ABSTRACT

Since December 2019 the world is experiencing a deadly disease caused by a novel coronavirus termed as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease associated with this virus is known as COVID-19. This paper focuses on COVID-19 based on freely available datasets including the ones in Kaggle repository. Data analytics is provided on a number of aspects of COVID-19 including the symptoms of this disease, the difference of COVID-19 with other diseases caused by severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), and swine flu. The impact of temperature on the spread of COVID-19 is also discussed based on the datasets. Moreover, data visualization is provided on the comparison of infections in males/females which shows that males are more prone to this disease and the older people are more at risk. Based on the data, the pattern in the increase of confirmed cases is found to be an exponential curve in nature. Finally, the relative number of confirmed, recovered and death cases in different countries are shown with data visualization.

Keywords: COVID-19; Coronavirus; SARS-CoV-2; MERS; pandemic; vaccine.
1. INTRODUCTION

Enveloped, single stranded positive-sense ribonucleic acid (RNA) viruses named coronaviruses contain one of the largest viral genomes which are around 32 kbp in length. They can infect humans as well as a wide range of animals [1]. The 2019 novel coronavirus termed as SARS-CoV-2 caused pneumonia outbreak in Wuhan, China resulting in the 2019-2020 coronavirus pandemic declared by World Health Organization (WHO). It belongs to the Orthocoronavirinae subfamily. It is distinct from Middle East respiratory syndrome (MERS) and severe acute respiratory syndrome corona virus (SARS-CoV) [2]. Wuhan’s Huanan Seafood Wholesale Market trades a variety of live animal species which includes fish, poultry, marmots, snakes and bats which linked the outbreak [3-5]. Researchers identified the highly identical genome similarity between SARS-CoV-2 and bat coronavirus and pointed to bat as the natural host [6,7]. The infected patients showed clinical manifestations of dry cough, fever, confusion, sore throat, rhinorrhea, chest pain, dyspnea, bilateral lung infiltrates on imaging, nausea, vomiting and diarrhea [4,6]. The disease caused by SARS-CoV-2 known as COVID-19 can be deadly. This happens when the severity of the disease onset results in massive alveolar damage with progressive respiratory failure with a 2% case fatality rate [3]. According to the WHO, an infected patient can spread the virus during close contact and via respiratory droplets while coughing, talking or sneezing. People can inadvertently transfer the pathogen to a mucous membrane by touching contaminated surface. Though the virus can be transmitted by the asymptomatic carrier [7], it is the most contagious when people are symptomatic. A recent study [8] reports that for the case of SARS-CoV-2, aerosol transmission may be possible in closed places when there is longer exposure to the virus. Generally symptoms may arise in patients between two to fourteen days, with an average of five days [9]. According to the Centers for Disease Control and Prevention, the standard diagnosis method for the identification of the virus in the patient is by reverse transcription polymerase chain reaction (rRT-PCR) from nasopharyngeal swab [10]. A combination of symptoms, risk factors and a CT scan showing features of pneumonia can diagnose the infection [11].

In order to prevent the spread of this infection, the WHO recommends frequent hand washing, keeping unwashed hands away from the face, social distancing, and covering coughs and sneezes with a tissue or inner elbow. Some national health authorities recommend masks for the suspects and their caregivers. So far no vaccine or antiviral treatment is available for COVID-19 [12-19]. In order to stop this pandemic, the route of transmission of the virus to humans via animals, identification of the reservoirs, the incubation period of the virus, the characteristics of the susceptible population and their survival rates need to be identified. The analysis of the clinical information regarding age, gender, source of the virus, incubation period, transmission route, treatment response will help researchers to develop vaccines against COVID-19. As of 13 April 2020, COVID-19 has affected more than 1,858,800 patients in 210 countries and territories around the world [12] and has become a major global health concern.

This paper analyzes COVID-19 based on the currently available data [12,20]. Analytics is provided on a number of aspects of COVID-19 including the symptoms, the difference with other viruses, and the impact of temperature. Moreover, data visualization is provided on the comparison of infections on male/females, and the pattern in the increase of confirmed cases and the relative number of confirmed/recovered/death cases in different countries.

The rest of the paper is organized as follows. Section 2 describes the different aspects of COVID-19 using tabular data. Section 3 provides visualization of how the infection has spread across the world using pie charts and bar charts. Section 4 provides concluding remarks.

2. FEATURES OF COVID-19

2.1 Symptoms of COVID-19

The common symptoms of COVID-19 are fever, cough, shortness of breath, muscle ache, headache, sore throat, rhinorrhea (runny nose), chest pain, diarrhea, nausea and vomiting. The patients will not have all the symptoms; rather, they will carry different combination of symptoms. In order to find the most influential symptoms, informatics is provided here using the dataset available in Kaggle repository [20]. Table 1 depicts different combination of symptoms. Note that only the complete data from [20] are used after removing the null values as a part of preprocessing. It can be seen from Table 1 that majority of the patients carry only the symptoms of fever which is 33%. Then there are 30.7%
Table 1. Percentage of patients having different combination of symptoms [20]

| Symptoms                                          | Cases | Percentage% |
|----------------------------------------------------|-------|-------------|
| Fever                                              | 90    | 33%         |
| Fever, Cough                                       | 47    | 17%         |
| Fever, Cough, Shortness of Breath                  | 13    | 4.8%        |
| Fever, Runny Nose                                  | 4     | 1.48%       |
| Throat Pain, Fever                                 | 9     | 3.33%       |
| Vomiting, Diarrhea, Fever, Headache, Cough         | 10    | 3.70%       |
| Fever, Headache                                    | 6     | 2.22%       |
| Fever, Pneumonia                                   | 2     | 0.74%       |
| Fever, Cough, Sputum                               | 6     | 2.22%       |
| More than one sign or symptoms                     | 83    | 30.7%       |
| **Total**                                          | **270** |             |

patients who have more than one sign and symptoms. Among all the symptoms, fever and cough are found to be the most common, indicating that the combination of fever and cough is one of the major indicators of carrying this virus.

Using the dataset in [20] we estimated how long it takes for the symptoms to build from the start of exposure. Table 2 shows the days of incubation period and the associated number of cases. In the dataset [20], many records do not have either the exposure start date or the symptoms start date. Probably many patients could not remember exactly when they were exposed to the virus or when their symptoms began. Hence, either or both of these dates for those patients were recorded as null. Considering only the records that have the exposure start date and the symptoms start date, we got 73 cases. Table 2 shows that incubation period of COVID-19 is between 1 to 14 days. Table 2 also shows that among a total of 73 cases, the highest number of cases is 11 which corresponds to 4 to 5 days. This is consistent with another study [21] where the authors estimated that in most cases, the coronavirus incubation period is about 5 days. However, a larger data sample must be studied to confirm the actual incubation period of COVID-19.

Table 3 shows the number of days from symptom onset to hospital visit. Of note, only the complete data from [20] were used after removing the null values as a part of preprocessing. Table 3 indicates that the number of patients getting hospitalized are the highest in day 1 rather than any other days. However, there are cases when hospitalization took place even after 14 days. This can be explained by the fact that the number of patients in day 1 was less in the months of December and January when people were still not fully concerned about COVID-19. However, the number of hospitalized patients increased rapidly from the end of January and the beginning of February when people were already aware of the symptoms of this disease. So after seeing the symptoms they got hospitalized within one day.

Table 2. Incubation period of covid-19 from exposure start to symptoms onset [20]

| Days | Cases |
|------|-------|
| 1    | 4     |
| 2    | 5     |
| 3    | 3     |
| 4    | 11    |
| 5    | 11    |
| 6    | 9     |
| 7    | 5     |
| 8    | 8     |
| 9    | 7     |
| 10   | 4     |
| 11   | 2     |
| 12   | 2     |
| 13   | 2     |
| 14   | 0     |
| **Total** | **73** |

2.2 Comparison with Other Viruses

The WHO said that the novel coronavirus originated from China to other countries around the world does not seem to be as “deadly as other coronaviruses including MERS and SARS”. WHO’s director general, Tedros Adhanom Ghebreyesus called a briefing on 17 February and said that 80% of patients with COVID-19 have a “mild disease and will recover” and he also added that “it is fatal in 2% of reported cases” [22]. In comparison, the 2003 outbreak of SARS had a case fatality rate of around 10%
(8098 cases and 774 deaths), while MERS killed 34% of people with the illness between 2012 and 2019 (2494 cases and 858 deaths) [14,15]. However, COVID-19 has so far resulted in more deaths (114,698 as of 13 April 2020) [12] than SARS and MERS combined (1632) [16]. In particular, the death rate has increased significantly from mid-February to date. As of 13 April, 2020, the total number of confirmed cases, deaths and recovered cases for COVID-19 are 1,858,800, 114,698 and 429,020, respectively [12]. Since many of the cases are not closed, the death rate cannot be calculated properly. Nevertheless, by calculating the death rate as a ratio of total deaths to total confirmed cases as of 13 April 2020, we obtain a death rate of 6.17% for COVID-19.

Table 3. Incubation period of covid-19 from symptom onset to hospital visit [20]

| Days | Cases |
|------|-------|
| 1    | 221   |
| 2    | 48    |
| 3    | 41    |
| 4    | 30    |
| 5    | 27    |
| 6    | 17    |
| 7    | 20    |
| 8    | 17    |
| 9    | 2     |
| 10   | 3     |
| 11   | 2     |
| 12   | 18    |
| 13   | 1     |
| 14   | 5     |
| Total| 452   |

2.3 Impact of Temperature on COVID-19

This section discusses the possible effect of temperature on the spread of COVID-19. There have been some studies [13,17-19] which report that the spread of COVID-19 may be related with temperature and humidity. According to a study [17], COVID-19 has spread in regions with low temperatures (3-7 degree Celsius). The work in [19] shows that the number of confirmed cases is higher in cool and dry conditions. Using the confirmed cases from [12] and using the temperature data reported in [23,24], we show how the COVID-19 has spread in countries with different temperatures. Table 5 presents the number of affected cases and temperature for some countries. In Table 5, the number of confirmed cases is obtained from [12], the temperature domain and the corresponding temperature range are collected using the maps in [23], and the average yearly temperature is obtained from [24]. Table 5 indicates that the virus has spread comparatively less in countries having tropical and sub-tropical climate than cold climate. In other words, the number of infections is more in countries having lower average temperature than countries with higher average temperature. For example, Oman with a tropical climate (24-34 degree Celsius) and yearly average temperature of 25.6 degree Celsius, has 484 confirmed cases. Countries like USA, UK, Italy, Spain, China and Germany have huge number of affected people where the temperature domain is no more than 18 degree Celsius or the average yearly temperature is less than 14 degree Celsius. However, the apparent spread of the disease in low temperature region does not confirm that the spread of COVID-19 depends on the temperature of a country. This is because many other conflicting factors may influence the temperature versus confirmed cases data. Hence, further investigation is required to find whether temperature has any impact on the spread of SARS-CoV-2 virus.

2.4 Informatics on COVID-19

This section provides visual exploratory data analysis on COVID-19 using a number of vertical bar charts and pie charts. Python language is used to generate the charts. Firstly, we presents the spread of COVID-19 among different genders and among different age groups using the dataset [20]. After preprocessing the dataset [20] based only on gender, we got 902 data samples. Next we preprocessed the dataset [20] based on both gender and age resulting in 829 data samples. Fig. 1(a) shows the number of confirmed cases among the males and females of different age groups using the 829 data samples [20]. Fig. 1(a) also depicts that males, aged between 50 to 60 are more infected than any other group of ages. Similarly the infected rate is also higher for females of that age group. Fig. 1(b) shows that the virus SARS-CoV-2 has affected more males than females for the dataset of 902 samples [20]. However, this does not confirm that females are less susceptible to COVID-19 than males, as many factors may influence the currently obtained data. Nevertheless more investigation is required to find whether any gender or any age group is more prone to this disease.
Table 4. Mortality rate of different viruses

| Virus            | Death Rate   |
|------------------|--------------|
| SARS-CoV-2       | 6.17% *      |
| SARS             | 9.6%         |
| MERS             | 34%          |
| Swine Flu        | 0.02%        |

*The number of infected people is increasing day by day*

Table 5. Infected people of different countries according to temperature domain [12,23,24]

| Country  | Temperature domain            | Temperature range (°C) | Average yearly temperature (°C) | Confirmed cases |
|----------|-------------------------------|------------------------|--------------------------------|-----------------|
| Switzerland | Cold Temperate                 | (0-10)                  | 5.50                            | 24,551          |
| China    | Mix of Warm, Cold and Polar Temperate | (10-18) or (0-10) | 6.95                            | 81,953          |
| UK       | Cold Temperate                 | (0-10)                  | 8.45                            | 73,758          |
| Germany  | Cold Temperate                 | (0-10)                  | 8.50                            | 122,171         |
| USA      | Mostly Warm Temperate, and Cold Temperate | (10-18) or (0-10) | 8.55                            | 502,876         |
| Belgium  | Cold Temperate                 | (0-10)                  | 9.55                            | 26,667          |
| Spain    | Warm Temperate                 | (10-18)                | 13.30                           | 158,273         |
| Italy    | Warm Temperate                 | (10-18)                | 13.45                           | 147,577         |
| Iran     | Sub Tropical, and Warm temperate | (18-24) or (10-18) | 17.25                           | 68,192          |
| South    | Mostly Warm Temperate          | (10-18)                | 17.75                           | 2003            |
| Africa   | Temperate                      |                        | 17.75                           |                 |
| India    | Tropical Temperate             | (24-34)                | 23.65                           | 7,600           |
| Saudi    | Tropical, and Sub              | (24-34) or (18-24)     | 24.65                           | 3651            |
| Arabia   | Tropical Temperate             | (18-24)                |                                 |                 |
| Oman     | Tropical Temperate             | (24-34)                | 25.60                           | 484             |
| Sudan    | Tropical, and Sub              | (24-34) or (18-24)     | 26.90                           | 17              |

Fig. 1(a). Number of infected patients based on age groups
Next, we present the rate of infected cases over time. The total cumulative count of affected people [12] by COVID-19 was initially small for a while. On 22 January 2020, the number of affected people was approximately 580 around the world. In the very next day it became 845. Within a short period, it increased very rapidly. On 01 February 2020, the number crossed 14000 that means 25 times of the number of initial count on 22 January 2020. At the end of February 2020, the number crossed 86000. That is almost 150 times of the initial count on 22 January 2020. Fig. 2 shows total confirmed cases till 9 April 2020 using the data in [12]. From Fig. 2 it can be seen that the confirmed cases is increasing exponentially. This means that the number of cases will be double in a given amount of time. Hence measures must be taken to stop this growth of infections.

Based on the data in [12], Fig. 3 enlists the data of comparative analysis of confirmed (C) cases around the world till 10th April, 2020. The pie chart depicts that 9 countries are the most
infected countries around the world. This representation demonstrates that USA has endured the largest number of infected patients (C= 478,366) and Spain with a confirmed case of 157,053 is in the next position after USA. Then the risk zone country is Italy with a confirmed case of 48,000. The highest country in this case is USA being 18% of the total confirmed cases across the world. Though the deadly virus hit China first but their prevention strategy and medical protection helped China to recover with time. Fig. 5 presents the recovery cases in the world based on the data in [12]. From Fig. 5, we can see that the recovery cases in China is the highest which is 21% of the total recovery cases in the world. This recovery rate in China is quite satisfactory under the present worldwide condition. Note that the data present in Fig. 5 does not give a complete scenario of the recovery rate in different countries. This is because indifferent countries, the number of new cases is different and many of them are being recovered.

With time the total number of affected people with COVID-19 is increasing. Fig. 6 is a bar chart showing the number of recovered cases and total confirmed cases based on the dataset in [12]. On the other hand, Fig. 7 is a bar chart illustrating the number of death cases and total confirmed cases based on the same dataset [12]. From Fig. 6 and Fig. 7, we can see that the number of infected patients in USA is alarming where almost 476000 people are affected. In USA, the recovered and the death numbers are 26050 and 17055, respectively. The number of deaths in Italy has crossed 18200 as of 10 April 2020. China was the first country which was diagnosed with COVID-19 case, but their recovery rate is considerably satisfactory. More than 77000 people have recovered in China from almost 82000 affected people.

![Covid-19 Confirmed Cases Around the World](image)

**Fig. 3. COVID-19 confirmed cases around the world till 10 April 2020**
Fig. 4. Percentage of deaths around the world till 10 April 2020

Fig. 5. Percentage of recovered cases around the world till 10 April 2020
3. CONCLUSION

The deadly novel coronavirus termed as SARS-CoV-2 has caused thousands of deaths across the world since December 2019. This COVID-19 has been declared as a pandemic and the whole world was not in more danger after World War II. Since no vaccine or antiviral treatment is available, the WHO has recommended that infection should be avoided by frequent hand washing, social distancing, keeping unwashed hands away from the face, and covering coughs and sneezes with a tissue or inner elbow. Since the number of confirmed cases and deaths are increasing every day, and the virus hotspot has changed several times, it is very difficult to completely describe the nature of COVID-19 with current data. Still this work provides data analytics and data visualization to describe different aspects of the disease using the currently available datasets. Based on the dataset used and the data analytics of this paper, it can be seen that the combination of fever and cough is one of the major indicators of carrying this virus. It is also found that many patients develop the symptoms within 14 days of exposure. Furthermore, the currently available data shows that males and elderly people are more affected by the disease. It is also shown that the number of confirmed infections is much more in countries which have low average temperature compared to countries with high
average temperature. It can be seen that although the disease started in China, currently China has managed to restrict the spread of the disease. On the other hand, Italy, Spain and the USA now have very high number of confirmed cases and deaths.

The discussions of this paper will contribute to the ongoing research on COVID-19. More investigation is required to gain a clear understanding of the disease and to find means to deal with it.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Randhawa GS, Soltsyak MP, El Roz H, de Souza CP, Hill KA, Kari L. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. Biorxiv; 2020.
2. Dey SK, Rahman MM, Siddiqi UR, Howlader A. Analyzing the Epidemiological Outbreak of COVID-19: A Visual Exploratory Data Analysis (EDA) Approach. Journal of Medical Virology; 2020.
3. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, Tai Y. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. The Lancet Respiratory Medicine; 2020.
4. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Agha R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery; 2020.
5. Lu H, Stratton CW, Tang YW. Outbreak of Pneumonia of Unknown Etiology in Wuhan China: the Mystery and the Mirage. Journal of Medical Virology; 2020.
6. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Yu T. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. The Lancet. 2020; 395(10223):507-513.
7. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, Wang M. Presumed asymptomatic carrier transmission of COVID-19. Jama; 2020.
8. Santarpia JL, et al. Transmission Potential of SARS-CoV-2 in Viral Shedding Observed at the University of Nebraska Medical Center, medRxiv 2020;03: 23.20039446. DOI:https://doi.org/10.1101/2020.03.23.20039446
9. Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Health. 2020;25(3): 278-280.
10. Available:https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-specimens.htmlLAST [Accessed on 29 March 2020]
11. Jin YH, Cai L, Cheng ZS, Cheng H, Deng T, Fan YP, Han Y. A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). Military Medical Research. 2020;7(1):4.
12. COVID-19 Coronavirus Pandemic. Available:https://www.worldometers.info/coronavirus/ LAST [Accessed on 13 April 2020]
13. Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A. Temperature and latitude analysis to predict potential spread and seasonality for COVID-19; 2020. [ISSRN 3550308]
14. Centers for Disease Control and Prevention. Frequently asked questions about SARS; 2005. Available:https://www.cdc.gov/sars/about/faq.html.Google Scholar
15. World Health Organization. Middle East respiratory syndrome coronavirus (MERS-CoV). Available:https://www.who.int/emergencies/mers-cov/en. Google Scholar
16. Mahase E. Coronavirus: Covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate; 2020.
17. Bukhari Q, Jameel Y. Will Coronavirus Pandemic Diminish by Summer?; 2020. [ISSRN 3556998]
18. Wang J, Tang K, Feng K, Lv W. High Temperature and High Humidity Reduce the Transmission of COVID-19; 2020. [ISSRN 3551767]
19. Araujo MB, Naimi B. Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate. Medrxiv; 2020.
20. COVID19_line_list_data.csv Available: https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset [Accessed on 28 March 2020]

21. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Lessler J. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. Annals of Internal Medicine; 2020.

22. Elisabeth M. Coronavirus: Covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate, BMJ. 2020;368:m641.

23. Sayre R, Karagulle D, Frye C, Boucher T, Wolff NH, Breyer S, Touval J. An assessment of the representation of ecosystems in global protected areas using new maps of World Climate Regions and World Ecosystems. Global Ecology and Conservation. 2020;21:e00860. DOI: https://doi.org/10.1016/j.gecco.2019.e00860

24. Average yearly temperature by country, Lebanese Economy Forum, Available: https://web.archive.org/web/20150905135247/http://lebanese-economy-forum.com/wdi-gdf-advanced-data-display/show/EN-CLC-AVRT-C/, Last [Accessed on 13 April, 2020]

© 2020 Khanam et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here: http://www.sdiarticle4.com/review-history/56278