Note: *CORONOSIS*: *Corona* Prognosis via a Global Lens to Enable Efficient Policy-making Both at Global and Local Levels

Ishrat Jahan Eliza
1605089@ugrad.cse.buet.ac.bd
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

Md Hasibul Husain Hisham
1605009@ugrad.cse.buet.ac.bd
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

Mohammad Nuwaisir Rahman
nuwaisir1998@gmail.com
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

Ajwad Akil
1605079@ugrad.cse.buet.ac.bd
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

Abir Mohammad Turza
abir1turza@gmail.com
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

Fahim Morshed
1605077@ugrad.cse.buet.ac.bd
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

Nazmus Sakib
nsakib280@gmail.com
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

Sriram Chellappan
sriramc@usf.edu
University of South Florida
Professor
South Florida, United States

A. B. M. Alim Al Islam
alim_razi@cse.buet.ac.bd
Bangladesh University of Engineering and Technology
Professor
Dhaka, Bangladesh

ABSTRACT

Epidemics and pandemics have been affecting human lives since time, and have sometimes altered the course of history. At this very moment, Coronavirus (COVID-19) pandemic has been the defining global health crisis. Now, perhaps for the first time in history, humanity as a whole has undergone major disruptions to life and some form of lockdown. New policies need to be forged by policy-makers for various sectors such as trading, banking, education, etc., to lessen losses and to heal quickly. For efficient policy-making, in turn, some prerequisites needed are historical trend analysis on the pandemic spread, future forecasting, the correlation between the spread of the disease and various socio-economic and environmental factors, etc. Besides, all of these need to be presented in an integrated manner in real-time to facilitate efficient policy-making. Therefore, in this work, we developed a web-based integrated real-time operational dashboard as a one-stop decision support system for COVID-19. In our study, we conducted a detailed data-driven analysis based on available data from multiple authenticated sources to predict the upcoming consequences of the pandemic through rigorous modeling and statistical analyses. We also explored the correlations between disease spread and diverse socio-economic as well as environmental factors. Furthermore, we presented how the outcomes of our work can facilitate both contemporary and future policy-making.

CCS CONCEPTS

• **Information systems** → *Decision support systems.*

ACM Reference Format:
Ishrat Jahan Eliza, Md Hasibul Husain Hisham, Mohammad Nuwaisir Rahman, Ajwad Akil, Abir Mohammad Turza, Fahim Morshed, Nazmus Sakib, Sriram Chellappan, and A. B. M. Alim Al Islam. 2022. Note: *CORONOSIS*: ***Corona*** Prognosis via a Global Lens to Enable Efficient Policy-making Both at Global and Local Levels. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '22), June 29-July 1, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3530190.3534831

1 INTRODUCTION

The COVID-19 pandemic has spawned problems across different aspects covering political [35], economical [29], social [29], psychological unrest [36], and immense panic [26] across the world. Mitigating these complex problems is a great challenge for world leaders and policy-makers, especially considering the scale of impact. In order to address these policy-making challenges and to provide useful insights from trends of disease spread, we developed a web-based integrated operational dashboard. Our work has two broad goals - 1) to design and develop a user-friendly real-time interactive dashboard, and 2) to predict future spreading of the pandemic, while also correlating pandemic spread with different socio-economic contexts we integrate all these in our dashboard.

Previously, transcribing raw data into information and visualizing them have been done for epidemics [76, 80] and pandemics [15, 77] including COVID-19 [31, 72, 81]. Besides, environmental

COMPASS '22, June 29-July 1, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9347-8/22/06…$15.00
https://doi.org/10.1145/3530190.3534831

conditions and socio-economic factors such as correlation of epidemics and pandemics with other factors such as humidity and temperature have been examined in previous epidemics and pandemics [9, 10, 37, 43, 46, 54, 68, 70]. Previous studies include long short-term memory network (LSTM) of a recurrent neural network (RNN) [84] as well as detection and prediction model based on Twitter data [3, 6, 7, 83] for Influenza pandemic and Dengue outbreak [57, 67]. There also exist map based [12, 87], machine learning based [66], image processing based [1, 38], and mathematical modeling based approaches to predict pandemics [39, 73, 74]. However, in order to make robust and sustainable policies in the case of a global pandemic, policy-makers need all the information together from a single source having more in-depth analyses and forecasting. Keeping this goal in our mind, we analyzed COVID-19 from different perspectives, such as social distancing, lockdown, trend of infection rates, statistical correlation, and more [18], and take all of them in a single platform. There exist related studies on real-time decision-making by fitting an epidemic model to observed and spatially-explicit infection data for the foot-and-mouth disease outbreak, influenza outbreak, Dengue epidemic, etc., [27, 59, 60, 63]. Additionally, in recent times, there are studies that focus on policy-making during COVID-19 pandemic [11, 75].

Going beyond, three tasks - 1) bringing sources of information together, 2) generating other relevant information out of them, and 3) presenting all useful information in an integrated manner - are of utmost significance for presenting data to policy-makers in an efficient manner. However, these tasks present several research challenges covering - 1) reliable collection of real-time pandemic spreading data from various sources (which can be in different formats), 2) conducting diverse analyses over different types of data relating to pandemic, environmental factors, socio-economic factors, etc., and 3) presenting outcomes of the analyses in an easy-to-understand manner and in real-time for future forecasting. This situation motivated us to create a fully functional and integrated real-time dashboard as shown in Figure 1, which would be a one-stop source for all such live useful information about the pandemic.

2 METHODOLOGY

Our methodology encompasses the process of collecting required data, platform design, relevant technologies to do so, and methods of analysis and analytics.

2.1 Methodology of System Development

The whole development architecture of the functional dashboard web application can be divided into four specific categories - 1) data collection, 2) analysis, 3) back end, and 4) front end (Figure 2).

2.1.1 Data Collection and Analysis. We collected global live update from Covid19.Mathdro.id API [48], data for global prediction from Corona-API [56] and historical data for country-wise prediction from covid19-api.org [4]. We collected data of medical facilities (BEDS/10M) from WHO [79], current death rate, and current active rate from NovelCOVID API [28], humidity from Kaggle [34], temperature, rainfall, GDP, literacy rate from DataBank [30], area, population, and population density from REST Countries API [5]. We faced challenges during data collection bacause of the instability of the APIs. For our analyses, we transformed the cumulative time series data to represent the new cases of each day. We used Microsoft Excel and Python programming language to perform preprocessing. After performing correlation analysis and prediction based on analytics, we pushed data into the database so that back end team could process them to visualize onto the front end.

2.1.2 Back End. The data that we obtained from external API calls or database query can be divided into three broad categories: (1) constantly changing data, (2) seldom changing data, and (3) almost constant or fixed data. We designed two types of URL routes or API endpoints to handle these different types of data. Instead of storing the dynamically changing data into the database, these routes bypassed the data to the receiving end of them (category 1). At the primary level, we filtered and pushed to the database some important data such as GDP, literacy rate, average annual temperature and rainfall, average humidity of the countries, population, country codes, and country area (category 3). We have been pushing prediction data to the database everyday (category 2). Besides, we populate our database daily with the daily statistical summary of COVID-19 of all countries in the world with the help of an automated script.

2.1.3 Front End. We used React and Material-UI to develop our user interface. We called the API routes in the respective React Components (Charts, Tables, and Maps) to retrieve data from the back end. We mainly used React-chartjs-3 plugin to draw the charts to represent predictions related to pandemic, analyses of different factors, and affects of those factors. We parameterized some of the graphs according to country using "InputForm" of Material-UI. We used Material-UI plugin for displaying informative tables that show predictions and latest conditions of different countries. There are two types of map to be displayed: 1) World Street Map and 2) Choropleth Map for which we used React-leaflet and Esri tile-server. World street map is necessary to visualize the country-wise spreading and choropleth map is necessary for visualizing spreading through other contexts such as, temperature, humidity, population density, etc.

2.2 Analysis

We analysed the data through a process of cleaning, transforming, and modeling them to discover useful information. Our data analysis process includes trend analysis and statistical analysis methods.

2.2.1 Trend Analysis. We inspected the trend of COVID-19 affected rate and COVID-19 death rate over time for different countries. After collecting corresponding data from various sources [41, 42], we fixed a range of consecutive days, Δt. Afterwards, we fitted the daily affected graph during this Δt time period for a particular country to a seventh degree polynomial $f(t)$ using the polyfit tool [23]. We did the same for daily death graph. We specifically used the degree of seven to come up with a solution of the bias-variance trade-off [13]. We solved this trade-off by our own methodology. We calculated the residual sum of squares (RSS) after fitting a country's daily affected graph to the polynomials of degree 2 to degree 15. We took the average of these residuals over the selected countries. We show these residuals in Figure 3(a). Besides, we calculated the runtimes of our overall calculations against the degree of polynomials and took the average over the selected countries. We show these

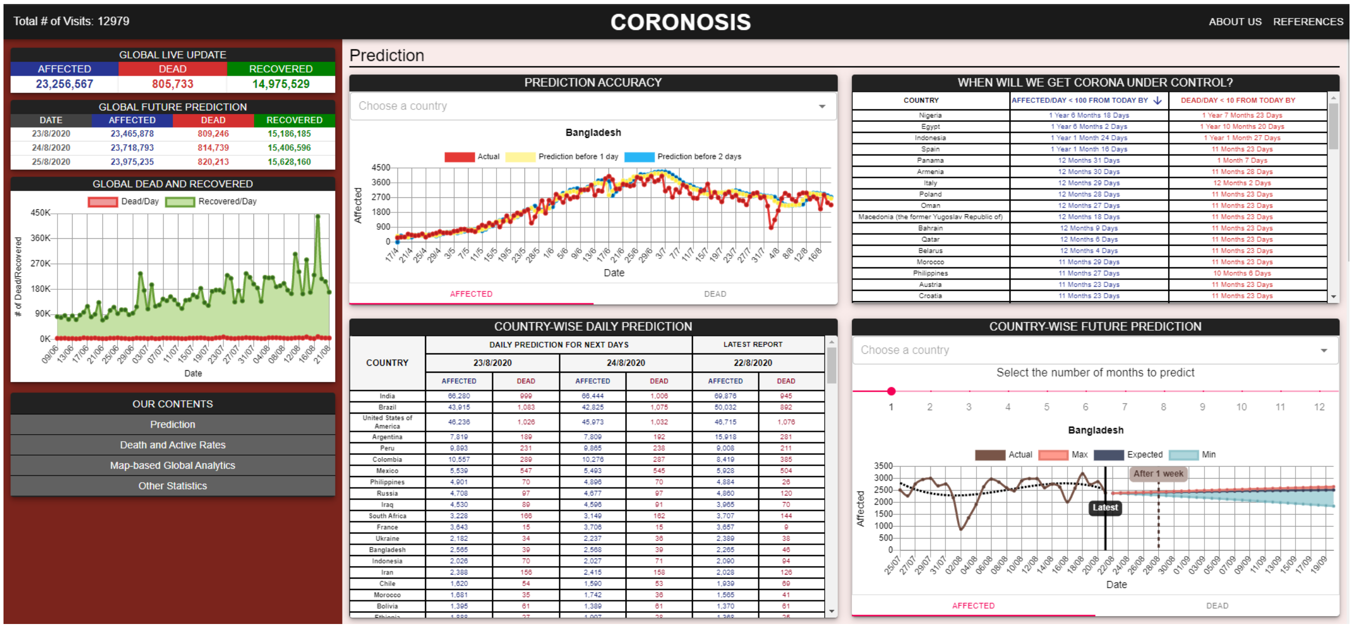


Figure 1: CORONOSIS - Version 3 (launched on July 20, 2020)

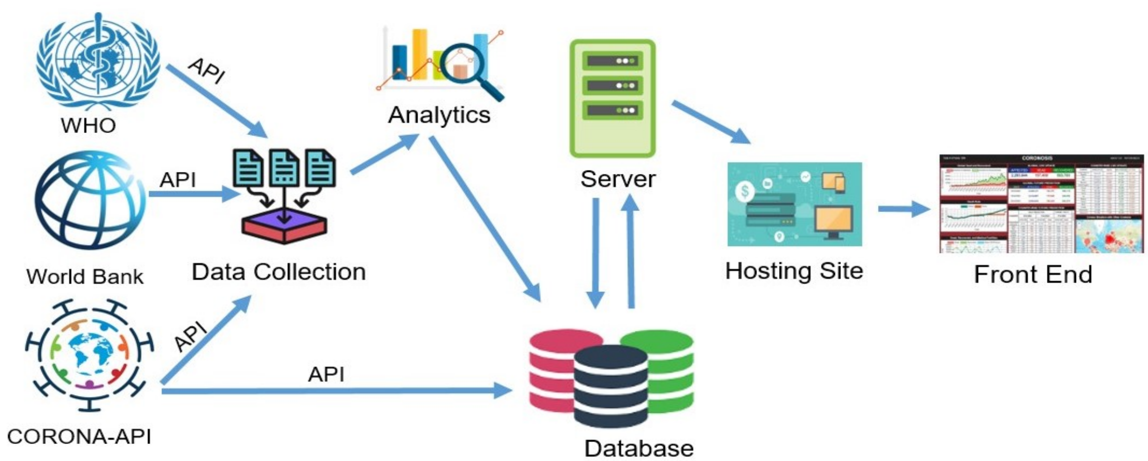
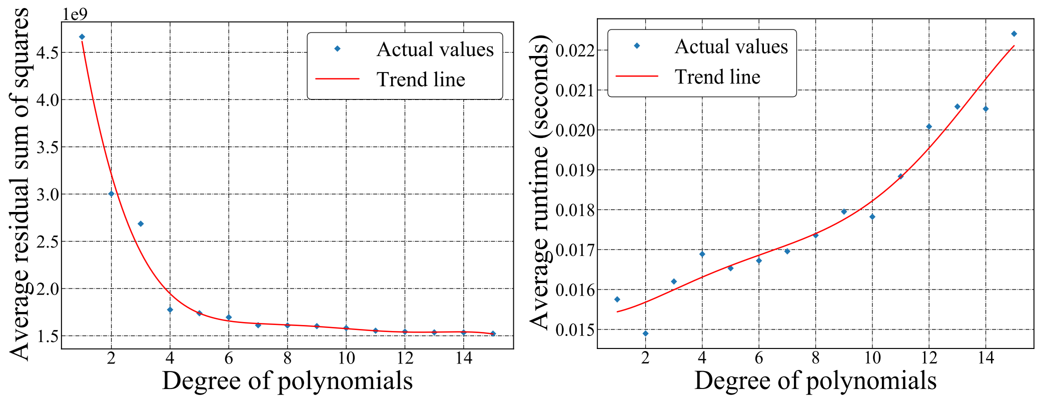


Figure 2: Methodology of system development in our study

runtimes in Figure 3(b). From Figure 3, it is evident that we have to make a trade-off in choosing the degree number. Increasing the degree can lead to the overfitting of the data [86]. Besides, there is a clear increasing trend on the runtime graph. In order to solve this trade-off between residual sum of squares and runtime, we took the degree of seven. From Figure 3, the RSS trend line takes the first flattening shape around the value of seven. Besides, the average runtime of the calculation is also much less at the degree of seven.

We determined the first order derivative $f'(t)$ of $f(t)$ using the polyder tool [22]. From $f'(t)$, we took the average value of affected rate or death rate. We defined *Average Rate* as the mean value of the sixth degree polynomial $f'(t)$ for different equidistant values of t taken between the range of Δt. We selected Argentina, Bangladesh, Brazil, Colombia, India, Mexico, Russia, South Africa, and United States to analyse the affected and death rate. We specifically selected these countries because these are the ones which are affected mostly by COVID-19 and still the infection is spreading [81]. We selected $\Delta t = 140$ days for our calculation, from March 27, 2020 to August 13, 2020. During this time period, all our selected countries have commonly reported significant numbers of COVID-19 cases [81].

2.2.2 Statistical Analysis. Our statistical analysis process [32, 49] included analyzing on COVID-19 datasets to find out correlation of spreading of COVID-19 with other contexts. We collected humidity and temperatures data from January 22, 2020 to April 01, 2020 and



(a) (b)

Figure 3: Average residual and runtime for degree 2 to degree 15

for 114 countries. To find correlation between spreading of COVID-19 and temperature/ humidity, we took country-wise daily affected cases as dependent variable Y and maximum temperature/ humidity five days prior [44] to that day as independent variable X. We created scatter plot of the two variables to see whether there is any existing trend or not. We calculated Pearson's correlation, Spearman's rank correlation, and Kendall's rank correlation coefficient using SciPy tool [19–21]. To observe a correlation with pollution, healthcare, food security, and population tests, we used average data of socio-economic factors and the latest data of COVID-19 cases. Our latest data for COVID-19 cases for analysis and analytics is of August 13, 2020. Every country has one data point for socio-economic analysis. We determined the median of correlations for getting a worldwide overview. We collected data of 108 countries for pollution index and 80 countries for healthcare index from Numbeo [50]. We collected food security index data of 113 countries from the Global Food Security Index [45]. Population tests data is collected from Novel COVID API [28] and death rate of every single country is calculated using the formula:

$$Death\ rate = \frac{Total\ number\ of\ deaths}{Total\ number\ of\ affected} \tag{1}$$

2.3 Analytics

Our analytics incorporated analyzing COVID-19 datasets followed by predicting the transmission dynamics of COVID-19. We started off with a basic compartmental model of epidemiology, the *SIR* model [55]. This model breaks down the population into three compartments; S: number of Susceptible, I: number of Infectious, and R: number of Removed individuals. Using these compartments, SIR predicts a contagious disease dynamics using a set of three ordinary differential equations [55]. But this model assumes constant population throughout the epidemic time. It also assumes homogeneous mixing of people and no reinfection of recovered individuals. Besides, the model does not take age, social mixing, and race into consideration. To address all these issues, first we broke down the Removed compartment into *Recovered* and *Dead* compartments. Then, we got the basic *SIRD* model which could address the difference of the recovered and dead individuals [61].

Then we interpreted the lockdown and social distancing measures. We aggregated the total population N into $M = 16$ groups with respect to 5-year age interval, ranging 1-80 years. Each group represents a class of corresponding aged individuals. We got the total population N as follows: [71].

$$N = \sum_{i=1}^{M} S_i(t) + I_i(t) + R_i(t) + D_i(t) = \sum_{i=1}^{M} N_i \tag{2}$$

Where S_i, I_i, R_i, and D_i is the number of susceptible, infectious, recovered, and dead individuals of class i.

Finally, we used social contact structures among all possible classes. We considered the social mixing patterns at home, school, working place, and other locations [58]. Let C be the contact matrix where element C_{ij} represents the average number of contacts made per day by an i-class individual with j-class individuals. We partitioned C_{ij} into C_{ij}^H, C_{ij}^S, C_{ij}^W, and C_{ij}^O, representing the contact structures at home, school, working place, and other locations respectively. We got the element C_{ij} as follows.

$$C_{ij} = \delta_1 C_{\mathrm{ij}}^{\mathrm{H}} + \delta_2 C_{\mathrm{ij}}^{\mathrm{S}} + \delta_3 C_{\mathrm{ij}}^{\mathrm{W}} + \delta_4 C_{\mathrm{ij}}^{\mathrm{O}} \tag{3}$$

Where δ_1, δ_2, δ_3, and δ_4 are the indicators of mixing patterns at different places. We vary these indicators according to different states of lockdown and social distancing [65]. Accordingly, we present the formulation of our mathematical model.

$$\begin{aligned}
\frac{dS_i}{dt} &= -\beta \sum_{j=1}^{M} C_{ij} \frac{I_j}{N_j} S_i(t) I_i(t) + \mu N_i - \Gamma S_i(t) + \zeta(t)\Gamma R_i(t) \\
\frac{dI_i}{dt} &= \beta \sum_{j=1}^{M} C_{ij} \frac{I_j}{N_j} S_i(t) I_i(t) - (\lambda_d + \lambda_r) I_i(t) - \Gamma I_i(t) \\
\frac{dR_i}{dt} &= \lambda_r I_i(t) - \Gamma R_i(t) \\
\frac{dD_i}{dt} &= \lambda_d I_i(t) - \Gamma D_i(t).
\end{aligned} \tag{4}$$

Where β represents the transmission rate of infection, λ_r and λ_d represent the recovery rate and death rate of the infected individuals respectively. To determine the constants (β, λ_d, and λ_r), we ran the basic SIRD model on the existing historical data of the corresponding country. µdenotes birth rate and (Γ) denotes death rate, making the model compatible with vital dynamics. (ζ(t)) denotes a time dependent reinfection rate.

3 RESULTS

Analyses with the processed data lead us to some interesting insights. Our analysis encompassed trend analysis, prediction analysis for some specific countries, and some statistical tests and the results are illustrated below.

3.1 Trend Analysis

We present our findings of trend analysis according to the methodology described in Section 2.2.1. Upon collecting appropriate data from various sources [41, 42], we performed the analysis using our numerical methods.

3.1.1 Affected Rate. We classified our selected nine countries into three categories based on *Average Affected Rate*, conforming to the *Average Rate* defined in Section 2.2.1. Countries having *Average Affected Rate* greater than 100 are classified as *High Rated*, *Average Affected Rate* greater than 45 but less than 100 are classified as *Medium Rated*, and *Average Affected Rate* less than 45 are classified as *Low Rated*. High rated countries are United States, Brazil, and India having 197.836, 334.190, and 463.903 *Average Affected Rate* respectively. Medium Rated countries are Mexico, Columbia, and Argentina having 45.981, 82.722, and 49.665 *Average Affected Rate* respectively. Low Rated countries are Bangladesh, Russia, and South Africa having 20.373, 38.454, and 39.952 *Average Affected Rate* respectively.

3.1.2 Death Rate. Similarly, we classified the nine countries into three categories based on *Average Death Rate*, conforming to the *Average Rate* defined in Section 2.2.1. Countries having *Average Death Rate* greater than 5.5 are classified as *High Rated*, *Average Death Rate* greater than 2 but less than 5.5 are classified as *Medium Rated*, and *Average Death Rate* less than 2 are classified as *Low Rated*. High rated countries are United States, India, and Brazil having 5.601, 6.249, and 7.566 *Average Death Rate* respectively. Medium rated countries are Colombia, and Mexico having 2.298, and 5.318 *Average Death Rate* respectively. Low rated countries are Bangladesh, Russia, Argentina, and South Africa having 0.269, 0.776, and 1.359, and 1.479 *Average Death Rate* respectively.

3.2 Correlation with Other Contexts

As mentioned in Section 2.2.2, we find out different correlations between COVID-19 transmission dynamics such as daily affected, daily recovered, and daily deaths with different environmental and socio-economic factors. To determine whether the correlation between the variables is significant, we compared the p-value [25] to our defined significance level. We used 0.05 as our significance level α[24, 82].

The correlation coefficients for environmental data showed skewed distribution, we used the median as a parameter of the overall correlation coefficient of the whole world [14]. The median values are small and positive which means, though daily affected cases of some countries show a statistically significant correlation with temperature, the overall impact of it on COVID-19 is negligible. The median value of the humidity-affected correlation coefficient of all the countries is significant and negative which means daily affected cases of some countries show a statistically significant correlation with humidity. If humidity increases, a decreasing trend of affected cases is found. The overall impact of it on COVID-19 is neither negligible nor strong enough to draw a conclusion. The results are summarized into the Table 1.

According to our result of correlations from Table 1, we find that the overall correlation coefficient for pollution and total affected is negative and implies very weak relationship. Spearman's coefficient and Kendall's coefficient are also very weak and unimportant [2]. Pearson's coefficient of food security-affected correlation is positive but weak. Spearman's and Kendall's coefficient values indicate that the variables are moderately correlated [62]. The Pearson's, Spearman's, and Kendall's correlation coefficient for healthcare index are negative and very weak. The overall value of the population tests-death rate correlation coefficient seems positive correlation but too small to draw a conclusion [47].

3.3 Future Prediction

We ran our mathematical model of future prediction on different countries by collecting affected, dead, recovered, and demographic data [51, 64] from various sources. We predicted the number of affected and the number of dead for the next 60 consecutive days from the latest data. We tweaked the lockdown and social distancing indicating parameters (δ_1, δ_2, δ_3, and δ_4) to get three different states: Lockdown Imposed and Social Distancing Maintained, Lockdown Released and Social Distancing Maintained, and Lockdown Released and Social Distancing Discarded. We present our results by running the model on Bangladesh, Brazil, India, Iraq, Mexico, Russia, South Africa, and United States. Iraq is showing a nearly exponential trend on the cases per day curve [81]. India and Brazil have seen a lot of cases and their curves are still giving a positive trend [81]. Meanwhile, Mexico and Bangladesh seem to flatten their curves, yet the infection is spreading to some extent [81]. Russia and South Africa seem to flatten their curves by that time and was giving a negative trend [81]. Besides, all of these countries are heavily populated, bearing a better demographic significance than less populated countries. Moreover, all these eight countries represent five clusters altogether, based on the causes of death and health risk factors [85]. We illustrate the results in Figure 4. We find that Brazil, India, Iraq, Mexico, Russia, and South Africa will get a negative slope in the graph within two months, if the authorities impose lockdown and the countrymen maintain social distancing. However, the curve of United States and Bangladesh still remains flattened to some extent, even after two months of imposed lockdown and maintained social distancing.

All countries except Mexico and Russia get an exponential trend on the graph, if the authorities release lockdown and the countrymen discard social distancing. We also measure the impacts when

Table 1: Correlation between COVID-19 datasets and different environmental, socio-economic factors

Correlating factors	Pearson's coefficient		Spearman's coefficient		Kendall's coefficient	
	Range	Median	Range	Median	Range	Median
Temperature and Daily Affected	-0.0019 to 0.6900	0.0866	-0.0009 to 0.8023	0.0930	-0.0050 to 0.6526	0.0777
Humidity and Daily Affected	-0.0093 to 0.5418	-0.4951	-0.0098 to 0.5189	-0.4366	-0.0018 to 0.4011	-0.3155
Pollution and Total Affected	-	-0.030	-	-0.004	-	0.001
Food security and Total Affected	-	0.190	-	0.502	-	0.334
Healthcare and Total Recovered	-	-0.062	-	-0.059	-	-0.045
Population tests and Death rate	-	0.135	-	0.170	-	0.116

the authorities release lockdown and the countrymen maintain social distancing. Our finding is, this is a balanced strategy. Indeed, it has a less effective spreading impact compared to imposed lockdown and maintained social distancing. Yet, this measure might be the key to solve the economic crisis [53] and mental health crisis [8] due to lockdown. We illustrate the impacts of lockdown and social distancing on daily number of deaths in Figure 5. It almost follows the same pattern as the affected curves.

Vital dynamics contribute to susceptible compartment with $\chi = (\mu - \Gamma) \times N$ number of people. Since χ is mostly positive [78], it only adds χnumbers of people to the compartment. However, it still remains a research question [17] whether newborn babies not inheriting COVID-19 from the mother are actually susceptible to the virus or not. Besides, in vital dynamics, we are considering the number of accident deaths that would happen in a normal continued lifestyle. Although, it might compensate the numbers of unreported deaths due to COVID-19 [88]. The reinfection parameter $\zeta(t)$ has very insignificant effects in the results. Although we considered the highest value of $\zeta(t)$ from the study [52], this value is very small and bear minimum consequences. If it were significant enough, we would get a clear oscillating trend in the graphs. However, if the reinfection rate increases in the future due to the mutation of the virus [89], this parameter may become significant.

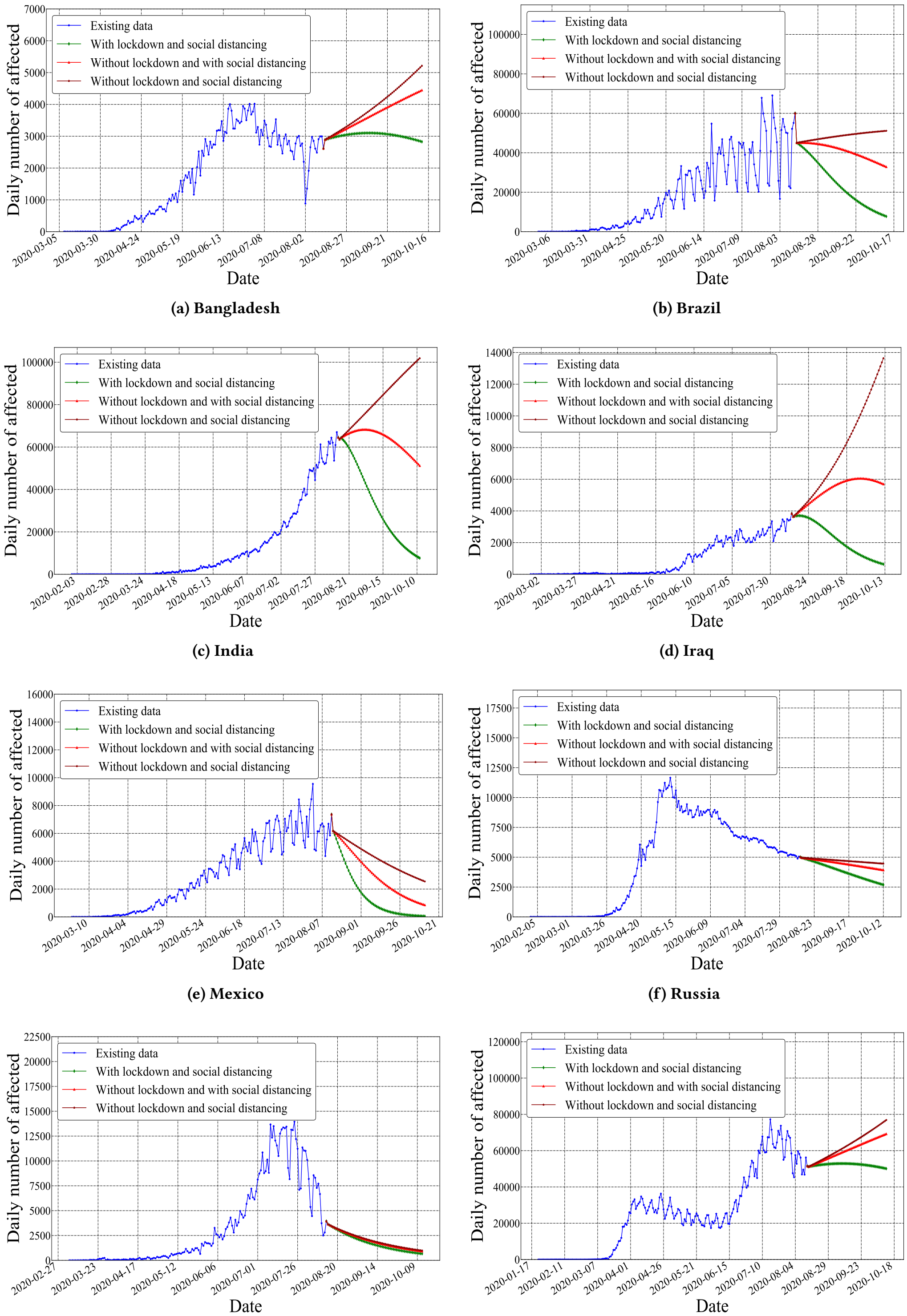
3.4 Prediction Accuracy

RMSPE gives large error scores due to having outliers [33, 40] in historical datasets. Hence, we adopted MAPE as our error measurement mechanism [69]. For some countries, we found some sudden up or down to the number of affected and death cases and reduced the effect of these outliers by omitting them. We calculated the absolute differences of COVID-19 cases of consecutive days and summed up the differences. After that, we took the mean of these and checked whether the individual difference is greater or equal to 1.65 times of mean or not [16]. If the difference of the values met up the benchmark, we marked it as an outlier and omitted the second date from the calculation. For instance, the absolute difference of affected cases for India of June 10, 2020 and June 09, 2020 is 10,523. The mean of absolute differences between these consecutive affected cases is 2,482. 1.65 times of 2,482 is 4,095 and 10,523 is greater than 4,095. Therefore, we marked the data of June 10, 2020 as an outlier and ignored the data. We did the same calculations for both 1-day and 2-day prediction, and for death cases as well. MAPE includes division of deviated value by the actual value. As division by 0 causes the error measurement to be undefined, we ignored those cases where the value of actual cases is 0. We found some negative value from the data source and we cut off the value forcefully to 0 as negative number of affected cases does not make any sense. Ignorance of these values caused a reduction in data points. We did our error calculations on 80%, 82%, 89%, 85%, 81%, 80%, 84%, and 84% of data for Bangladesh, Brazil, India, Iraq, Mexico, Russia, South Africa, and United States respectively. The values of average percentage error of 1-day prediction and 2-day prediction for Bangladesh, Brazil, India, Iraq, Mexico, Russia, South Africa, and United States are illustrated in Table 2.

We performed error measurements on death cases prediction values for Bangladesh, Brazil, India, Iraq, Mexico, Russia, South Africa, and United States. As mentioned earlier, we did some preprocessing to reduce outliers, zero numbers of cases, and negative numbers of cases. We did our error calculations on 80%, 80%, 91%, 82%, 80%, 80%, 79%, and 80% of data for Bangladesh, Brazil, India, Iraq, Mexico, Russia, South Africa, and United States respectively. The values of error of 1-day prediction and 2-day prediction for death cases for Bangladesh, Brazil, India, Iraq, Mexico, Russia, South Africa, and United States are illustrated in Table 2.

4 CONCLUSION AND FUTURE WORK

Effective policy-making is considered to be the most important task during the COVID-19 pandemic, as it determines the overall functionality of a region during and after the pandemic. An important basis of effective policy-making is insightful data, which is even more important in the case of a pandemic due to its spanning over the spatio-temporal domains. Accordingly, in this work, we focused on how to provide authentic and reliable comprehensive data to policy-makers. To do so, we explored correlations between the pandemic spreading and a number of different socio-economic and environmental contexts. Besides, we demonstrated a robust future prediction of the pandemic over different regions of the world in different time frames. We presented outcomes of all the work in an integrated manner in our newly-developed dashboard. Our study is completely data-driven. However, due to insufficiency and unavailability of some crucial data, we could not make our work more comprehensive. For example, we were in a great dearth of sophisticated hospitalization data, lockdown schemes, and schedule data, asymptomatic patient data, data of innate immunity against COVID-19, etc. We could make our prediction model more robust using these data. In future, we aim to perform such a more rigorous study.



(a) Bangladesh

(b) Brazil

(c) India

(d) Iraq

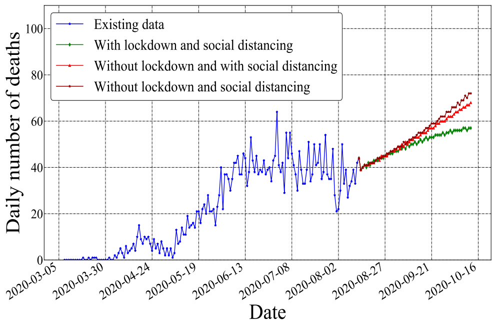
(e) Mexico

(f) Russia

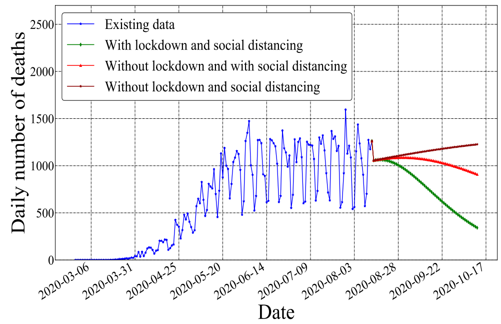
(g) South Africa

(h) United States

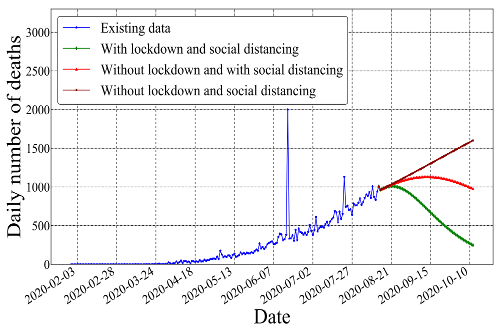
Figure 4: Prediction of daily number of affected for the next 60 days from the data as found on August 13, 2020



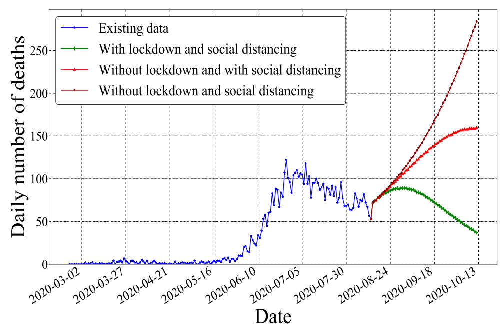
(a) Bangladesh



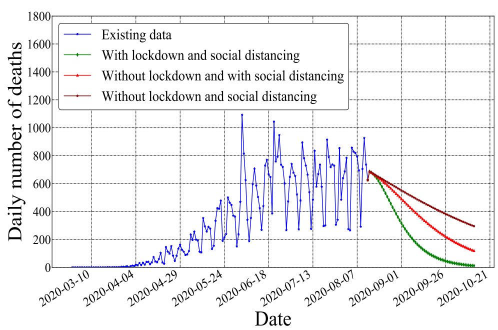
(b) Brazil



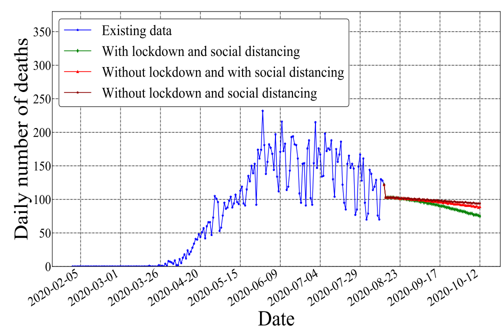
(c) India



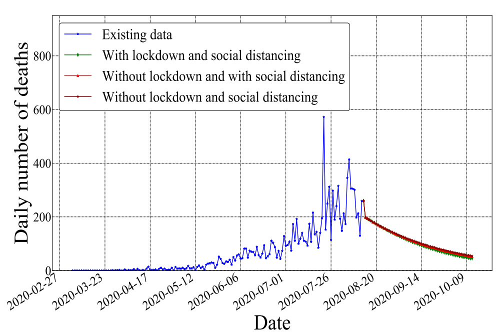
(d) Iraq



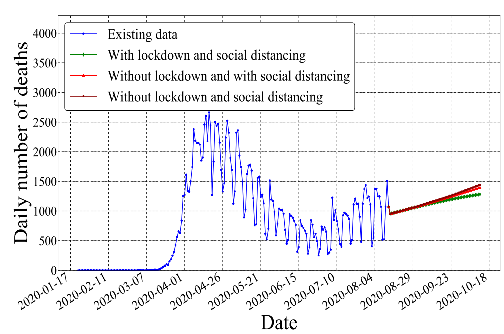
(e) Mexico



(f) Russia



(g) South Africa



(h) United States

Figure 5: Prediction of daily number of deaths for the next 60 days from the data as found on August 13, 2020

Table 2: Average percentage error scores for 1-day and 2-day predictions

Country	Affected		Dead	
	1-day prediction	2-day prediction	1-day prediction	2-day prediction
Bangladesh	12%	14%	26%	25%
Brazil	36%	42%	29%	33%
India	12%	12%	19%	20%
Iraq	7%	10%	13%	14%
Mexico	24%	25%	41%	45%
Russia	6%	6%	20%	21%
South Africa	22%	24%	38%	40%
United States	14%	16%	40%	46%
	Avg = 17%	Avg = 19%	Avg = 28%	Avg = 31%

ACKNOWLEDGMENTS

The work has been conducted at and supported by Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh.

REFERENCES

[1] Ahmed Abdullah Farid, Gamal Selim, and Hatem Khater. 2020. A Novel Approach of CT Images Feature Analysis and Prediction to Screen for Corona Virus Disease (COVID-19). *International Journal of Scientific and Engineering Research* 11 (March 2020), 1141. https://doi.org/10.14299/ijser.2020.03.02

[2] Haldun Akoglu. 2018. User's guide to correlation coefficients. *Turk J Emerg Med* 18(3) (September 2018), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

[3] Ali Alessa and Miad Faezipour. 2018. A review of influenza detection and prediction through social networking sites. *Theoretical Biology and Medical Modelling* 15 (February 2018), 2. https://doi.org/10.1186/s12976-017-0074-5

[4] COVID-19 API. 2020. Free COVID-19 API. https://documenter.getpostman.com/view/10877427/SzYW2f8n?version=latest. [last accessed August 13, 2020].

[5] apilayer. 2019. restcountries. https://restcountries.eu/. [last accessed August 13, 2020].

[6] Jason Asher and Leidos. 2017. Forecasting Ebola with a Regression Transmission Model. *Epidemics* 22 (February 2017), 50–55. https://doi.org/10.1016/j.epidem.2017.02.009

[7] Gagangeet Aujla. 2014. Prediction Model for Influenza Epidemic Based on Twitter Data. *International Journal of Advanced Research in Computer and Communication Engineering* 3 (July 2014), 7541–7545.

[8] Md. Hasan Banna, Abu Sayeed, Satyajit Kundu, Enryka Christopher, M.Tasdik Hasan, Musammet Begum, Tapos Kormoker, Shekh Dola, Md Hassan, Sukanta Chowdhury, and md. shafiqul islam Khan. 2020. The impact of the COVID-19 pandemic on the mental health of the adult population in Bangladesh: a nationwide cross-sectional study. *International Journal of Environmental Health Research* 0 (08 2020), 1–12. https://doi.org/10.1080/09603123.2020.1802409

[9] Muhammad Farhan Bashir, Benjiang Ma, . Bilal, Bushra Komal, Muhammad Bashir, Duojiao Tan, and Madiha Bashir. 2020. Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Science of The Total Environment* 728 (April 2020), 138835. https://doi.org/10.1016/j.scitotenv.2020.138835

[10] . Bilal and Bushra Komal. 2020. Correlation between environmental pollution indicators and COVID-19 pandemic: A brief study in Californian context. *Environmental Research* 187 (May 2020), 109652. https://doi.org/10.1016/j.envres.2020.109652

[11] Gian Borasio, Claudia Gamondi, Monika Obrist, Ralf Jox, and for ch. 2020. COVID-19: decision making and palliative care. *Swiss Medical Weekly* 150 (March 2020). https://doi.org/10.4414/smw.2020.20233

[12] André Botha and Wynand Dednam. 2020. A simple iterative map forecast of the COVID-19 pandemic. arXiv:2003.10532 [physics.soc-ph]

[13] Erica Briscoe and Jacob Feldman. 2011. Conceptual complexity and the bias/variance tradeoff. *Cognition* 118, 1 (2011), 2–16.

[14] James Dean Brown. 1997. Skewness and kurtosis. *Shiken: JALT testing & evaluation SIG* 1 (1997), 20–23.

[15] Visual Capitalist. 2020. Visualizing the History of Pandemics. https://www.visualcapitalist.com/history-of-pandemics-deadliest/. [last accessed March 14, 2020].

[16] Chris Chatfield. 1988. *Problem Solving: A Statistician's guide.* A Chapman and Hall Book, United States.

[17] Yan Chen, Hua Peng, Lin Wang, Yin Zhao, Lingkong Zeng, Hui Gao, and Yalan Liu. 2020. Infants Born to Mothers With a New Coronavirus (COVID-19). *Frontiers in Pediatrics* 8 (2020), 104. https://doi.org/10.3389/fped.2020.00104

[18] Bernard Choi, Tikki Pang, Vivian Lin, Pekka Puska, Gregory Sherman, Michael Goddard, Michael Ackland, Peter Sainsbury, Sylvie Stachenko, Howard Morrison, and Clarence Clottey. 2005. Can scientist and policy makers work together? *Journal of epidemiology and community health* 59 (September 2005), 632–7. https://doi.org/10.1136/jech.2004.031765

[19] The Scipy community. 2008-2009. scipy.stats.kendalltau. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kendalltau.html. [accessed Dec 19, 2019].

[20] The Scipy community. 2008-2009. scipy.stats.pearsonr. https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.pearsonr.html. [last accessed May 11, 2014].

[21] The Scipy community. 2008-2009. scipy.stats.spearmanr. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.spearmanr.html. [accessed Dec 19, 2019].

[22] The SciPy community. 2020. numpy.polyder. https://numpy.org/doc/stable/reference/generated/numpy.polyder.html. [last accessed August 5, 2020].

[23] The SciPy community. 2020. numpy.polyfit. https://numpy.org/doc/stable/reference/generated/numpy.polyfit.html. [last accessed August 5, 2020].

[24] Tukur Dahiru. 2008. P – VALUE, A TRUE TEST OF STATISTICAL SIGNIFICANCE? A CAUTIONARY NOTE. *Ann Ib Postgrad Med* 6(1) (June 2008), 21–26.

[25] Kenneth P. Burnham David R. Anderson and William L. Thompson. 2000. Null Hypothesis Testing: Problems, Prevalence, and an Alternative. *The Journal of Wildlife Management* 64(4) (October 2000), 912–923. https://doi.org/10.2307/3803199

[26] Anneliese Depoux, Sam Martin, Emilie Karafillakis, Raman Preet, Annelies Wilder-Smith, and Heidi Larson. 2020. The pandemic of social media panic travels faster than the COVID-19 outbreak. *Journal of Travel Medicine* 27 (March 2020). https://doi.org/10.1093/jtm/taaa031/5775501

[27] Arin Dutta. 2008. The effectiveness of policies to control a human influenza pandemic : a literature review. *Policy Research Working Paper Series* 1 (February 2008). https://doi.org/10.1596/1813-9450-4524

[28] JULIAN PUFLER ELITEDAMYTH RAVY, ETHAN WINTERS. 2020. Novel COVID API. https://disease.sh/v2/countries. [last accessed August 13, 2020].

[29] Olaniyi Evans. 2020. Socio-economic impacts of novel coronavirus: The policy solutions. *BizEcons Quarterly, Strides Educational Foundation* 7 (March 2020), 3–12.

[30] The World Bank Group. 2020. DataBank. https://databank.worldbank.org/home.aspx. [last accessed August 13, 2020].

[31] Han. 2020. CoronaTracker. coronatracker.com. [last accessed May 28, 2020].

[32] William Lee Hays. 1973. *Statistics for the Social Sciences.* Holt and Rinehart and Winston, Canada.

[33] Amy Overbay Jason W. Osborne. 2004. The Power of Outliers (and Why Researchers Should Always Check for Them). *Practical Assessment, Research & Evaluation* 9(1) (January 2004). https://doi.org/10.7275/qf69-7k43

[34] kakumanu Sudhir. 2020. Country-wise weather data for covid19. https://www.kaggle.com/ksudhir/weather-data-countries-covid19. [last accessed April 3, 2020].

[35] Solmaz Filiz Karabag. 2020. An Unprecedented Global Crisis! The Global, Regional, National, Political, Economic and Commercial Impact of the Coronavirus Pandemic. *Journal of Applied Economics and Business Research* 10(1) (March 2020), 1–6.

[36] Suliman Khan, Rabeea Siddique, Hongmin Li, Ashaq Ali, Muhammad Shereen, Nadia Bashir, and Mengzhou Xue. 2020. Impact of coronavirus outbreak on psychological health. *Journal of Global Health* 10 (April 2020). https://doi.org/10.7189/jogh.10.010331

[37] Krishnamurthy Konduru, Amy Shurtleff, Sina Bavari, and Gerardo Kaplan. 2018. High degree of correlation between Ebola virus BSL-4 neutralization assays and pseudotyped VSV BSL-2 fluorescence reduction neutralization test. *Journal of virological methods* 254 (January 2018), 1–7. https://doi.org/10.1016/j.jviromet.2018.01.003

[38] Rahul Kuamr, Ridhi Arora, Vipul Bansal, Vinodh Joseph, Himanshu Buckchash, Javed Imran, Narayanan Narayanan, Ganesh Pandian, and Balasubramanian Raman. 2020. Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers. https://doi.org/10.1101/2020.04.13.20063461

[39] Toshikazu Kuniya. 2020. Prediction of the Epidemic Peak of Coronavirus Disease in Japan, 2020. *Journal of Clinical Medicine* 9 (March 2020), 789. https://doi.org/10.3390/jcm9030789

[40] Sang Kyu Kwak and Jong Hae Kimcorresponding. 2017. Statistical data preparation: management of missing values and outliers. *Korean J Anesthesiol* 70(4) (August 2017), 407–411. https://doi.org/10.4097/kjae.2017.70.4.407

[41] Global Change Data Lab. 2020. Coronavirus (COVID-19) Cases. https://ourworldindata.org/covid-cases. [last accessed August 13, 2020].

[42] Global Change Data Lab. 2020. Coronavirus (COVID-19) Deaths. https://ourworldindata.org/covid-deaths. [last accessed August 13, 2020].

[43] Béatrice Labrosse, Mathieu Tourdjman, Raphaël Porcher, Jerome Legoff, Xavier Lamballerie, François Simon, Jean-Michel Molina, and François Clavel. 2010. Detection of Extensive Cross-Neutralization between Pandemic and Seasonal A/H1N1 Influenza Viruses Using a Pseudotype Neutralization Assay. *PloS one* 5 (June 2010), e11036. https://doi.org/10.1371/journal.pone.0011036

[44] Stephen Lauer, Kyra Grantz, Qifang Bi, Forrest Jones, Qulu Zheng, Hannah Meredith, Andrew Azman, Nicholas Reich, and Justin Lessler. 2020. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine* 172 (March 2020), 577–582. https://doi.org/10.7326/M20-0504

[45] The Economist Intelligence Unit Limited. 2020. Global Food Security Index. https://foodsecurityindex.eiu.com/index. [last accessed August 13, 2020].

[46] Anice C. Lowen and John Steel. 2014. Roles of Humidity and Temperature in Shaping Influenza Seasonality. *Journal of Virology* 88, 14 (2014), 7692–7695. https://doi.org/10.1128/JVI.03544-13 arXiv:https://jvi.asm.org/content/88/14/7692.full.pdf

[47] MaŁgorzata Tartanus Marcin Kozak, Wojtek Krzanowski. 2012. Use of the correlation coefficient in agricultural sciences: problems, pitfalls and how to deal with them. *Anais da Academia Brasileira de Ciências* 84(4) (December 2012), 1147–1156. https://doi.org/10.1590/S0001-37652012000400029

[48] Mathdroid MIT License. 2020. https://covid19.mathdro.id/api. [last accessed August 13, 2020].

[49] Lincoln E. Moses. 1986. *Think and Explain with Statistics.* Addison-Wesley, Australia.

[50] Numbeo. 2009-2020. Numbeo API Documentation. https://www.numbeo.com/api/doc.jsp. [last accessed August 13, 2020].

[51] U.S. Department of Commerce. 2019. International Programs. https://www.census.gov/data-tools/demo/idb/informationGateway.php. [last accessed August 11, 2020].

[52] Alexander Victor Okhuese. 2020. Estimation of the Probability of Reinfection With COVID-19 by the Susceptible-Exposed-Infectious-Removed-Undetectable-Susceptible Model. , e19097 pages. https://doi.org/10.2196/19097

[53] Peterson Ozili and Thankom Arun. 2020. Spillover of COVID-19: impact on the Global Economy. https://doi.org/10.2139/ssrn.3562570

[54] Bernadett Palyi, Nora Magyar, Judit Henczko, Balint Szalai, Agnes Farkas, Thomas Strecker, Maria Takacs, and Zoltan Kis. 2018. Determining the effect of different environmental conditions on Ebola virus viability in clinically relevant specimens. *Emerging microbes & infections* 7 (03 2018), 52. https://doi.org/10.1038/s41426-018-0043-z

[55] Sweta Pathak, Alakes Maiti, and G. Samanta. 2010. Rich dynamics of an SIR epidemic model. *Nonlinear Analysis: Modelling and Control* 15 (March 2010), 71–81. https://doi.org/10.15388/NA.2010.15.1.14365

[56] Sascha Perkowski. 2020. about-corona. https://about-corona.net/. [Online; accessed 8-June-2020].

[57] Suani Pinho, Claudia Ferreira, Lourdes Esteva, Florisneide Barreto, V Silva, and M.G.L. Teixeira. 2010. Modelling the dynamics of dengue real epidemics. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 368 (December 2010), 5679–93. https://doi.org/10.1098/rsta.2010.0278

[58] Kiesha Prem, Alex R. Cook, and Mark Jit. 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology* 13, 9 (September 2017), 1–21. https://doi.org/10.1371/journal.pcbi.1005697

[59] William J. M. Probert, Chris P. Jewell, Marleen Werkman, Christopher J. Fonnesbeck, Yoshitaka Goto, Michael C. Runge, Satoshi Sekiguchi, Katriona Shea, Matt J. Keeling, Matthew J. Ferrari, and Michael J. Tildesley. 2018. Real-time decision-making during emergency disease outbreaks. *PLOS Computational Biology* 14, 7 (July 2018), 1–18. https://doi.org/10.1371/journal.pcbi.1006202

[60] Ritchie-Dunham, James L., Méndez Galván, and Jorge F. 1999. Evaluating epidemic intervention policies with systems thinking: a case study of dengue fever in Mexico. *Syst. Dyn. Rev.* 15 (1999), 119–138. https://doi.org/10.1002/(SICI)1099-1727(199922)15:2<119::AID-SDR163>3.0.CO;2-G

[61] Ahmad Sedaghat, Ali Mostafaeipour, and Seyed Oloomi. 2020. Prediction of COVID-19 Dynamics in Kuwait using SIRD Model. *Integrative Journal of Medical Sciences* 7 (June 2020). https://doi.org/10.15342/ijms.7.170

[62] Samithamby Senthilnathan. 2019. Usefulness of Correlation Analysis. https://doi.org/10.2139/ssrn.3416918

[63] Farhana Shahid, Shahinul Hoque Ony, Takrim Rahman Albi, Sriram Chellappan, Aditya Vashistha, and A. B. M. Alim Al Islam. 2020. Learning from Tweets: Opportunities and Challenges to Inform Policy Making During Dengue Epidemic. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 065 (May 2020), 27 pages. https://doi.org/10.1145/3392875

[64] Esther Shaulova. 2020. Global No.1 Business Data Platform Insights and facts across 600 industries and 50+ countries. https://www.statista.com/. [last accessed August 11, 2020].

[65] Rajesh Singh and Ronojoy Adhikari. 2020. Age-structured impact of social distancing on the COVID-19 epidemic in India (Updates at https://github.com/rajeshrinet/pyross).

[66] Ram Singh, Meenu Rani, Akshaya Bhagavathula, Ranjit Sah, Alfonso Rodriguez-Morales, Himangshu Kalita, Chintan Nanda, Shashi Sharma, Yagya Sharma, Ali Rabaan, and Jamal Rahmani. 2020. The Prediction of COVID-19 Pandemic for top-15 Affected Countries using advance ARIMA model. *JMIR public health and surveillance* 6 (May 2020), e19115. https://doi.org/10.2196/19115

[67] Padet Siriyasatien, Sudarat Chadsuthi, Katechan Phd, and Kraisak Kesorn. 2018. Dengue Epidemics Prediction: A Survey of the State-of-the-Art based on Data Science Processes. *IEEE Access* 6 (September 2018), 1–39. https://doi.org/10.1109/ACCESS.2018.2871241

[68] Stephen Smith, Patrick Sullivan, Beth Gardner, Juan Aparicio, Almira Hoogesteijn, and Carlos Castillo-Chávez. 2003. Identification of Geographical Factors Associated with Early Spread of Foot-and-Mouth Disease in the 2001 Uruguayan Epidemic. *American journal of veterinary research* 64 (January 2003), 1–31. https://doi.org/10.2460/ajvr.2003.64.1519

[69] R. Draxler Tianfeng ChaiR. 2014. Root mean square error (RMSE) or mean absolute error (MAE)? *Geoscientific Model Development Discussions* 7(1) (January 2014), 1525–1534. https://doi.org/10.5194/gmdd-7-1525-2014

[70] Ramadhan Tosepu, Joko Gunawan, Devi Effendy, La Ahmad, Hariati Lestari, Hartati Bahar, and Pitrah Asfian. 2020. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Science of The Total Environment* 725 (April 2020), 138436. https://doi.org/10.1016/j.scitotenv.2020.138436

[71] Sherry Towers and Zhilan Feng. 2012. Social contact patterns and control strategies for influenza in the elderly. *Mathematical biosciences* 240 (August 2012), 241–9. https://doi.org/10.1016/j.mbs.2012.07.007

[72] Johns Hopkins University. 2020. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://coronavirus.jhu.edu/map.html. [last accessed August 13, 2020].

[73] Robert Verity, Lucy Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick Walker, Han Fu, Amy Dighe, Jamie Griffin, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Anne Cori, Zulma M. Cucunubá, Rich FitzJohn, Katy Gaythorpe, and Neil Ferguson. 2020. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases* 20 (March 2020), 630–631. https://doi.org/10.1016/S1473-3099(20)30243-7

[74] Huwen Wang, Zezhou Wang, Yinqiao Dong, Ruijie Chang, Xiaoyue Yu Chen Xu, Shuxian Zhang, Lhakpa Tsamlag, Meili Shang, Jinyan Huang, Ying Wang, Gang Xu, Tian Shen, and Xinxin Zhang & Yong Cai. 2020. Phase-adjusted estimation of the number of Coronavirus Disease 2019 cases in Wuhan, China. *Cell Discovery* 6 (February 2020), 1–8. https://doi.org/10.1038/s41421-020-0148-0

[75] Christopher Weible, Daniel Nohrstedt, Paul Cairney, David Carter, Deserai Crow, Anna Durnova, Tanya Heikkila, Karin Ingold, Allan Mcconnell, and Diane Stone. 2020. COVID-19 and the policy sciences: initial reactions and perspectives. *Policy Sciences* 53 (April 2020), 225–241. https://doi.org/10.1007/s11077-020-09381-4

[76] J. Weng, Y. Xu, and A. R. Sharma. 2012. Epidemic analysis and visualization based on Digital Earth spatio-temporal framework. In *2012 IEEE International Geoscience and Remote Sensing Symposium.* IEEE, Munich, Germany, 7220–7223.

[77] WHO. 2009. Maps of the spread of influenza A(H1N1). https://www.who.int/csr/disease/swineflu/h1n1_maps_may/en/. [last accessed 2009].

[78] WHO. 2016. Global Health Observatory data repository. https://apps.who.int/gho/data/node.main.CBDR107?lang=en. [last accessed August 13, 2020].

[79] WHO. 2016. Global Health Observatory visualizations. https://apps.who.int/gho/data/node.wrapper.imr?x-id=97. [last accessed August 13, 2020].

[80] WHO. 2020. Ebola virus disease. https://www.who.int/health-topics/ebola/#tab=tab_1. [last accessed May 28, 2020].

[81] Worldometers.info. (n.d.). COVID-19 CORONAVIRUS PANDEMIC. https://www.worldometers.info/coronavirus/. [last accessed August 13, 2020].

[82] Donald S. Likosky Xiaoting Wu. 2020. p-Values and Their Unintended Consequences. *J Extra Corpor Technol* 52(2) (June 2020), 88–89. https://doi.org/10.1182/ject-522editorial

[83] Hongxin Xue, Yanping Bai, Hongping Hu, and Haijian Liang. 2019. Regional level influenza study based on Twitter and machine learning method. *PLOS ONE* 14, 4 (April 2019), 1–23. https://doi.org/10.1371/journal.pone.0215600

[84] Chao-Tung Yang, Yuan-An Chen, Yu-Wei Chan, Chia-Lin Lee, Yu-Tse Tsan, Wei-Cheng Chan, and po-yu Liu. 2020. Influenza-like illness prediction using a long short-term memory deep learning model with multiple open data sources. *The Journal of Supercomputing* 1(1) (February 2020), 1–27. https://doi.org/10.1007/s11227-020-03182-5

[85] Selen Yilmaz Isikhan and Dilek GÜLEÇ. 2018. The Clustering of World Countries Regarding Causes of Death and Health Risk Factors. *Iranian Journal of Public Health* 47 (October 2018), 1520–1528.

[86] Xue Ying. 2019. An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series* 1168 (February 2019), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

[87] Choujun Zhan, Chi Tse, Yuxia Fu, Zhikang Lai, and Haijun Zhang. 2020. Modeling and Prediction of the 2019 Coronavirus Disease Spreading in China Incorporating Human Migration Data. https://doi.org/10.1101/2020.02.18.20024570

[88] Shi Zhao, Salihu Musa, Qianying Lin, Jinjun Ran, Guangpu Yang, Weiming Wang, Yijun Lou, Lin Yang, Daozhou Gao, Daihai He, and Maggie Wang. 2020. Estimating the Unreported Number of Novel Coronavirus (2019-nCoV) Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak. *Journal of Clinical Medicine* 9 (February 2020), 388. https://doi.org/10.3390/jcm9020388

[89] Zhongming Zhao, Xiaozhuang Wu, Yixi Zhong, Keqin Zhang, Ya-Ping Zhang, Eric Boerwinkle, and Yun-Xin Fu. 2004. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC evolutionary biology* 4 (July 2004), 21. https://doi.org/10.1186/1471-2148-4-21