



The COVID-19 pandemic: prediction study based on machine learning models

Zohair Malki¹ · El-Sayed Atlam^{1,2} · Ashraf Ewis^{3,4} · Guesh Dagnew⁵ · Osama A. Ghoneim⁶ · Abdallah A. Mohamed^{1,7} · Mohamed M. Abdel-Daim^{8,9} · Ibrahim Gad²

Received: 27 May 2020 / Accepted: 5 April 2021 / Published online: 10 April 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

COVID-19 was first discovered in Wuhan, China in December 2019. It is one of the worst pandemics in human history. Recent studies reported that COVID-19 is transmitted among humans by droplet infection or direct contact. COVID-19 pandemic has invaded more than 210 countries around the world and as of February 18th, 2021, just after a year has passed, a total of 110,533,973 confirmed cases of COVID-19 were reported and its death toll reached about 2,443,091. COVID-19 is a new member of the family of corona viruses, its nature, behaviour, transmission, spread, prevention, and treatment are to be investigated. Generally, a huge amount of data is accumulating regarding the COVID-19 pandemic, which makes hot research topics for machine learning researchers. However, the panicked world's population is asking when the COVID-19 will be over? This study considered machine learning approaches to predict the spread of the COVID-19 in many countries. The experimental results of the proposed model showed that the overall R2 is 0.99 from the perspective of confirmed cases. A machine learning model has been developed to predict the estimation of the spread of the COVID-19 infection in many countries and the expected period after which the virus can be stopped. Globally, our results forecasted that the COVID-19 infections will greatly decline during the first week of September 2021 when it will be going to an end shortly afterward.

Keywords Machine learning model · COVID-19 pandemic · Artificial intelligence · Prediction

Introduction

COVID-19 was first discovered in Wuhan, China in December 2019. The World Health Organization (WHO) later declared the new emerging disease as a pandemic (Huang et al. 2020). Recent studies reported that COVID-19 is transmitted among humans by droplet infection or direct contact (Lai et al., 2020a).

The WHO has specified that the main human-to-human transmission mechanism varies, but still can be generalized as direct contact with an infected person through shaking hands, exposure to droplets coming out during coughing or sneezing, and by traveling to an affected area and attaining the virus in one or other way. The core symptoms of COVID-19 highly vary, ranging from being severely affected to being asymptomatic and the infected individuals can experience

Responsible Editor: Philippe Garrigues

✉ El-Sayed Atlam
satlam@yahoo.com

- ¹ College of Computer Science and Engineering, Taibah University, Yanbu, Saudi Arabia
- ² Faculty of Science, Tanta University, Tanta, Egypt
- ³ Department of Public Health and Occupational Medicine, Faculty of Medicine, Minia University, Minia, Egypt
- ⁴ Department of Public Health, Faculty of Health Sciences – AlQunfudah, Umm AlQura University, Meccah, Saudi Arabia

- ⁵ Department of Computer Science, Institute of Technology, Dire Dawa University, Dire Dawa, Ethiopia
- ⁶ Faculty of Computers and Informatics, Tanta University, Tanta, Egypt
- ⁷ Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Shibin El Kom, Egypt
- ⁸ Department of Zoology, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia
- ⁹ Pharmacology Department, Faculty of Veterinary Medicine, Suez Canal University, Ismailia 41522, Egypt

from mild to very severe respiratory illnesses. High fever, cough, sore throat and muscular pain were the primary symptoms in most of the symptomatic cases. However, severe cases suffer from pneumonia, micro-coagulopathies, and septic shock. Rapid clinical deterioration of the cases can lead to death (Qiu et al. 2020; Wu et al. 2020). Mostly, old-aged people and those who have pre-existing medical conditions e.g., diabetes mellitus, chronic respiratory disease, or cancer are more likely to experience manifestations and consequences of COVID-19 infection World Health Organization (WHO) (2020).

As of February 18th, 2021, a total of 110,533,973 confirmed cases of COVID-19 were reported and its death toll reached about 2,443,091 (Worldometer 2020). However, the available information about COVID-19 is being built up and its nature and characteristics are being discovered especially, its very quick ability to change its nature evolving new variants based on its accelerated genetic mutations. Therefore, thoroughgoing observational studies are being performed to establish facts about COVID-19 to find out treatment or a vaccine that may help in ending its pandemic (Yang et al. 2020).

Many research studies on COVID-19 are published and others are on the lane, and floods of huge data about it are constantly accumulating, without reaching a strong prediction about the transmission and end of the pandemic (Yang et al. 2020). In our current study, we deployed machine learning approaches for predicting the spread of the virus in several selected countries. Yet, the same approach can be applied for predicting the spread of COVID-19 infection in any other country, since the nature of the virus is nearly the same everywhere.

This study has the following major contributions:

- It presents the machine learning model as a method for predicting the transmission of COVID-19 pandemic in an easily understandable way using statistical visualization graphs e.g., normal distribution.
- It determines the predictive value of the technique with quality and density of collected data of WHO.
- It provides the governments and health authorities with the required information that helps in planning and decision-making. The resulting predictions will reduce the population's anxiety and prepares their mentality for accepting and dealing with the next phases of the pandemic.

The current paper's organization is: Related work is presented in the “[Related work](#)” section. Methodology and proposed approaches are presented in the “[Methodology](#)” section. Then the experimental observations and the discussion are presented in “[Experimental evaluation](#)” and “[Discussion](#)” sections, respectively. The paper ends up with conclusions and future work in the “[Conclusion](#)” section.

Related work

Machine learning and Artificial Intelligence (AI) models are essentially used to improve the prediction accuracy of diagnosis and the screening of non-infectious diseases. Moreover, machine learning approaches are also widely used in the analysis and prediction of COVID-19 survival rate, the discharge time of patients based on clinical data. Lai et al. (2020a, b, c) considered the scourge idea of COVID-19 in regard to the every day aggregate list, death rate, and cooperative status of the countries' healthcare and economy. Punn et al. (2020) have proposed the utilization of machine learning and deep learning models to understand the COVID-19 pandemic based on the data taken from the Johns Hopkins dashboard.

Dandekar and Barbastathis (2020) recommended a mixture model that comprises of first-standards epidemiological conditions and an information-driven neural organization to gauge the stopping of the transmission of the COVID-19 infection. They used a neural network model to predict for four locations namely Wuhan city, Italy, South Korea, and the USA. Finally, for the USA, they predicted the currently infected growth curve and predicted a halting of infection by April 20th, 2020.

The WHO rules for the anticipation of the COVID-19 infection showed that it enters the human body through the eyes, nose, or mouth. Along these lines, it gave a few prescribed insurances on the whole settings to avoid getting the infection, for example, trying not to contact the face with unwashed hands, washing hands with soap and water for at least 20 s, or cleaning hands completely with gels, or tissues. It likewise suggested physical distancing of at least one and a half meter or even working from home can diminish the danger of contamination World Health Organization (WHO) (2020).

Former studies developed methods to achieve accurate and time-efficient predictions of the transmission of COVID-19. However, these studies lack some promising features that are mainly related to their low accurate predictive results and lacking the promising features that enable the prediction of the highest possible accuracy of the confirmed cases with COVID-19.

Methodology

In the following subsections, we described the measures used to slow the spread of disease as in “[Measures used to slow the spread of COVID-19](#)” subsection, datasets used to validate the proposed method in “[Dataset description](#)” subsection and the description of our proposed method is presented in details in “[The predictive machine learning model](#)” subsection.

Measures used to slow the spread of COVID-19

The COVID-19 pandemic is rapidly spreading all over the world, while there is no clear picture of how and why the virus is spreading among the people and involves more countries. The number of infected cases is doubling and the healthcare systems are suffering even in the developed countries rather than the developing ones. To date, it is clear that about 85–90% of cases pass without the need for hospitalization; however, about 10% of the COVID-19 patients require hospitalization and intensive care services. Many countries are trying their best to avoid worsening the situation by “flattening the curve” of the pandemic by preventing and delaying the spread of the virus to keep a large portion of the population not infected at the same time. Therefore, many countries imposed lockdown measures to contain the spread of the COVID-19. WHO called all countries to implement comprehensive precautions and apply preventive measures aiming at slowing down transmission and flattening the curve for saving lives and buying time till the development of effective vaccines and/or specific treatments.

Dataset description

The data used in this study were collected from official data repositories such as Johns Hopkins University, WHO and Worldometer official website. These data shows the daily total COVID-19 confirmed positive cases, daily and total deaths, and the total and daily recoveries. Table 1 shows a sample of the highest and lowest countries arranged in descending order by the number of confirmed cases. The table depicts the time-series summary for confirmed and recovered cases as well as the deaths of COVID-19 from twelve countries namely the United States of America (USA), India, Brazil, Russia, United Kingdom (UK), France, Italy, Turkey, Spain, Germany, Saudi Arabia, and Vanuatu.

According to the WHO, the first corona-virus that was detected in the Chinese city of Wuhan in December 2019 has infected more than 110,533,973 people in at least 210 countries and territories globally. Of those infected individuals, more than 2,443,091 people died. China was the first country that has more than 89,806 reported infections. The Chinese government completely locked down major cities, restricted the movements of millions, and suspended business operations for a period of time in order to contain the COVID-19 pandemic. As for the time of preparing this study, things are getting worse, and the disease is spreading rapidly around the world, with countries like Spain, Italy, France, Germany, and the UK reporting more than 2,071,615 cases each. Other countries like Saudi Arabia and South Africa have also seen a recent spike beyond 365,325 cases, while most world countries (ex. Vanuatu) have less than 100 confirmed cases as shown in Table 1.

The predictive machine learning model

This study is mainly developed on a decision tree algorithm on the COVID-19 global real-time data. The core idea is to utilize supervised machine learning algorithms for time-series forecasting. The algorithms proposed for this work namely: decision tree algorithm and linear regression, are powerful models in predicting sequence and time-series data-related problems.

Experimental evaluation

In this section, we present and discuss the experimental results of the proposed method. The experimental results are presented visually and tabular. Moreover, a comparison with results from other previous epidemics will be discussed.

Experimental data

Currently, it is feasible to predict for how long the outbreak of COVID-19 will last and how the epidemic will unfold. This is because of the new features exhibited by COVID-19 and a lot of uncertainties remain problematic. Some domain experts remain optimistic that the transmission will gradually decrease during the northern hemisphere summer, as they consider that COVID-19 will be like the epidemics of seasonal influenza. With the help of machine learning, we developed a predictive model using the available data of COVID-19 found in famous data repository websites.

The WHO uses empirical values to show the rate of confirmed cases, mortality rate, recovery rate, and growth rate. Equation (1) was developed to compute the rate of confirmed change and the mortality rate is computed based on Eq. (2). The recovery rate of patients is computed based on Eq. (3), and Eq. (4) is used to calculate the growth rate of the pandemic.

$$CC[i] = \text{Confirmed}[i] - \text{Confirmed}[i-1] \quad (1)$$

$$\text{Mortality Rate}[i] = \text{Deaths}[i] / \text{Confirmed}[i] \quad (2)$$

$$\text{Recovery Rate}[i] = \text{Recoveries}[i] / \text{Confirmed}[i] \quad (3)$$

$$GR[i] = \text{Growth Rate}[i] = CC[i] / \text{Confirmed}[i] \quad (4)$$

Nowadays, the USA has the majority of confirmed cases, with over 28 million infections as of February 18, 2021. Table 2 presents sample data recorded from January 22, 2020 to January 19, 2021 for the USA that shows only a single patient was detected on the first date and alarmingly increased to 24,246,830 on January 19, 2021. Likewise, the recovery, death, confirmed changes, mortality rates, recovery rate, and growth rate of the USA are described.

Table 3 presents sample data recorded globally from January 22 (2020-01-22) to January 19 (2021-01-19). It

Table 1 A sample of the highest and lowest countries arranged in descending order

Country	Date	Lat	Long	Confirmed	Recoveries	Deaths
USA	2021-01-19	40.000000	-100.000000	2,424,830	6,298,082	401,553
India	2021-01-19	20.593684	78.962880	10,595,639	10,245,741	152,718
Brazil	2021-01-19	-14.235000	-51.925300	8,573,864	7,618,080	211,491
Russia	2021-01-19	61.524010	105.318756	3,574,330	2,970,450	65,632
UK	2021-01-19	270.029898	-482.924666	3,476,804	8363	91,643
France	2021-01-19	77.103595	-118.075614	2,996,784	217,745	71,482
Italy	2021-01-19	41.871940	12.567380	2,400,598	1,781,917	83,157
Turkey	2021-01-19	38.963700	35.243300	2,399,781	2,277,987	24,328
Spain	2021-01-19	40.463667	-3.749220	2,370,742	150,376	54,173
Germany	2021-01-19	51.165691	10.451526	2,071,615	1,757,713	48,997
Saudi Arabia	2021-01-19	23.885942	45.079162	365,325	357,004	6335
Vanuatu	2021-01-19	-15.376700	166.959200	1	1	0

evidently presents that the spread of COVID-19 grows alarmingly from 540 confirmed cases in January 22, 2020 to 95,390,046 on January 19, 2021. The number of recovered people on 22 January was limited to 28 and increased to 52,370,571 on January 19, 2021. The number of dead people by the COVID-19 on January 22 was 17 and alarmingly increased to 2,037,575 on January 19, 2021.

Almost every country and union territory have declared a lock-down time to prevent the outbreak of COVID-19. Figure 1 shows the lock-down day for mainland China and the USA. As it is presented in the same Figure, China had an effective lock-down following the outbreak of COVID-19. China has declared to put Wuhan City, the center of the outbreak, on lock-down on January 23, 2020. Before the lock-down time, the growth rate of the pandemic was 0.361 and decreased to 0.020 after the lock-down and hence China is considered as a model for the lock-down as the spread of the virus is getting flattened over time. Although it is not as effective as mainland China, the US growth rate of the COVID-19 has declined after the lock-down. The growth rate for the USA was 0.277 before the lock-down and declined to 0.176 after the lock-down. Figure 2 shows the global rates for confirmed, recoveries and deaths.

Experimental results

The proposed method has forecasted the possible confirmed cases for the upcoming 7 days for the USA. Experimental results showed that the confirmed cases are exponentially increasing from a few hundreds of thousands to nearly two and a half million. Our observation at this particular point is that the prediction is not as optimal as we have used few numbers of records in our deep learning model that is a challenging problem to train deep learning models using few datasets. Figure 3 depicts the forecasting of confirmed cases for the globe. Similarly, the confirmed cases of the pandemic are forecasted as seen in Fig. 4 that indicates the predicted values are close to the test values, while Figs. 5 and 6 present the forecasting of deaths for the globe. To validate the performance of the proposed method, we used root mean square error on each of the three attributes namely confirmed cases, recoveries and death. Table 4 shows the final prediction of the proposed model for all attributes. Table 5 shows the root mean square error of the experimental results of the proposed model for the specified attributes. This table shows that the overall R2 is 0.99 from the perspective of confirmed cases, and R2 values for deaths, recoveries are 0.99 and 0.99, respectively.

Table 2 A sample of confirmed, recoveries, death, confirmed changes, mortality rates, recovery rate, and growth rate in the USA

Date	Confirmed	Recoveries	Deaths	Confirmed change	Mortality rate	Recovery rate	Growth rate
2020-01-22	1	0	0	0.0	0.000000	0.000000	0.000000
2020-01-23	1	0	0	1.0	0.000000	0.000000	1.000000
2020-01-24	2	0	0	0.0	0.000000	0.000000	0.000000
...
2021-01-17	23,936,773	0	397,600	20,634.0	0.059834	0.222628	0.012717
2021-01-18	24,078,772	0	399,003	19,056.0	0.059468	0.223178	0.011597
2021-01-19	24,246,830	0	401,553	NaN	0.059087	0.228092	NaN

Table 3 A sample of confirmed, recoveries, death, confirmed changes, mortality rates, recovery rate, and growth rate worldwide

Date	Confirmed	Recoveries	Deaths	Confirmed change	Mortality rate	Recovery rate	Growth rate
2020-01-22	540	28	17	89.0	0.031481	0.051852	0.164815
2020-01-23	629	30	18	274.0	0.028617	0.047695	0.435612
2020-01-24	903	35	25	446.0	0.027685	0.038760	0.493909
...
2021-01-17	94,290,354	51,669,727	2,011,705	506,567.0	0.021335	0.547985	0.005372
2021-01-18	94,796,921	51,978,127	2,020,858	593,125.0	0.021318	0.548310	0.006257
2021-01-19	95,390,046	52,370,571	2,037,575	NaN	0.021360	0.549015	NaN

Comparison with state-of-the-art methods

As shown in Table 6, we performed a comparative study using the most up-to-date methods. A proposed model is compared with various state-of-the-art models (random forest, ARIMA, and deep learning) and the accuracy of the machine learning models on the training dataset is evaluated using root mean square error (RMSE) and mean absolute error (MAE). Substantial results are obtained when comparing the proposed decision tree model’s experimental results to those of the leading state-of-the-art approaches. Due to the good performance of the decision tree model, it may be extended and used to forecast other countries.

Comparison with other epidemics

Table 7 presents the most known viruses in the past 20 years such as severe acute respiratory syndrome (SARS) in 2002–

2003 (Hu et al. 2017), H1N1 influenza in 2009–2010 (Lathouwers et al. 2017). Middle East respiratory syndrome (MERS) coronaviruses in 2012–2017 (Chu et al. 2019), Ebola in 2013–2016 (Ebola 2020), and COVID-19 in 2019–2020 Lai et al. (2020a, b, c). Unlike other diseases, COVID-19 is still spreading worldwide. The rate of spread for COVID-19 is still lower than the most known pandemics. Moreover, this virus has infected more people than recent outbreaks such as SARS or Ebola, and it does not hit the scale of the most massive modern pandemics such as H1N1 or the seasonal flu. Every year, the seasonal flu infects millions of people, and it is not life-threatening for most people who have to go infected. In contrast, the total reported cases of Ebola are less than 30,000, but it was treated as a crisis because the big number of sick people are dead. Currently, COVID-19 is deadly than the normal flu, but its mortality rate of 6.87% is lesser compared to the mortality rates of other outbreaks such as MERS or Ebola which recorded 34.40 and 39.53%, respectively.

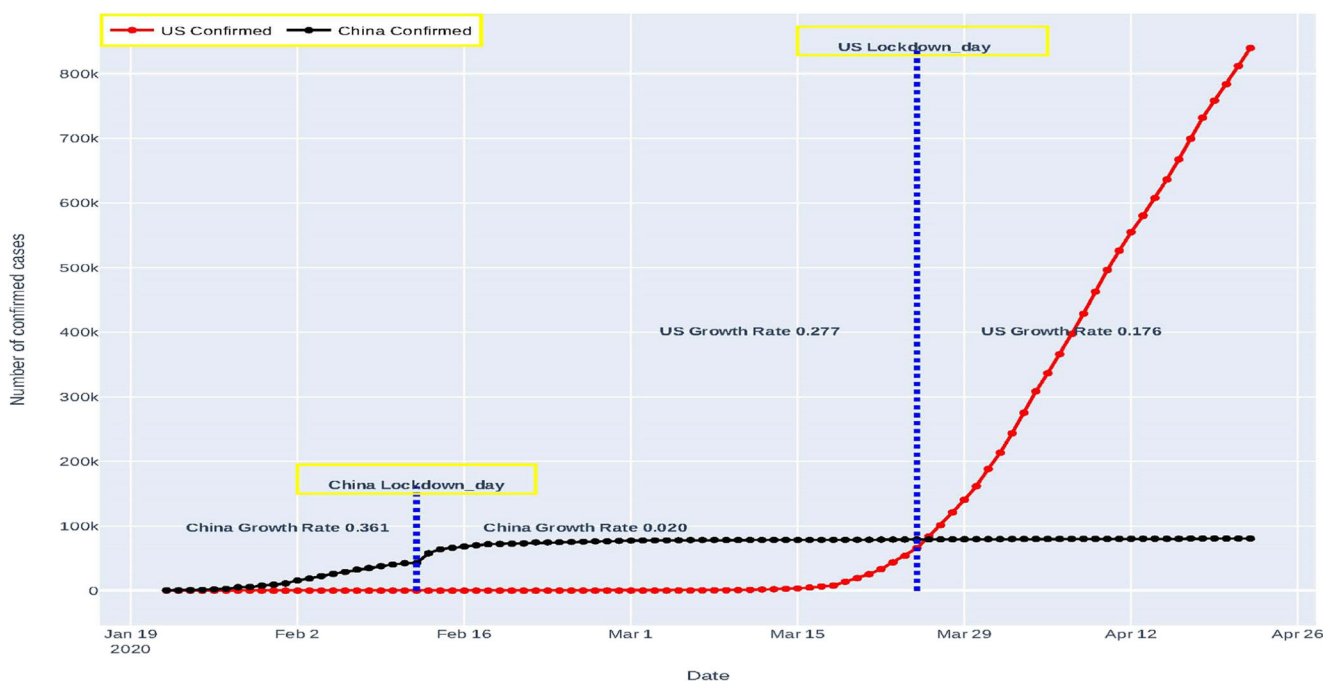


Fig. 1 Lock-down days for the USA and mainland China

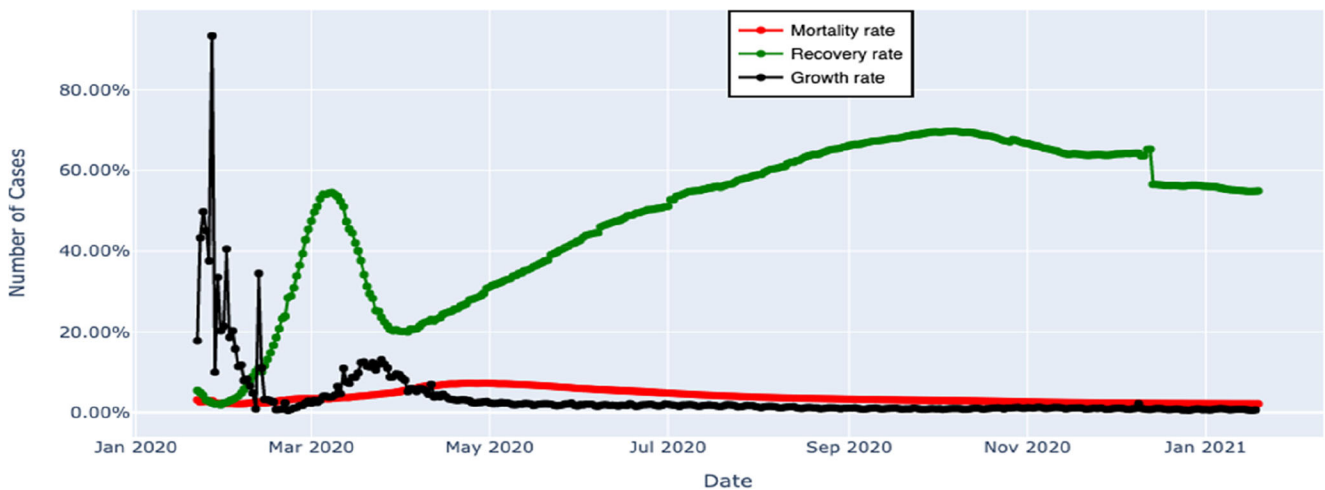


Fig. 2 Worldwide rates

Almost, many common key symptoms exist in all pandemics such as cough, fever, and shortness of breath. Moreover, people of all ages are prone to infection COVID-19 and the other pandemics are deadliest among older patients with the weaker immune system. The mortality rate multiplied rapidly as patients got older, to a high percentage among patients over 65. In comparison to SARS, Ebola, and MERS coronaviruses, which were identified in the past 20 years, COVID-19 is likely more highly transmissible but not as deadly, the researchers noted. SARS had a mortality rate of 9.6%, MERS has a rate of 34.4%, and Ebola has a rate of

39.53%. Unlike SARS and MERS, hospital-based outbreaks do not seem to be a hallmark of COVID-19 at this time.

Discussion

The COVID-19 pandemic has become the biggest threat to human beings in many aspects such as health-wise, financial markets and economic crisis. Major financial institutions and banks have stopped forecasting the global economy, with the organization for economic cooperation and development

Table 4 The experimental results for the expected values of the different attributes

Date	Confirmed	Predicted confirmed	Deaths	Predicted deaths	Recoveries	Predicted recovered
2021-01-01	83,408,277	83,408,277.0	1,811,176	1,811,176.0	46,770,872	46,770,872.0
2021-01-02	84,027,168	84,027,168.0	1,819,284	1,819,284.0	47,073,882	47,073,882.0
2021-01-03	84,544,789	84,544,789.0	1,826,405	1,826,405.0	47,325,731	47,325,731.0
2021-01-04	85,085,071	85,812,929.0	1,836,470	1,836,470.0	47,600,661	47,600,661.0
2021-01-05	85,812,929	85,812,929.0	1,851,716	1,851,716.0	47,910,471	47,600,661.0
2021-01-06	86,583,923	86,583,923.0	1,866,575	1,866,575.0	48,215,743	48,529,412.0
2021-01-07	87,431,624	87,431,624.0	1,881,002	1,881,002.0	48,529,412	48,529,412.0
2021-01-08	88,243,896	88,243,896.0	1,896,212	1,896,212.0	48,819,361	48,819,361.0
2021-01-09	88,999,272	88,999,272.0	1,908,786	1,908,786.0	49,146,579	49,146,579.0
2021-01-10	89,582,189	89,582,189.0	1,916,851	1,916,851.0	49,407,426	49,146,579.0
2021-01-11	90,191,209	90,191,209.0	1,926,931	1,916,851.0	49,684,961	49,684,961.0
2021-01-12	90,888,320	91,630,462.0	1,944,089	1,960,304.0	50,020,565	50,020,565.0
2021-01-13	91,630,462	91,630,462.0	1,960,304	1,960,304.0	50,377,442	50,377,442.0
2021-01-14	92,377,506	92,377,506.0	1,975,446	1,975,446.0	50,737,194	50,737,194.0
2021-01-15	93,135,429	93,135,429.0	1,990,287	1,990,287.0	51,051,058	46,770,872.0
2021-01-16	93,747,244	93,747,244.0	2,003,151	2,003,151.0	51,364,439	51,364,439.0
2021-01-17	94,290,354	94,796,921.0	2,011,705	2,011,705.0	51,669,727	47,325,731.0
2021-01-18	94,796,921	94,796,921.0	2,020,858	2,037,575.0	51,978,127	51,978,127.0
2021-01-19	95,390,046	95,390,046.0	2,037,575	2,037,575.0	52,370,571	47,600,661.0

Table 5 The performance of the decision tree model for the different attributes

Country	Feature	R2	MAPE	MAE	MPE	RMSE	CORR
Worldwide	Confirmed	0.993	0.160	0.047	0.040	0.085	0.996
	Deaths	0.993	0.089	0.054	0.011	0.084	0.996
	Recoveries	0.103	1.549	0.588	-1.342	0.947	0.552
USA	Confirmed	0.995	0.107	0.049	-0.038	0.068	0.998
	Deaths	0.997	0.292	0.027	-0.259	0.056	0.998
	Recoveries	0.000	1.000	0.999	-1.000	1.000	NaN
Brazil	Confirmed	0.989	0.073	0.058	0.025	0.106	0.994
	Deaths	0.980	0.522	0.096	-0.400	0.142	0.990
	Recoveries	0.992	0.067	0.056	0.012	0.089	0.996
India	Confirmed	0.995	0.248	0.050	-0.200	0.073	0.997
	Deaths	0.997	0.053	0.031	-0.008	0.059	0.998
	Recoveries	0.998	0.070	0.031	-0.013	0.049	0.999
Spain	Confirmed	0.977	0.207	0.098	-0.081	0.152	0.988
	Deaths	0.971	0.213	0.110	0.110	0.171	0.985
	Recoveries	1.000	NaN	0.000	NaN	0.000	NaN
Italy	Confirmed	0.996	0.113	0.038	-0.086	0.062	0.998
	Deaths	0.995	0.285	0.042	-0.218	0.069	0.998
	Recoveries	0.997	0.405	0.029	0.359	0.051	0.999
France	Confirmed	0.982	0.277	0.069	-0.124	0.133	0.991
	Deaths	0.987	0.231	0.077	-0.069	0.116	0.993
	Recoveries	0.991	0.081	0.059	-0.028	0.097	0.995
UK	Confirmed	0.994	0.126	0.044	0.096	0.075	0.997
	Deaths	0.991	0.125	0.063	-0.082	0.094	0.996
	Recoveries	0.992	0.101	0.062	-0.030	0.092	0.996
Germany	Confirmed	0.996	0.050	0.040	0.006	0.060	0.998
	Deaths	0.997	0.120	0.029	-0.030	0.057	0.998
	Recoveries	0.993	0.168	0.059	0.086	0.084	0.996
Russia	Confirmed	0.991	0.308	0.055	-0.189	0.094	0.996
	Deaths	0.996	0.142	0.039	-0.009	0.063	0.998
	Recoveries	0.997	0.077	0.031	-0.061	0.054	0.999
Turkey	Confirmed	0.996	0.051	0.033	-0.004	0.065	0.998
	Deaths	0.998	0.046	0.028	0.022	0.048	0.999
	Recoveries	0.994	0.541	0.041	0.506	0.080	0.997

being one of the latest to do so. Fear of the COVID-19 has negatively affected the global economy; mainly the markets are badly hit, worldwide, with stock prices and bond fall steeply. Even though the global economy was expected to grow by 2.9% in 2020, recent economic prediction forecasts only 2.4%. For instance, the manufacturing sector in China has been hit massively by COVID-19 pandemic. Such a slowdown in Chinese manufacturing activity has hurt countries with close economic links with China; many of those countries are from Asia Pacific economic corridors such as Vietnam, Singapore, and South Korea. The good news from China is, factories have resumed operations. In summary, due to the outbreak of the COVID-19 pandemic, close to 1.6

billion children worldwide are absent from school, many business sectors have lost their customers in the USA such as restaurants and aviation. Therefore, the next “Estimation of slowdown COVID-19” section will discuss when COVID-19 is going to be over.

Estimation of slowdown COVID-19

Many countries around the world implemented an effective shutdown in order to contain the fast-spreading COVID-19. Restrictions on daily life for millions of people, such as school closures, large-scale social distancing, and bans on public gatherings, have been put in place. Because it was not easy

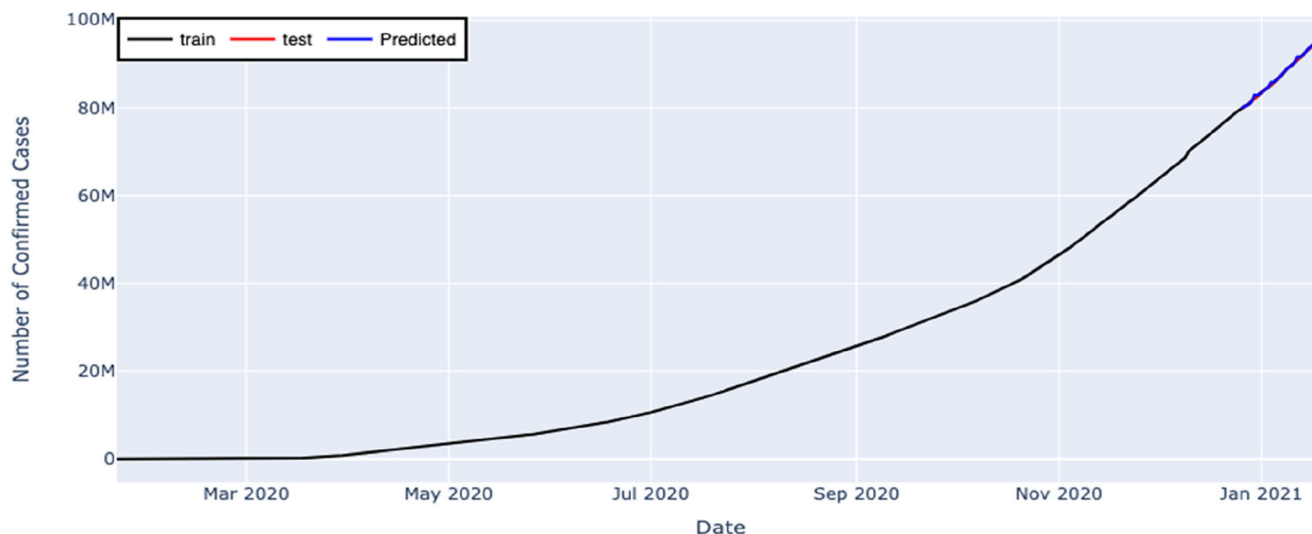


Fig. 3 Comparison between train, test and predicted for the confirmed cases globally

to know exactly when a vaccine for COVID-19 becomes available, these protective measures will be extended for the next few months. However, health experts are much more cautious. Lifting lock-down restrictions in order to alleviate the economic and social harm that results from long-term lock-down could open the door to future waves of the COVID-19 pandemic.

In our assessment, we have developed a predictive model that can forecast the time period that the COVID-19 can be suppressed. The proposed method depicts the possible stoppage of the pandemic using the normal distribution. It specifically presents the statistical estimation of the slow down period of the pandemic which is extracted based on the concept of normal distribution. The following equations explain how to calculate the area under the curve between $\mu + 2\sigma$ and $\mu + 3\sigma$. Therefore, we selected the period that the virus can stop between $\mu + 2\sigma$ and $\mu + 3\sigma$.

$$\begin{aligned}
 p(\mu + 2\sigma < X < \mu + 3\sigma) &= p\left(\frac{\mu + 2\sigma - \mu}{\sigma} < z < \frac{\mu + 3\sigma - \mu}{\sigma}\right) \\
 &= p\left(\frac{2\sigma}{\sigma} < z < \frac{3\sigma}{\sigma}\right) = p(2 < z < 3) = 2.1\%
 \end{aligned}$$

Table 8 presents the possible period that the virus can slow down from being infectious in the top countries. Table 8 shows the prediction of the deadline for India and the results show that the predicted number of confirmed cases will be 548,318 on August 05, 2021, and after three months, that is, on November 15, 2021, the number of confirmed cases will remain 156. As can be seen from Fig. 7, experimental results of the predictive model, the USA will have 2,379,799 confirmed cases on August 17, 2021, and three months (on November 30, 2021), the expected number of positive cases in 1147 patients.

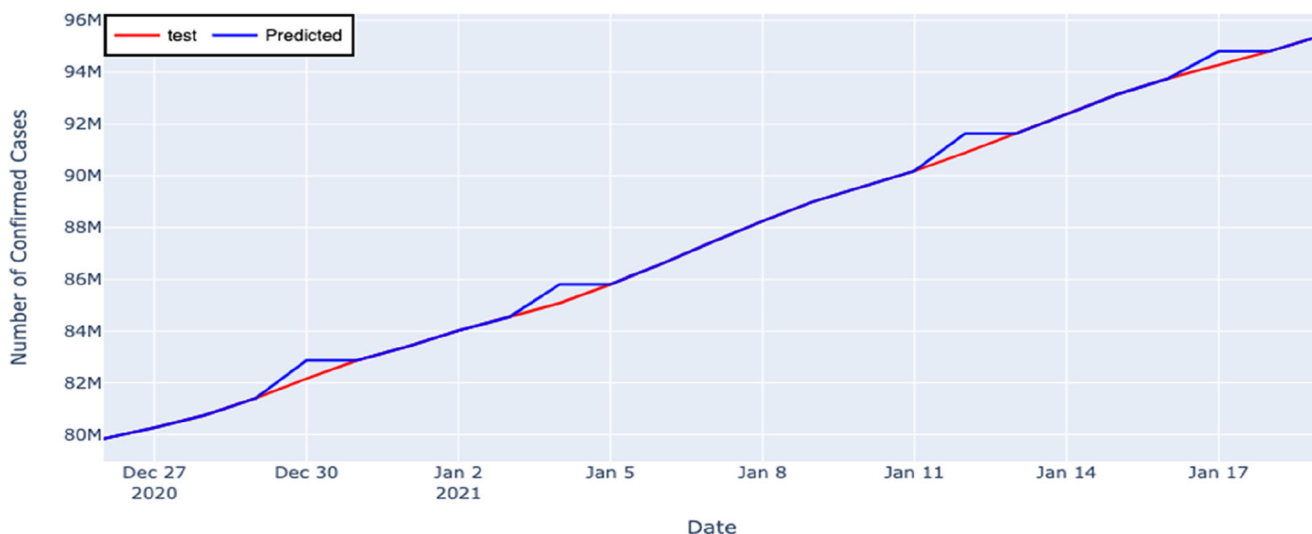


Fig. 4 The forecasted data of confirmed cases for the globe

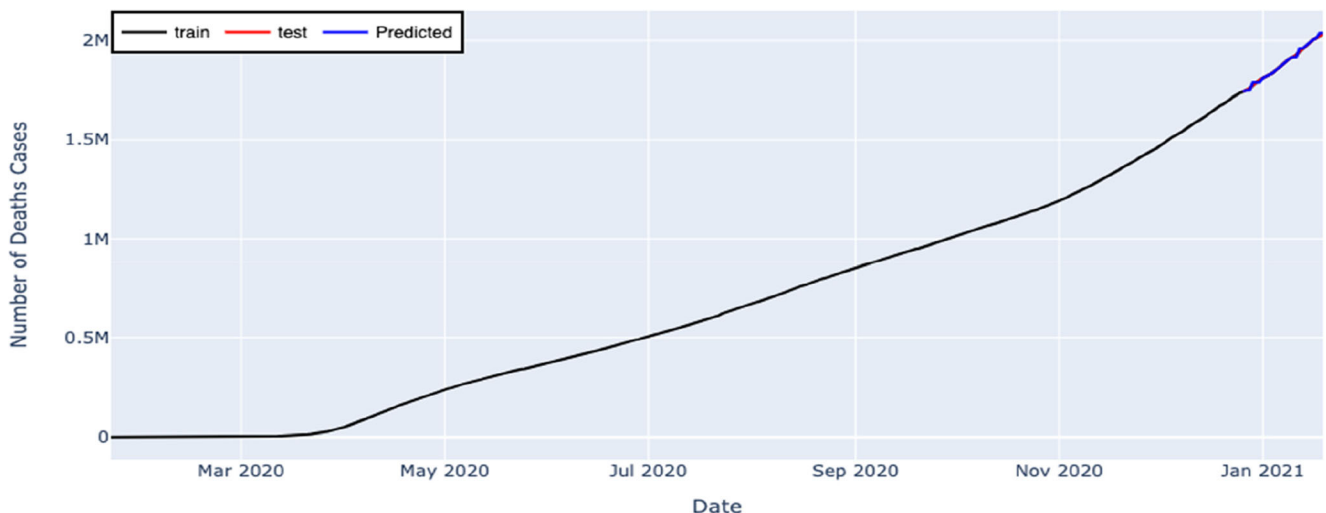


Fig. 5 Comparison between train, test and predicted data for global deaths

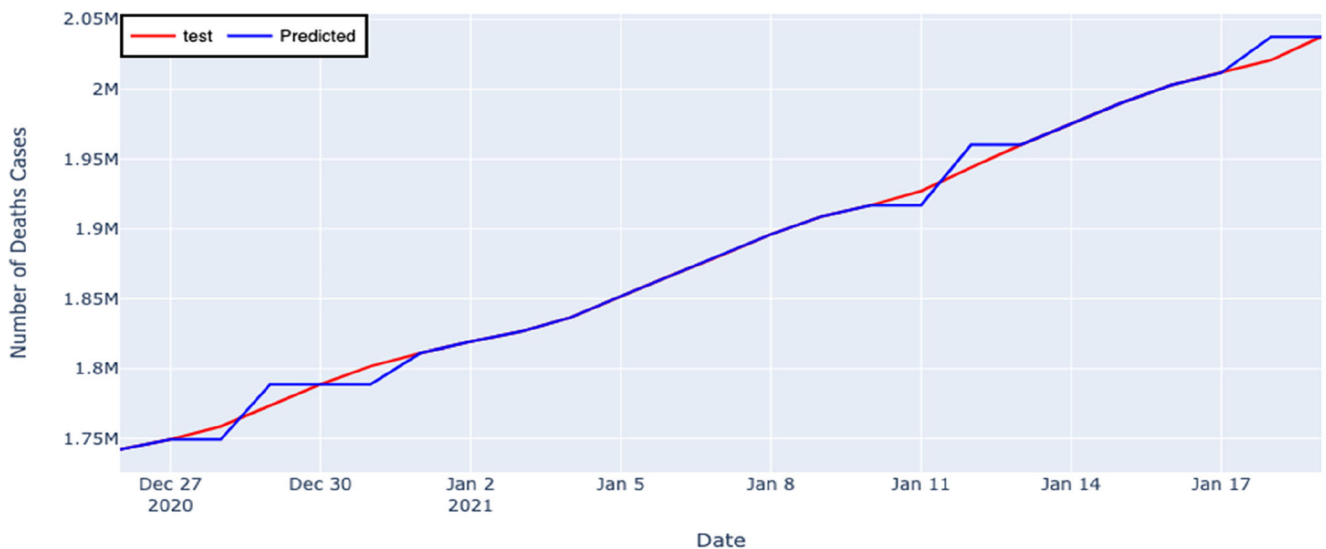


Fig. 6 Comparison between test and predicted data for global deaths

Conclusion

A machine learning model has been developed to predict the estimation of the spread of the COVID-19 infection in many countries and the expected period after which the virus can be

stopped. Globally, our results forecasted that the COVID-19 infections will greatly decline during the first week of September 2021 when it will be going to an end shortly afterward. Moreover, we can apply our proposed model to other countries that are affected by the COVID-19. Additionally,

Table 6 Comparison between the proposed model and latest state-of-the-art techniques

	Other models			The proposed model		
	Country	Metrics	Value	Country	Metrics	Value
Machine learning (Random Forest) (Z. Malki et al. 2020)	Worldwide	MAE	368.821	Worldwide	MAE	0.047
ARIMA Bayyurt and Bayyurt (2020)	Spain	RMSE	379.89	Spain	RMSE	0.152
Deep learning (LSTM) Direkoglu and Sah (2020)	Worldwide	MAE	30758	Worldwide	MAE	0.047

Table 7 Comparison with other epidemics (CIDRAP 2020; Healthline 2020; Kelly-Cirino et al. 2019; Helmy et al. 2020; Organization WH 2020; Sohrabi et al. 2020; Yosra et al. 2020)

Epidemic	COVID-19	SARS	EBOLA	MERS	H1N1
Start year	2019	2003	2014	2012	2009
End year	2021	2004	2016	2017	2010
Confirmed	95,390,046 Global population	8096 29 countries	28,646 10 countries	2494 27 countries	6,724,149 Global population
Deaths	2,037,575	774	11,323	858	284,000
Mortality	2.14	9.56	39.53	34.40	4.22
Key symptoms	Cough, fever, shortness of breath	Fever, respiratory symptoms, cough, malaise	Fever, aches and pains, weakness, diarrhea, vomiting	Cough, fever, aches, shortness of breath sore, throat, headache	Fever, chills, cough, body aches
First detection	December 2019 in Wuhan, China	November 2002 in Guangdong province of China	December 2013 in Guinea	2012 in Saudi Arabia	January 2009 in Mexico
Most affected groups	Adults over 65 with underlying health conditions, children's	Patients ages 60 and older 55% death rate	Children 20% death rate	Patients ages 60	Children 47% death rate people ages 65 11%
Treatment/vaccine	Vaccine	Vaccine	None	None	Antivirals/vaccine

Table 8 Expected deadline for the selected countries

Country	First case	Top point	Start date	End date	Start value	End value
US	2020-01-22	2021-01-19	2021-08-17	2021-11-30	2,379,799.0	1147.0
Brazil	2020-02-26	2021-01-19	2021-06-23	2021-09-26	1,926,824.0	19,638.0
India	2020-01-30	2021-01-19	2021-08-05	2021-11-15	548,318.0	156.0
Spain	2020-02-01	2021-01-19	2021-08-02	2021-11-12	248,970.0	17,963.0
Italy	2020-01-31	2021-01-19	2021-08-03	2021-11-14	240,436.0	35,713.0
France	2020-01-24	2021-01-19	2021-08-14	2021-11-27	201,853.0	2293.0
UK	2020-01-31	2021-01-19	2021-08-03	2021-11-14	284,812.0	5467.0
Germany	2020-01-27	2021-01-19	2021-08-09	2021-11-21	194,458.0	5795.0
Russia	2020-01-31	2021-01-19	2021-08-03	2021-11-14	640,246.0	147.0
Turkey	2020-03-11	2021-01-19	2021-06-01	2021-08-31	222,402.0	98,674.0

our model could also evaluate the effect of the public health guidelines, infection control, and lock-down decisions that were taken to stop the COVID-19 pandemic. Future work

could focus on applying a deep learning model by using big data as training data. Moreover, our proposed model can apply to specific countries.

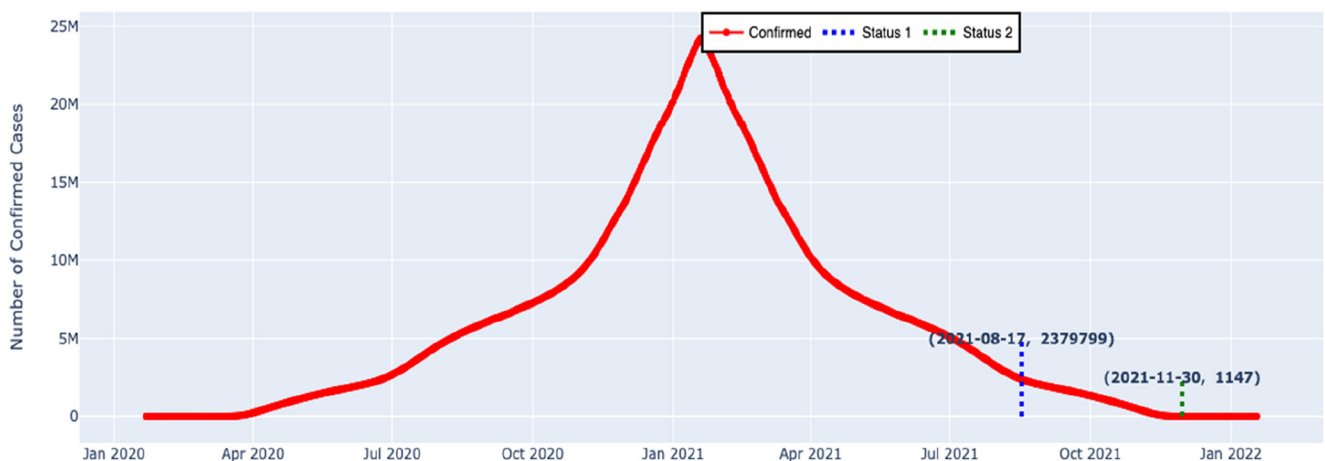


Fig. 7 Expected deadline for US COVID-19

Author contribution Zohair Malki: Supervision and Project administration.

El-Sayed Atlam: Methodology, idea, Writing- Reviewing, Editing and Supervision.

Ashraf Ewis: Project administration and Reviewing.

Guesh Dagnev: Data curation, Writing- Original draft preparation.

Osama Ghoneim: Original draft preparation and Data collection

Abdallah A. Mohamed: Visualization and Editing

Mohamed M. Abdel-Daim: Investigation, Reviewing and Editing.

Ibrahim Gad: Formal analysis, Idea, Methodology, Software.

Funding “The authors have all the responsibility for all funding and did not have any support from any organization”

Data availability The data used in this study were collected from the official data repositories such as Johns Hopkins University, WHO and Worldometer official website

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

URL: <https://www.worldometers.info/coronavirus/>.

Declarations

Ethics approval and consent to participate Not applicable

Consent for publication All authors agreed on this submission.

Competing interests The authors declare no competing interests.

References

- Bayyurt L. and Bayyurt B. (2020) Forecasting of COVID-19 cases and deaths using ARIMA models. medrxiv <https://doi.org/10.1101/2020.04.17.20069237>, [10.1101/2020.04.17.20069237](https://doi.org/10.1101/2020.04.17.20069237)
- Chu DKW, Hui KPY, Perera RAPM, Miguel E, Oladipo JO, Traore A, Fassi-Fihri O, Chan MCW, Zhou Z, So RTY, Chevalier V, Peiris JSM (2019) A52 MERS corona-viruses from camels in Africa exhibit region-dependent genetic diversity. *Virus Evolution* 5 [10.1093/ve/vez002.051](https://doi.org/10.1093/ve/vez002.051), <https://doi.org/10.1093/ve/vez002.051>
- CIDRAP, (2020) CIDRAP - Center for Infectious Disease Research and Policy <https://www.cidrap.umn.edu/news-perspective/2013/01/study-puts-global-2009-pandemic-h1n1-infection-rate-24> (Accessed April 2020)
- Dandekar R, Barbastathis G, (2020) Quantifying the effect of quarantine control in covid-19 infectious spread using machine learning. medRxiv
- Direkoglu C. and Sah M (2020) Worldwide and regional forecasting of coronavirus (covid-19) spread using a deep learning model <https://doi.org/10.1101/2020.05.23.20111039>, [10.1101/2020.05.23.20111039](https://doi.org/10.1101/2020.05.23.20111039)
- Ebola (2020) First Ebola vaccine approved. *Nat Biotechnol* 38:6–6. <https://doi.org/10.1038/2Fs41587-019-0385-7>
- Healthline (2020) How deadly is the coronavirus compared to past outbreaks flu pandemic [https://www.healthline.com/health-news/how-deadly-is-the-coronavirus-compared-to-past-outbreaks#2009-\(H1N1\)-flu-pandemic](https://www.healthline.com/health-news/how-deadly-is-the-coronavirus-compared-to-past-outbreaks#2009-(H1N1)-flu-pandemic)
- Helmy, Y. A., Fawzy M., Shehata, A.A. (2020) The covid-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *Journal of Clinical Medicine* 9
- Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, Xie JZ, Shen XR, Zhang YZ, Wang N, Luo DS, Zheng XS, Wang MN, Daszak P, Wang LF, Cui J, Shi ZL (2017) Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog* 13:e1006698. <https://doi.org/10.1371/journal.ppat.1006698>, [10.1371/journal.ppat.1006698](https://doi.org/10.1371/journal.ppat.1006698)
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X et al (2020) Clinical features of patients infected with 2019 novel corona virus in Wuhan, China. *Lancet* 395:497–506
- Kelly-Cirino C, Mazzola LT, Chua A, Oxenford CJ, Van Kerkhove MD (2019) An updated roadmap for MERS-CoV research and product development: focus on diagnostics. *BMJ Glob Health* 4:e001105
- Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR (2020a) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (covid-19): the epidemic and the challenges. *International journal of antimicrobial agents*, 105924. <https://doi.org/10.1016/j.ijantimicag.2020.105924>
- Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR (2020b) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-483 2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *International Journal of Antimicrobial Agents* 55:105924. <https://doi.org/10.1016/2Fj.ijantimicag.2020.105924>, [10.1016/j.ijantimicag.2020.105924](https://doi.org/10.1016/j.ijantimicag.2020.105924)
- Lai, C.C., Wang, C.Y., Wang, Y.H., Hsueh, S.C., Ko, W.C., Hsueh, P.R., 2020c. Global epidemiology of coronavirus disease 2019: disease incidence, daily cumulative index, mortality, and their association with country healthcare resources and economic status. *International Journal of Antimicrobial Agents*, 105946
- Lathouwers E, Wong EY, Luo D, Seyed kazemi S, Meyer SD, Brown K (2017) HIV-1 resistance rarely observed in subjects using darunavir once-daily regimens across clinical studies. *HIV496 ClinicalTrials* 18:196–104. <https://doi.org/10.1080/2F15284336.2017.1387690>, [10.1080/15284336.2017.1387690](https://doi.org/10.1080/15284336.2017.1387690)
- Malki Z, Atlam ES, Hassanien AE, Dagnev G, Elhosseini MA, Gad I (2020) Association between weather data and COVID-19 pandemic predicting mortality rate: machine learning approaches. *Chaos, Solitons & Fractals* 138:110137. <https://doi.org/10.1016/j.chaos.2020.110137>, [10.1016/2Fj.chaos.2020.110137](https://doi.org/10.1016/2Fj.chaos.2020.110137)
- Organization WH, et al. (2020) Rational use of personal protective equipment for coronavirus disease (COVID-19): interim guidance, 27 February 2020. Technical Report. World Health Organization
- Punn NS, Sonbhadra SK, Agarwal S (2020) Covid-19 epidemic analysis using machine learning and deep learning algorithms. medRxiv
- Qiu H, Wu J, Hong L, Luo Y, Song Q, Chen D (2020) Clinical and epidemiological features of 36 children with coronavirus disease 2019 (covid-19) in Zhejiang, China: an observational cohort study. *Lancet Infect Dis* 20:689–696
- Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R (2020) World Health Organization declares global emergency: a review of the 2019 novel coronavirus (covid-19). *International Journal of Surgery*
- World Health Organization (WHO) (2020) Coronavirus <https://www.who.int/health-topics/coronavirus> Accessed April 13, 2020
- Worldometer 2020 COVID-19 CORO-NAVIRUS PANDEMIC <https://www.worldometers.info/coronavirus/> Accessed April 13, 2020
- Wu J, Liu J, Zhao X, Liu C, Wang W, Wang D, Xu W, Zhang C, Yu J, Jiang B et al. (2020) Clinical characteristics of imported cases of covid-19 in jiangsu province: a multi-center descriptive study. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*
- Yang P, Liu P, Li D, Zhao D (2020) Corona virus disease 2019, a growing threat to children? *The Journal of Infection*
- Yosra AH, Mohamed F, Ahmed E, Ahmed S, Scott PK, Awad AS (2020) The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *J Clin Med* 9: 1255

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.