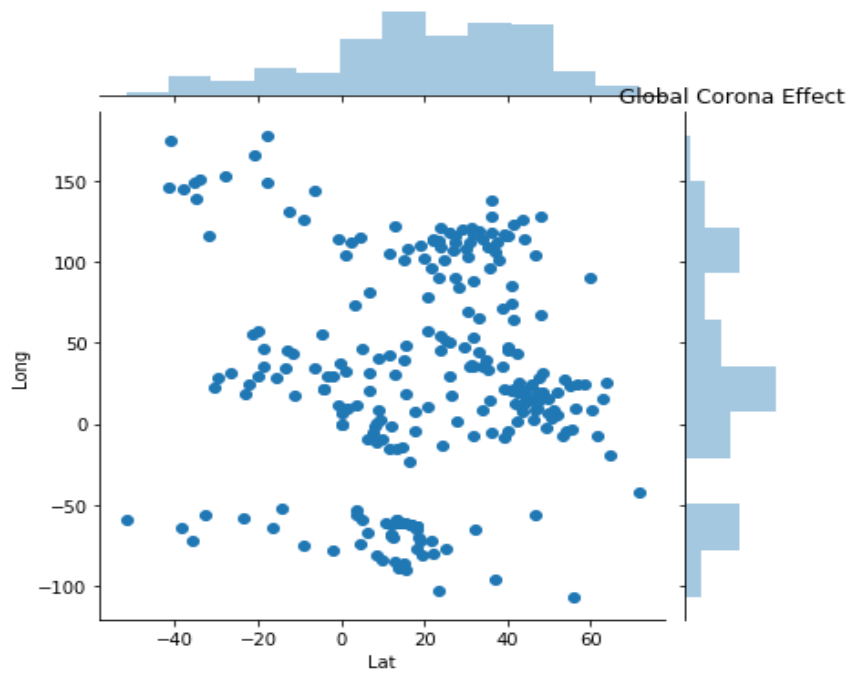


Figure 2 Global corona effect

It can be observed from Figure 2 that the number of recovered cases in some geographical areas of the world is high, as there exist dense clusters in comparison to the number of infected cases. Whereas, there are some dispersed data points in Figure 2, indicating a low recovery rate of coronavirus patients in comparison to the number of infected cases. Figure 2 demonstrates that distribution of recovery rate of coronavirus patients for

particular geographical region relies upon the mentality of citizens, food, health and innovative culture, healthcare and on-ground implementation.

We also analyzed the recovery rate of coronavirus patients in different countries on a given date from all months. Our analysis finds that there is a sudden growth and fall in recovery cases for different countries, as shown in Figure 3.

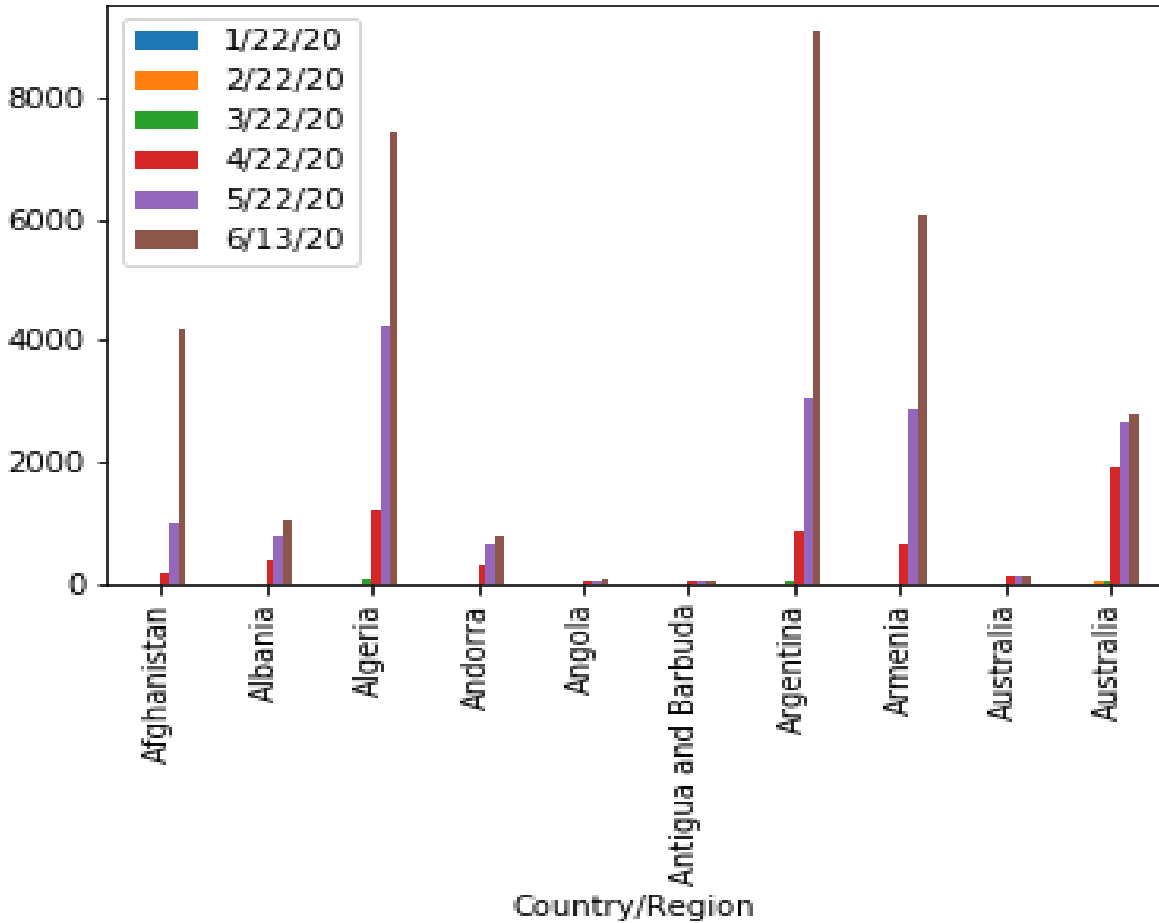


Figure 3 Recovery rate trend of coronavirus patients in different countries

We analyzed the recovery rate of coronavirus patient in the beginning days (Jan 2020) and last span (June 2020) considered in this study. We noticed that during the initial stages, few people are being affected by the coronavirus and the recovery rate was about 40% as presented in

Figure 4. However, with widespread of coronavirus across the world, the recovery rate of coronavirus patient's decreased abruptly to less than 2% as shown in Figure 5.

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 7, Issue 8, August 2020

However, a hike in recovery rate of coronavirus patient is observed in different countries in July 2020, leading to a recovery rate of 30% to 40% as presented in Figure 6.

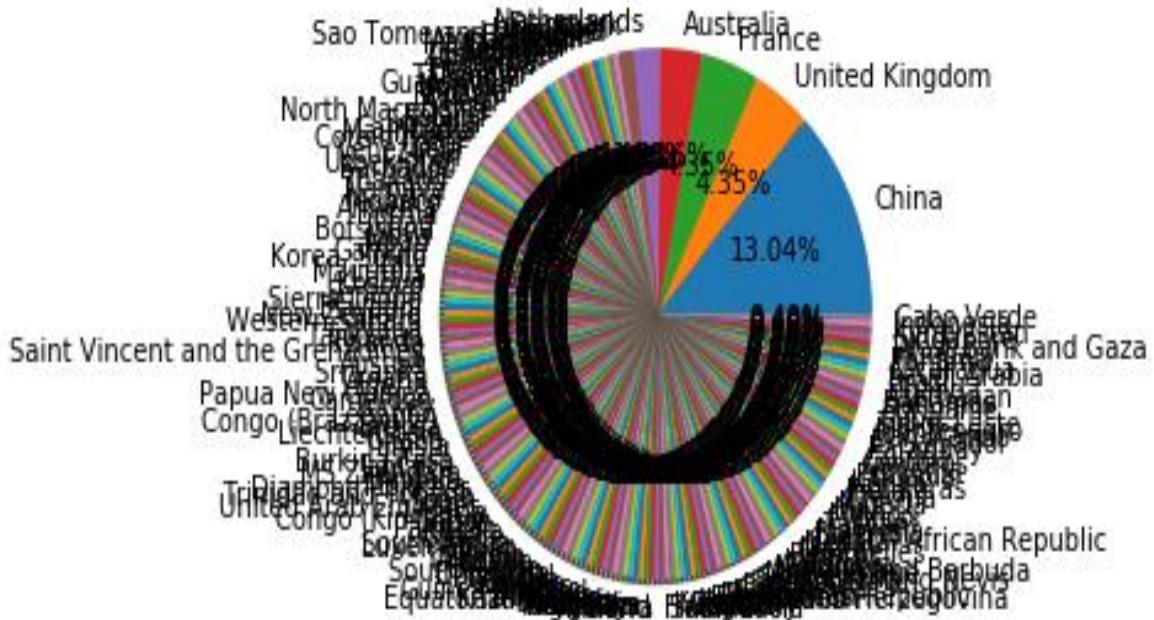


Figure 6 Recovery rate of corona patients in July 2020

It has also noticed that in May and June 2020, the recovery rate of coronavirus patients has increased for some countries whereas decreased for some other countries as depicted in Figure 7.

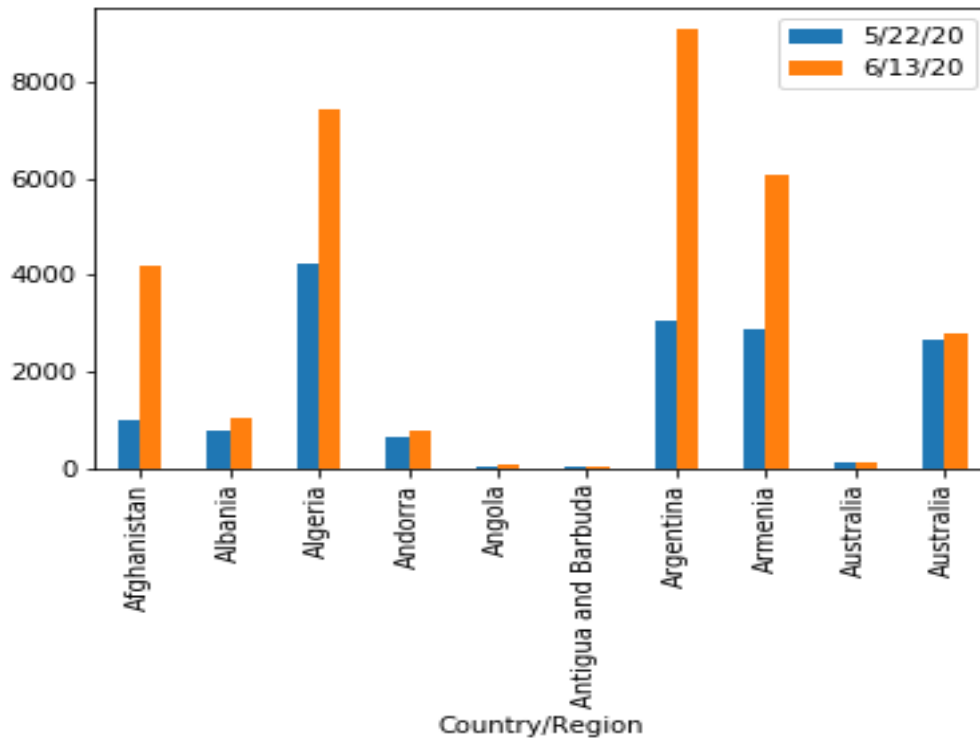


Figure 7 recovered cases of coronavirus patients in May and June 2020

**International Journal of Engineering Research in Computer Science and Engineering
(IJERCSE)**

Vol 7, Issue 8, August 2020

This indicates that the number of recovery case of coronavirus patients is increasing in comparison to infected cases now. Consequently, key findings of this study help in understanding trends in recovery cases of coronavirus patients that helps to plan strategies for providing adequate medical supporting equipment and reduces the increased load of medical staff.

5. Conclusion

In this paper, we have analyzed the recovery rate of coronavirus patients using machine-learning techniques based on COVID-19 data. An exploratory data analysis of the increase in the number of recovery cases in the different geographical region across the globe has been done. The recovery rate of coronavirus patients is presented for different countries in comparison to the infected cases. It has been observed that initial lockdown across the globe has resulted in slow down the spread of coronavirus. But, later there is a sudden spread of coronavirus in all countries, leading to an increased load on medical staff and lack of adequate medical supporting equipment in the absence of any antiviral or vaccine of coronavirus till date. We noticed that there is a decrease in the recovery rate of coronavirus patients to 2% (Approximately) after widespread of coronavirus in June 2020. However, in July 2020, there is a hike in recovery rate of coronavirus patients, about 30% - 40% in comparison to the infected cases. Therefore, the findings of this study can help in understanding the trend in coronavirus spread across the globe and its recovery rate. It can help by providing a piece of valuable information to healthcare authorities and workers to design appropriate strategies for reducing the death toll and understanding the recovery rate of coronavirus patients in different regions of the world.

References

[1] Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S., Lau, E.H., Wong, J.Y., and Xing, X. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* 382, 1199–1207.

[2] Mahase, E. (2020). Coronavirus: Covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *BMJ* 368, m641.

[3] World Health Organization (2020). Coronavirus Disease 2019 (COVID-19): Situation Report, 61. <https://apps.who.int/iris/handle/10665/331605?show=full>.

[4] Coronavirus Update (Live),

<https://www.worldometers.info/coronavirus/>
Last accessed on Aug 05, 2020

[5]

[6] Wu, Z., and McGoogan, J.M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 323, 1239–1242.

[7] Livingston, E., and Bucher, K. (2020). Coronavirus disease 2019 (COVID19) in Italy. *JAMA* 323, 1335.

[8] Capricornus EinBlick, COVID-19 Time Series Analysis, <https://medium.com/@thecapricornuseinblick/covid-19-time-series-analysis-e6f3f2235e43> last accessed Aug. 05, 2020.

[9] COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, <https://github.com/CSSEGISandData/COVID-19> last accessed on Aug 05, 2020.

[10] Arora, P., Kumar, H., & Panigrahi, B. K. (2020). Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, Solitons & Fractals*, 110017.

[11] Mohfw/home.2020. <https://www.mohfw.gov.in/> Last accessed on Aug, 05, 2020).

[12] Organization W.H. et al. Q& A on coronaviruses. 2020b.

[13] Grenfell R., Drew T.. Here's why it's taking so long to develop a vaccine for the new coronavirus. *Science Alert* Archived from the original on 28 2020.

[14] Wang L , Li J , Guo S , Xie N , Yao L , Cao Y , et al. Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm. *Sci Total Environ* 2020:138394 .

[15] Gupta S ,Raghuwanshi GS , Chanda A . Effect of weather on COVID-19 spread in the US: a prediction model for India in 2020. *Sci Total Environ* 2020:138860.

[16] AhmarAS , del Val EB . SutteARIMA: short-term forecasting method, a case: COVID-19 and stock market in Spain. *Sci Total Environ* 2020:138883.

[17] CeylanZ . Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci Total Environ*

2020:138817.

- [18] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals 2020;134:109761.