



4P Model for Dynamic Prediction of COVID-19: a Statistical and Machine Learning Approach

Khandaker Tabin Hasan¹ · M. Mostafizur Rahman¹ · Md. Mortuza Ahmmed¹ · Anjir Ahmed Chowdhury¹ · Mohammad Khairul Islam¹

Received: 20 July 2020 / Accepted: 21 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Around the world, scientists are racing hard to understand how the COVID-19 epidemic is spreading and growing, thus trying to find ways to prevent it before medications are available. Many different models have been proposed so far correlating different factors. Some of them are too localized to indicate a general trend of the pandemic while some others have established transient correlations only. Hence, in this study, taking Bangladesh as a case, a 4P model has been proposed based on four probabilities (4P) which have been found to be true for all affected countries. Efficiency scores have been estimated from survey analysis not only for governing authorities on managing the situation ($P(G)$) but also for the compliance of the citizens ($P(P)$). Since immune responses to a specific pathogen can vary from person to person, the probability of a person getting infected ($P(I)$) after being exposed has also been estimated. And the vital one is the probability of test positivity ($P(T)$) which is a strong indicator of how effectively the infected people are diagnosed and isolated from the rest of the group that affects the rate of growth. All the four parameters have been fitted in a non-linear exponential model that partly updates itself periodically with everyday facts. Along with the model, all the four probabilistic parameters are engaged to train a recurrent neural network using long short-term memory neural network and the followed trial confirmed a ruling functionality of the 4Ps.

Keywords 4P model · COVID-19 · Kinetics-modeling-fitting · Non-linear exponential model · LSTM

Introduction

From the beginning of civilizations, the human race has witnessed epidemics of various forms and degrees. The recorded history of Greek epidemic dated back to 430–427 BCE which claimed the lives of an estimated 25–35 Greeks according to the contemporary historian Thucydides. Before the outbreak of COVID-19, there were many occurrences of epidemics, most notably The Black Death (1346–1353) that wiped out nearly half the population of Europe and Spanish Influenza (1918–1919) caused about 500 million deaths around the world. A detailed account of the history of

epidemic and pandemic can be found in [1]. Throughout history, it has been observed that inadequate knowledge about the disease, misinformation and misconception among the populous, and improper handling of the situation caused more damage than it could have [2].

No wonder, in the case of COVID-19, we have experienced various confusing or sometimes conflicting information about the epidemic that affects the proper functioning of government. As succeeding populations are getting infected through social contact, the countries who have taken early measures have turned out to be the successful ones in containing the outbreak. Before any scientific correlations could be found among factors of contagion, basic measures could have saved most of the lives that were lost. Moreover, government's inaction and citizen's ignorance have turned the current situation at stake. While the nature of the virus and its contagiousness are yet to be determined, still proper actions by the legislative body or government and compliance of citizens could significantly reduce the spread and have better containment. A study on social distancing policies driven by public awareness was conducted and

This article belongs to the Topical Collection *Data-Driven Artificial Intelligence approaches to Combat COVID-19*
Guest Editors: Mufti Mahmud, M. Shamim Kaiser, Nilanjan Dey, Newton Howard, Aziz Sheikh

✉ Khandaker Tabin Hasan
tabin@aiub.edu

Extended author information available on the last page of the article.

voluntary actions were found to have the strongest causal impact on reducing social interactions which resulted in the decline of the rate of infection by 37% after 15 days [3]. Mahmud et al. [4] considered their model on social dynamics that excludes government roles in control. The number of tests per population size is very important for allowing authorities to isolate and treat the infected individual avoiding further spread. In this paper, we cast down our eyes on an epidemic prediction model for COVID-19 that considers the role of governance and citizen's consciousness along with other necessary variables [5]. This model appraises the concerned audience how their actions and behaviors change the outcome of the onrush by getting user inputs of the aggregated parameters on an interactive Web user console available at [Covid-19 Bangladesh Projection](#). Therefore, this work is not just a model that works on a computer and reports the result in an ambiguous way that the general public do not understand, it is also a functional system that communicates information with the concerned parties.

The hard part of the model is to initialize the parameter values accurately enough to fit the reported data that can reasonably project the future. Though accuracy at an early phase is not what we are concerned with, we have focused on its correct behavior based on social and pathological dynamics. The model updates its parameter on a regular basis when new data are received. Testing of hypothesis and methodology have been described in the "Methodology" section followed by model inception and implementation.

Social Dynamics

Scherer and Cho (2003) showed how one's social networking behavior, as well as perception, impacts the health behavior of the respondent [5]. The government of several countries remain unsuccessful in assuring their people that they have control over the situation. In the interim, both the number of newly infected cases and deaths continue to increase every day. Pessimistic decisions may ascend either when a hazard is publicized aggressively or when a dreadful circumstance is presented with extravagant believability [6]. The decelerate response of several countries to the COVID-19 sternness can be explained through geographic collocation. Findings of a research work led by Fischhoff et al. (2003) revealed that the hazard discernment concerning a hostile affair is induced by propinquity to the hazard [7]. In another study, it was observed that the perceptions and behaviors toward the hazard concerning avian influenza (H5N1) virus striking Europe between 2005 and 2006 were significantly associated with the proximity to hazards [8].

Besides, the inability of the political leaders to have people in their confidence to combat COVID-19 would have

an influential impact on people to be defiant to the government's order regarding social distancing and lockdown [9]. The government should clarify the motives of a lockdown to its people, particularly to the younger ones who have been found to breach the lockdown rules constantly [10]. Confidence in mainstream media has been found to be correlated with support for lockdown measures as well [11]. In a country like Bangladesh having misgovernance, the government has traditionally been failing to build proper awareness among the people during any emergency period and manage the crisis politically [12].

Prediction Models and Their Components

The SIR model [13] is one of the olden compartmental models in epidemiology projecting infectious diseases like COVID-19 [14] and numerous diversified derivations came out from it. The principal SIR model comprises three compartments.

S: The number of susceptible individuals. It is the number of a population who is at risk of being infected and after being infected. The susceptible individuals are then shifted to the next infectious compartment *I* after being infected. The transmission rate from susceptible to infected is assumed to be $\beta SI/N^2$ where *N* is the total population and the transmission parameter β is the average number of individuals that one infected individual will infect per unit time.

I: The number of infectious individuals. The population of this compartment are the individuals who have been infected and are capable of infecting susceptible individuals. A segment of the population of this infected class will be shifted to the next removal compartment at a recovery rate γ , so that $1/\gamma$ is the average period during which an infected individual remains infectious.

R: The number of removed individuals. The individuals who have been infected and have either recovered from the disease or died entered the removed compartment.

The basic reproductive number R_0 , which is inured to quantify the transmission of pathogens, is the ratio β/γ . That means R_0 is the average number of people infected by an infected individual over the disease infectivity period, in a susceptible population [15]. The model assumes that each transferring person has an equal probability of being contracted by others regardless of social practice and rule in place. The model considers the infected people residing in quarantine as similar as those who are not in quarantine. Hence, both are assumed to have equivalent transmission rates β which might not be true in real cases. Finally, the assumption that β stays constant all the way through the duration of pandemic also might not be veritable. The SIR model is weak to translate the gestation period when an

affected person is moving ubiquitously without showing any symptoms whatsoever.

SEIR model is the most widely used epidemic model derived from the SIR model. In addition to SIR, the SEIR model introduced an intermediate compartment E for exposed population for the incubation period during which susceptible individuals have been infected but are not yet infectious themselves, and σ is the incubation rate at which the dormant individuals become infectious. The SEIRS model is used to allow recovered individuals to return to a susceptible state at a rate ξ at which they return to the susceptible status due to the loss of immunity [16].

Whether a person is susceptible to a specific pathogen partly depends on the person's immune response. The person could have natural active immunity or might have developed antibodies through a non-deliberate contact with the pathogen. Natural passive immunity that comes from the mother is also possible given the length of the COVID-19 epidemic [18]. Furthermore, overcoming the population's natural immunity, the type of strain is another numerator for the susceptible. Body mass index (BMI) is said to have an influence on developing or aggravating the disease, but the studies are complicated by the nutrition factors that should oppose the effect of BMI, and we chose not to consider this in our model [19].

After getting infected (I in SIR model), an uninfected person contributes to the transmission rate in a population N, where N should be a population in proximity and have means to come in close contact, either active or passive and not just a mass in a location. By passive contact, we mean places and artifacts shared by multiple persons at different times within an effective interval: this is how long the SARS-CoV-2 can survive given surface and aerosol stability [20]. R, the removal by recovery or death, appears to have little significance since the necessary transmission is already made within the I compartment, and N has to be observed in a real social context. In our computation, we have considered the median incubation period 5.2 days when the infected I infects others [21]. In the SEIR model, instead of susceptible (N), exposed individual (E) bears greater significance in calculating the rate of infection. Analyzing all conceivable notions and dynamics, the basic reproductive quantity has been fragmented into several practical factors described in kinetics-modeling-fitting.

A popular way of risk assessment was proposed by statistician Fergus Simpson who attempted to estimate the risk of being captured by the virus in terms of the number of dice [22]. According to him, in the first week when a virus initiates to spread across a country, one person out of ten million catches the virus each day, equivalent to a nine-dice risk for being infected by the virus and which is very low. But, if the virus can spread unrestrictedly, the accompanying risk will increase fast. By week 2, the risk

increases approximately six times compared to week 1, equivalent to an eight-dice risk. In the third week, it parallels seven-dice risk, equals six-dice risk by week 4, and so on. But he also cautioned that his formula might not work where the checkout process of positive cases being performed each day is very low with respect to the population size. Hence, his approach has not been taken into consideration in this study. Although it was anticipated by several researchers initially that the density of the population had a positive correlation with the spread of COVID-19 [23], some recent studies show that the assumption might not be completely evident [17, 24]. Health expenditure by the government was identified as a significant determinant of deaths caused from COVID-19 [25, 26], whereas topographical locations along with their climatic circumstances were detected as prominent factors for the diffusion of the virus [27]. Although the temperature has been observed to be a relevant factor of COVID-19 transmission according to the results of some studies [28], there are some counter-arguments as well [29, 30]. Males, as well as older people, were observed to be exceedingly vulnerable to COVID-19 [31, 32]. Even though a few studies have been conducted on the overall COVID-19 situation along with its various aspects regarding Bangladesh so far [33], none of them has unambiguously focused on the projection of it. In their study, Azad and Hussain (2020) have attempted to fit the available data to date in Bangladesh applying several conventional models like the exponential model, Richards model, logistic model, compartmental model, and Gompertz model, but these traditional models do not take crucial parameters associated with the actions and behaviors [34].

LTSM Network

LSTM: Among the state-of-the-art deep learning methods, recurrent neural network (RNN) has been proven to be the most robust for prediction as it can automatically excerpt the necessary features from the training samples, delivering the activation from the previous time step as the load for the present time step and the network self-connections. Long short-term memory (LSTM) is one of the most powerful and well-known RNN which is skilled to learn order dependence in sequence prediction state. We have tried out many other models along with the LSTM model, e.g., adaptive neuro fuzzy inference system (ANFIS) and multilayer perceptron (MLP) models provide some results but not as good as LSTM models. From Table 1, we can clearly understand that the ANFIS model is not so accurate compared to the LSTM model.

On the other hand, Table 2 is for the result of the MLP model.

Although the error rate is lower in MLP model, the resulting correlation coefficients indicate negative association

Table 1 MLP training results

1st, 2nd, 3rd, 4th units	Batch size	RMSE	MAPE	Corr Coff.
[5,10,15,20]	8	515.1363024	12.8653879	-0.12
[15,20,25,30]	8	499.73015	12.3597964	-0.28
[15,20,25,30]	8	658.3847978	15.5378021	-0.17

among the parameters. We have also checked with the STGCN model that works better in link prediction on adjacent matrix, but it has not fitted our dataset. Due to inadequate accuracy of both the MLP and ANFIS models, we have finally decided to apply the LSTM model for making prediction based on our dataset as shown in Table 5.

Methodology

In conventional compartmental models in epidemiology like SIR, SEIR initiates with the size and density of the population [35]. But after the outbreak of COVID-19, questions are being raised by the epidemiologists and the public health experts whether the population size as well as its density along with urbanization matter or not. The population size of the USA is 5.5 times more than Italy. If the population size underwrites the infected numbers, then nearly 5.5 times more infected cases would be expected in the USA in comparison to Italy. Likewise, Sweden has experienced almost a similar death rate as Ireland despite of having lesser population density. Similarly, Spain has undergone about an identical mortality rate as Italy although the population density of Italy is more than twice that of Spain. Moreover, urban cities like Shanghai, Seoul, and Singapore with enormous population densities have shown better performance in combatting the COVID-19 situation than many other cities which have comparatively a low population density. In this study, three different measures of correlation coefficients, namely, Pearson's product-moment correlation coefficient, Kendall's correlation coefficient, and Spearman's rank correlation coefficient, are used to evaluate the strength of the bivariate relationship between the confirmed COVID-19 positive cases as well as deaths, population density along with size, and urbanization. Results

Table 2 ANFIS training results

MF type	No. of MF	RMSE	MAPE	Corr Coff.
Triangular	3	10,365.8	93.28	0.52
Trapezoidal	3	8300.61	58.076	0.48
Gaussian	3	8135.48	70.97	0.42

are displayed graphically in Fig. 1, indicating weak correlations between the variables pairwise.

To discard the aforementioned infeasibility of population size and density as well as difficulties in estimating the transmission rate kinetics in SIR and SEIR models, a probabilistic approach has been adopted in this study where government control, people's acquiescence to the norms and rules of COVID-19, test positivity, and infection transmission frequency have been encompassed. Updated values from statistical survey and analysis (until 03 Aug 2020) of all four parameters can be found on our Web portal¹ with downloadable daily data sheet at *Dashboard* bar. The secondary data used in this study have been extracted from the WHO (url: <http://www.who.int>) and the Institute of Epidemiology, Disease Control and Research (IEDCR) (url: <https://iedcr.gov.bd>), while the primary data regarding people's compliance along with government's control have been collected from the respondents through a sample survey. A separate survey has been utilized to collect the information about the number of people expected to come in contact with an infected person if s/he moves outside home for whatever reason. Finally, a modified exponential regression model has been developed to fit the observed data encompassing the probability of citizen's compliance accompanied by the government's control, test positivity, and infection transmission frequency and to prove the dynamicity of all the probabilistic parameters. LSTM is used to train a neural network making a comparison with the cumulative positive COVID-19 cases of Bangladesh.

Kinetics-Modeling-Fitting

Identification of the relevant causes accompanied by interventions is the must to limit the spread of an epidemic. Epidemiological models work as a guide to plan regarding the outbreak of an epidemic. In the case of extremely infectious COVID-19, most of the affected nations have planned social distancing along with self-isolation measures after vigilant observation of the kinetics of the virus's evolution. With the multifaceted socio- and pathological dynamics of COVID-19, it is hard to predict when and how it grows and ends [36]. To date, different groups of scientists have come up with many different models with wide deviations [37]. Selecting too many parameters would increase the risk of non-sampling error due to imprecise and inadequate data, e.g., to estimate the number of people commuting en masse at a given time and location with acceptable accuracy in a country like Bangladesh could be very difficult. As mentioned earlier, this issue has been encountered through a sample survey covering the people aged between 21 and 50 which is the highest COVID-19

¹Link: <https://covid19.aiub.edu/>

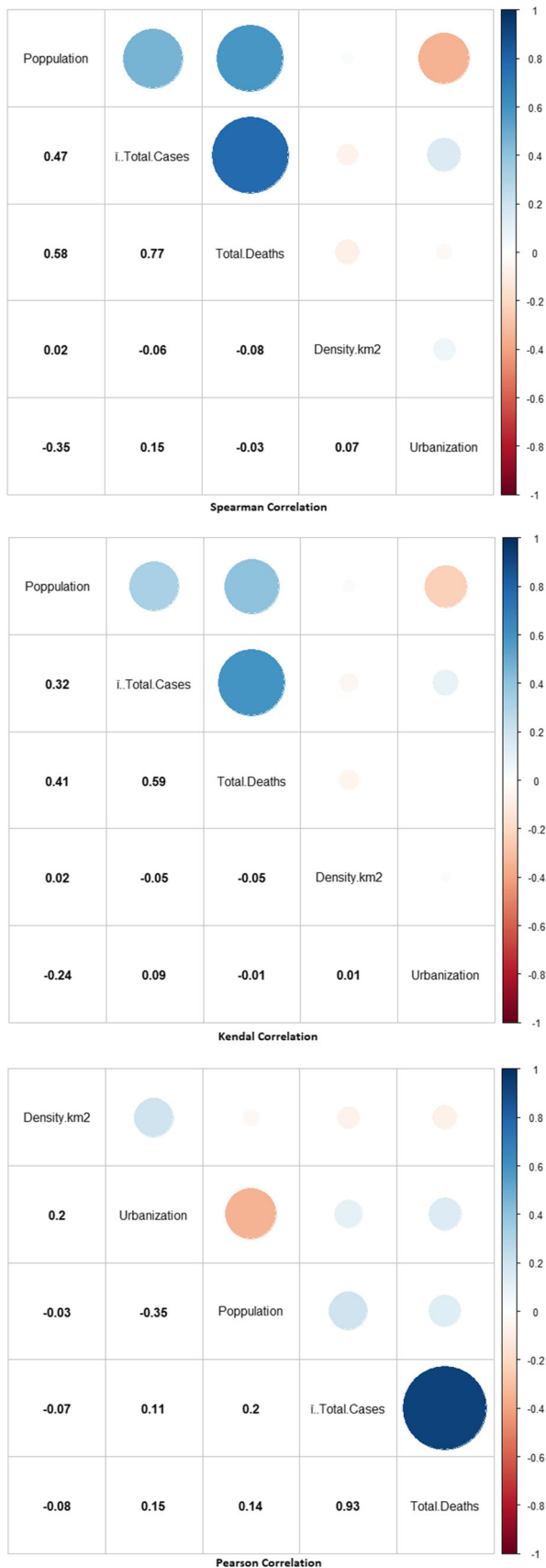


Fig. 1 Correlation coefficients between the selected variables

affected age group in Bangladesh. The survey also covers the working people of both sexes who have to go out for their job. The principal focus of this study has been on pandemic predictability by factors that influence the outcome of an exponential model by fitting four probabilities (4P) which can be adjusted over time when a new observation comes in. State’s control and citizen’s compliance:

$P(G)$ = Probability of the state’s control over the situation.

This encompasses:

1. State making the necessary laws to contain the pandemic.
2. State enforces the law in strict order.
3. State’s healthcare infrastructure and improvement where necessary.
4. Public awareness program for motivation.

$P(P)$ = Probability of the citizen’s compliance to the imposed laws and health regulations.

This includes:

1. Complying to laws regulating social distancing, such as the use of public transport, social gathering, and shopping.
2. The practice of sanitization.
3. Wearing protective gears when necessary.
4. Perception of the situation.

$P(G)$ has direct control over $P(P)$ whereas $P(P)$ has passive control over $P(G)$, which indicates an asymmetric relationship between them. However, poor scores of both $P(G)$ and $P(P)$ would only degrade the situation. Hence, both must function together in order to have an impact on the overall situation. Therefore, a joint probability of $P(G)$ and $P(P)$, namely $P(C)$, has been designed for the study which is inversely proportional to the propagation rate of a pandemic.

$$P(C) = P(G)P(P)$$

Where $P(C)$ is the joint factor of state and its citizens. To calculate the probability of overall control $P(C)$, data collected through sample survey have been utilized as mentioned before. The survey provides data about the public compliance (P), and government control (G) being scaled between 1 and 10 (1 indicates the lowest degree of acquiescence, whereas 10 indicates the highest one) along with information whether the respective respondent is affected with COVID-19 or not. Exactly 1119 respondents have participated in the survey, based on which, $P(C)$ has been estimated as follows:

$$P(C)=P(G)P(P)$$

Where $P(C)$ is the joint factor of state and its citizens.

$P(P)$ = weighted probability of public compliance

$$= \frac{\sum_{i=1}^{10} d_i p_i}{10}; d_i = 1, 2, 3, \dots, 10$$

Where, p_i = unweighted probability of public compliance:

$$= \frac{\text{number of respondents on compliance having specific scale value}}{\text{total number of respondents}}$$

Also,

$P(G)$ = weighted probability of government control

$$= \frac{\sum_{i=1}^{10} d_i p_i}{10}; d_i = 1, 2, 3, \dots, 10$$

Where, p_i = unweighted probability of government control:

$$= \frac{\text{number of respondents on control having specific scale value}}{\text{total number of respondents}}$$

Applying the aforementioned formulas, we have estimated $P(P) = 0.324664879$ and $P(G) = 0.295084897$ from the survey data of this study. Finally, we have estimated

$$P(C) = P(P)P(G) = 0.095803703.$$

Pathogen's Reproduction Number and Contagiousness

The reproduction number, R_0 , is a complicated factor to compute that indicates how contagious an infectious disease is [14, 38]. It states how many individuals can get contracted by a single infected person with a previously unknown disease. There are three known variants of SARS-CoV-2 (A, B, C) and it is changing the nucleotide sequence of the genome, and we do not know which variant is prevailing in the South Asian region [39]. Nine new mutations have been reported in Bangladesh according to an unpublished report [40]. Also, it is not determined yet why some people who are more affected are showing serious symptoms, while others remain immune or asymptomatic. It also remains undiscovered why and how this has formed serious havoc in some parts of the world whereas the continent like Africa is only mildly affected. So, it is appallingly difficult to determine the contagiousness here in Bangladesh. Since it depends on the type of pathogen and how people are interacting with each other in certain social settings, we have considered this part in $P(C)$. Many scientists have hypothesized the possibility of herd immunity [41] or active natural immunity (ANI), which is very hard to quantify as a factor of R_0 . By far, since the virus is new to our understanding and no vaccine (natural artificial immunity) has yet been developed, we have no idea how an individual's immune would respond at first exposure [42]. On the other hand, the probability of a person catching infection once being exposed, designated as $P(I)$, is an essential factor to understand the velocity and magnitude of the contagion.

$P(I)$ = Probability of a person getting infected after being exposed.

Where, $P(I)$ is primarily a function of:

1. Strength of pathogen type, S_V : Value not known
2. Human natural immunity, I_N : Value not known
3. Viral load, L_V : Value not known

V_L is the threshold of viral load to which a target is exposed to be infected. Viral load is a measure of virus particles, also called infectious dose, that is the amount of viruses needed to establish an infection. For influenza viruses, people need to be exposed to as little as 10 virus particles to get infected, while as many as thousands for other human viruses. W. David Hardy mentioned that "The virus is spread through very casual interpersonal contact [43]." We did not find an estimation of how many virus particles of SARS-CoV-2 are needed to trigger the infection, but COVID-19 is clearly very contagious, probably because few particles are needed for causing the infection leading to a low infectious dose or viral load [44]. Since the values of the above components are yet to be reliably established, we have passively determined the $P(I)$ through statistical studies on various data, which is proportional to the growth of a pandemic and varies from region to region.

Thus, we can say:

$P(I)$ is the configuration of the strength of pathogen type, human natural immunity, and viral load and we have tried to determine $P(I)$ by the rate of propagation in a sample population who are possibly being exposed and part of them got infected.

The probability of infection $P(I)$ has been estimated through:

$$P(I) = \frac{f_{i+IP}}{N\varphi_i}; i = 1, 2, 3, \dots$$

Where, N = number of people expected to come in contact to an infected person:

$$\varphi_i = \sum_{i^{\text{th}} \text{ day}}^{(i+IP-1)^{\text{th}} \text{ day}} \text{+ve cases}$$

$$f_{i+IP} = \text{+ve cases of } (i + IP)^{\text{th}} \text{ day}$$

In this study, the value of N is estimated to be 10 from a different survey conducted on 359 respondents who have gone outside the home for some/any reason with the risk of being infected by an infected person and IP is the incubation period.

Test Positivity

Test positivity is the ratio between the number of daily positive cases and tests done. A high percentage of the population being tested positive may assume that the right people are being tested and more tests would accumulate more positive results. Though the association between the number of tests and positive cases is not linear, it helps to contain the disease at an early stage [45]. This proportion

Table 3 List of countries with high to low test positivity

Country	Test positivity (%)	Total cases	Tests/million Pop
Brazil	49.67	376,669	3461
Bangladesh	14.00	35,585	1538
USA	11.43	1,706,226	44,587
India	4.60	144,950	2200
S. Korea	1.35	11,206	16,121
Australia	0.57	7118	48,885

Data accessed: May 25, 2020

will be headed down from either upsurge in the number of tests increases or drop in contagiousness. Two types of testing approach have been observed around the world depending on the state's capacity and policy: (1) reactive tests where only people with acute symptoms are being tested and (2) proactive tests where subjects are being tested at random. Countries that have gone through reactive testing obtained high test positivity by testing only people with acute symptoms while less symptomatic or asymptomatic cases are free to roam contracting others upon contact. Conversely, countries with proactive testing, where people are being tested in every suspected scenario at random got a lower test positivity. The results in Tables 3 and 4 show how the containment of the pandemic is proportional to the test positivity. Table 4 is exhibiting an increase over total cases for 29 days. Table entries in bold indicate high Test Positivity resulting in higher number of cases. Besides, it also shows that the accomplishing situations in some countries have strong correlations with the percentage of test positivity. The daily trend can be found on various authentic data sources like in [46, 47].

Here are the following tables.²

The World Health Organization (WHO) reported incubation period (time from exposure to the development of symptoms) for COVID-19 between 2 and 10 days [46]. The mean incubation period was 5.2 days (95% confidence interval) [21]. In our model, test positivity $P(T)$ acts as the probability of getting positive cases out of the total number of tested cases and it is regularly updated with the moving average test positivity of incubation period (IP).

$P(T)$ = Moving average of the probability of getting positive cases out of total samples being tested (Fig. 2).

Without running into a complex set of incomprehensible parameters, we have tried to keep the model simple with a core equation apprehended by 4Ps that can be learned over time reflecting a closer prediction of reality.

Finally, the regression model fitted with the data becomes:

$$I(t) = \alpha \left(1 - e^{-\mu \frac{P(I)P(T)}{P(g)P(p)} t^n} \right)$$

here α , μ , and n are constants.

²source: www.worldometers.info/coronavirus

Table 4 List of countries with high to low test positivity

Country	Test positivity	Total cases	Tests/million (Pop)	Increase over 29 days (%)
Brazil	44.86	1,151,478	11,436	206
Bangladesh	18.36	119,198	3,831	235
USA	8.25	2,416,991	86,834	42
India	6.33	456,115	5038	215
S. Korea	1.05	12,484	23,057	11
Australia	0.36	7492	82,614	5

Data accessed: June 23, 2020

LSTM

LSTM network is an enhanced version of RNN. It has memory blocks instead of neurons that are connected through layers. Each block has its own components which make it smarter than classical neurons. It also has memory to store the recent sequences. There are some gates in the block that actually control the block's current state and the output. Block uses the sigmoid activation function to check that the gates are triggered or not (Fig. 3).

This equation basically selects information which can be passed to the cell. The data from the input side of previous memory which is to be ignored is decided by the forget gate by the following equation:

$$f_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \quad (1)$$

The control gate controls the update of the cell by the following formula:

$$\tilde{C} = \tanh(W_c * [h_{t-1}, x_t] + b_c)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (2)$$

The output layer updates both hidden layer h_{t-1} and output and is given by:

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$

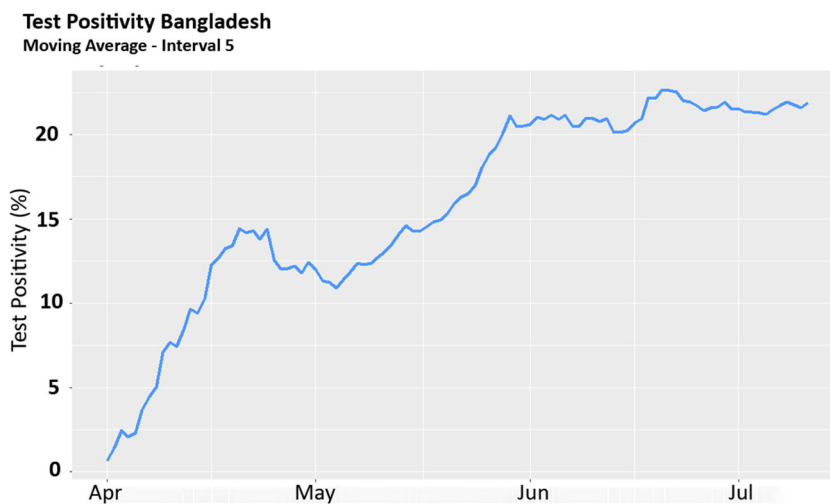
$$h_t = o_t * \tanh(C_t) \quad (3)$$

\tanh is used to normalize the values into range -1 to 1 . The weight matrices are W and the activation function is σ which is taken as the sigmoid.

Results and Discussions

Scientists are working on different facts of COVID-19 in Bangladesh as well as other countries of the world to analyze and forecast the propagation using the conventional SIR and SEIR models initializing some implicit compartmental rate constants [48, 49]. Hazardous civic interaction

Fig. 2 Test positivity, moving average from April 1 to July 12, 2020



plays the role of diffusing transmissible viral particles which is possible to control by being acquiescent with the health norms and law and order of the government. Test positivity is the vibrant contagion trend indicator which indicates how obviously the suspected are being infected. But these vital factors have been ignored by most pandemic models. In this proposed model, these key factors have not only been addressed but have also been integrated with their implication in fitting the propagation trail. Those who worked with traditional compartmental models considering the whole population as susceptible and density as a multiplier of the rate constants overlooked the correlation between the factors and the eruptions. Some studies showed the correlation between the degree of contagion and body mass index (BMI), but they ignored the nutrition facts. Evidently in this study, the population size and density are ignored due to the insignificant correlation aspect. The reproduction number and implicit compartmental rate constants which are certainly difficult to measure due to the insufficiency of reliable

data in Bangladesh are circumvented. We have explicitly addressed the impact of citizen’s awareness about the hazard of COVID-19, norms and rules of health concern, government laws, obligations, and citizen’s compliance with them. Moreover, we have also considered the test positivity and the probability of being infected.

Figure 4 shows how precisely the model is superimposing with real values. The growth rates of new positive cases, both estimated from model and real cases, are shown in Fig. 5. Though the growth rate is declining, its sluggish gradient indicates the lengthier presence of coronavirus here in Bangladesh.

The scenario may change upon how cognizant the citizens are about this contagious microbe. As mentioned earlier, the joint probability $P(C)$ of state’s control and citizen’s compliance regulates the propagation rate. The propagation has been estimated in $\pm 5\%$ controlled compliance situation. Figure 6 shows how only 5% decrease in $P(C)$ and 5% increase in $P(T)$ force the growth curve

Fig. 3 The structure of LSTM unit [51]

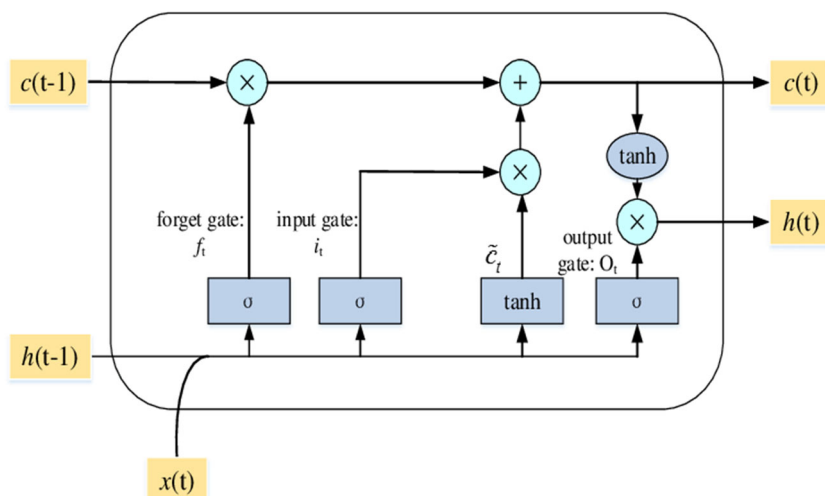


Fig. 4 Projection coherent with real positive cases (April 1 to June 29, 2020)

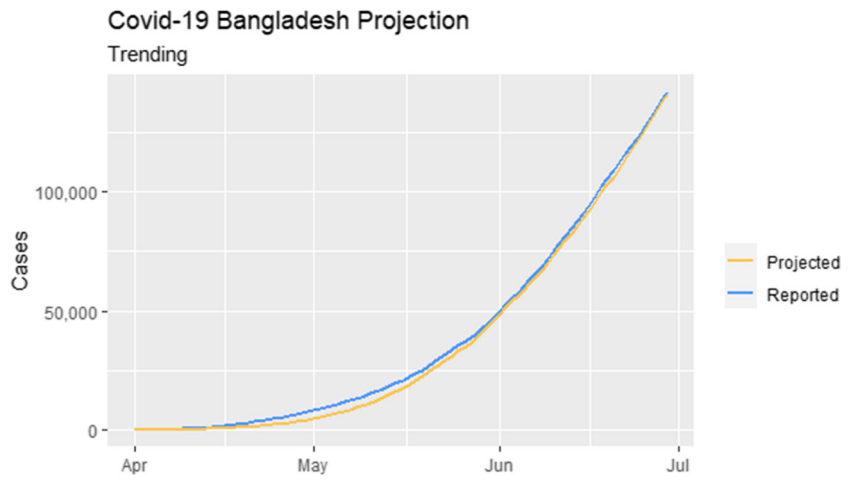


Fig. 5 Growth rate of positive cases

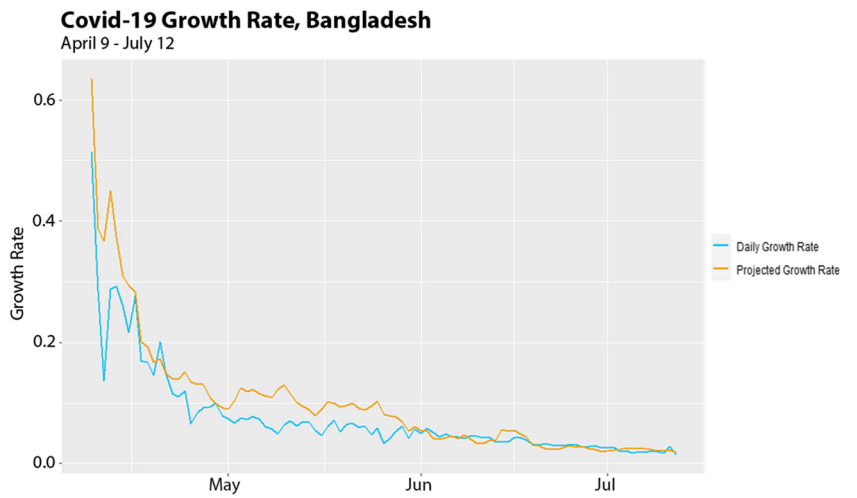


Fig. 6 5% decrease in P(C) and 5% increase in P(T) from the estimated values (April 1 to June 29, 2020)

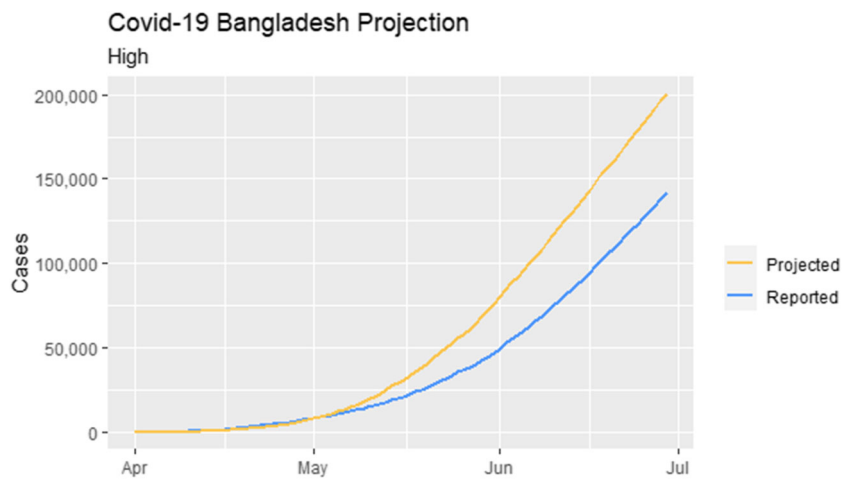


Fig. 7 5% increase in P(C) and 5% decrease in P(T) from the estimated values (April 1 to June 29, 2020)

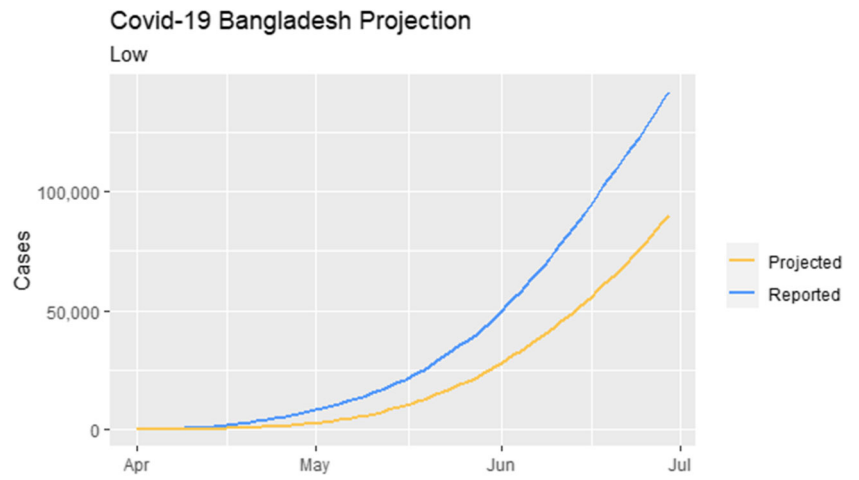


Fig. 8 Daily new positive cases, reported vs. projected

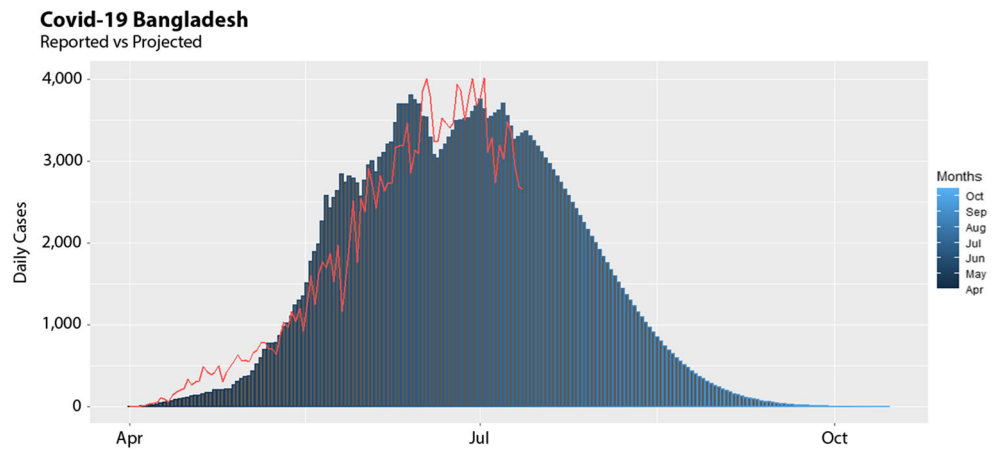


Fig. 9 Propagation projection with reported cases until July 12, 2020

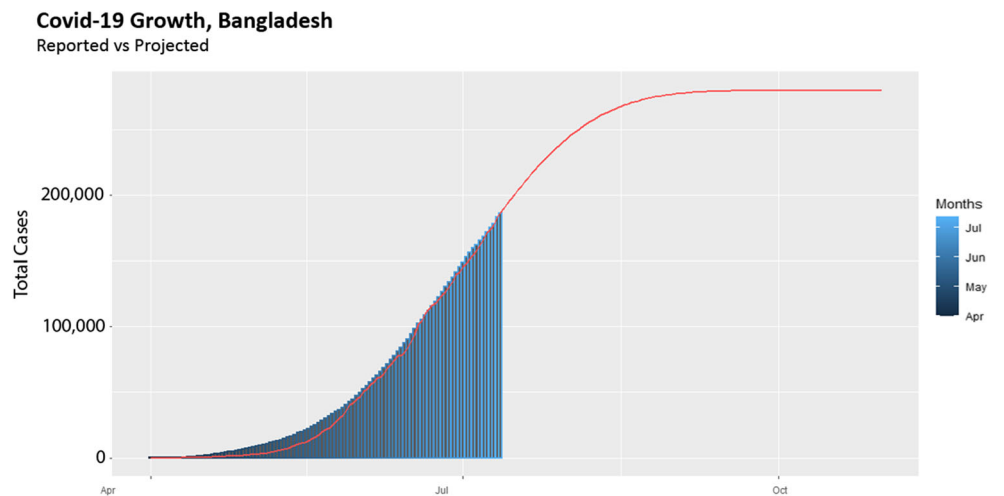


Table 5 LSTM training results

1st, 2nd, 3rd, 4th units	Batch size	RMSE	MAPE	Corr Coff.
[5,10,15,20]	32	10,368.318	5.96	1
[15,20,25,30]	32	10,518.03	6.11	1
[15,20,25,30]	8	8768.84	4.37	1

upward. Similarly, Fig. 7 shows exactly the opposite: 5% increase in $P(C)$ and 5% decrease in $P(T)$ from the estimated values. We have kept $P(I)$ unchanged in both cases.

The daily positive cases are estimated from the projected propagation by the first derivative and considered an average of the incubation period. Figure 8 shows the comprehensive scenario of daily positive cases with the estimated one. An estimated mean absolute prediction error (MAPE) of 0.20 indicates reasonable prediction by our 4P model.

Figure 9 shows the estimated flattening of the curve on last calculated $P(C)$, $P(T)$, and $P(I)$ which is subject to change. On June 26, the USA observed an all-time surge in new coronavirus cases which were about to get flattened during the first quarter of June following some premature reopening, public unrest, civil disobedience, etc., that evidently lowered the country's $P(C)$ score. Reports on critical events are available on the JHU Timeline of COVID-19 [50]. Although improving, the test positivity in the USA is still above 12% on average during this writing. In Bangladesh, a similar surge could be expected during the Islamic festival Eid-UI-Adha at the end of July. However, in every possible scenario, we continue conducting our study to get the values of the determinants and update the projection accordingly. Since there will no periodic version

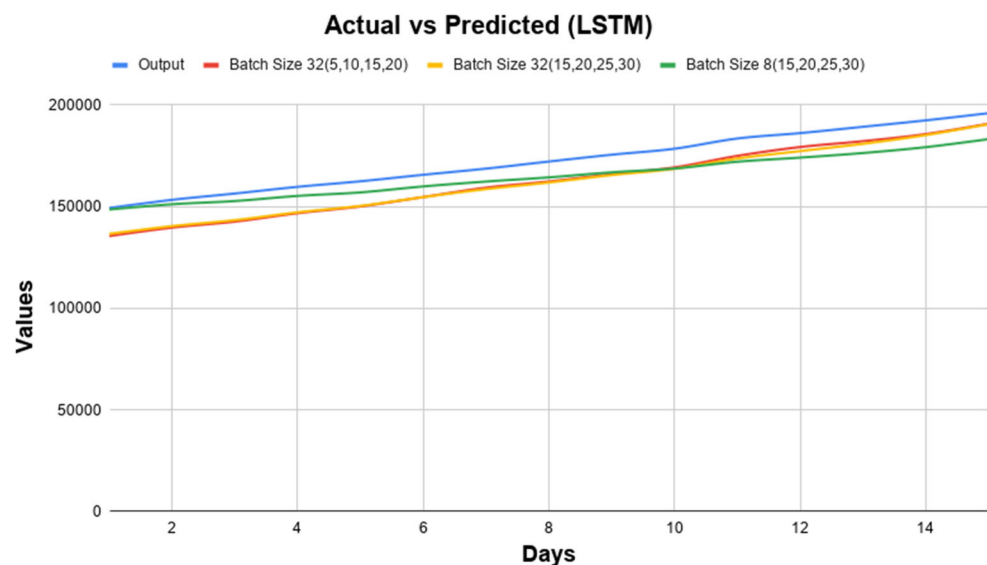
of the same paper, we have deployed our model available at [Covid-19 Bangladesh Projection](#).

The LSTM method is used to forecast the COVID-19 cases in Bangladesh. From Table 5, we can see that there are three results with its different parameters where the prediction accuracy is quite high. Numerous settings have been tested but not all of them have provided satisfactory results. Reasonable results have only been provided by the number 3 settings. From Fig. 10, it is also noticeable that the accuracy of number 3 result is good. In that, the 1st layer unit is 15, 20, 25, and 30 with the drop out 0.1, 0.1, 0.25, and 0.2, respectively.

Conclusion

Nature plays its role on mankind as history tells us and human wisdom and action that ultimately determine the outcome. In this study, the emphasis has been given more on human actions rather than on the SARS-CoV-2 as well as a human as a host for which very little is known so far. Empirical determination of its association with the probability of an exposed person to get infected omitting ambiguous factors has been accomplished. For successful containment of the epidemic, effective social distancing, hygiene, large-scale testing, and isolation are recommended to ensure at the earliest possible time. The 4P model provides a strong premise in decision making by demonstrating the causality of the epidemic over which the state alongside its citizens has control and the machine learning outcome concretely concludes that here four probabilistic parameters are not just some fitting parameters, rather it is sufficient to train a machine learning model. The roles of the four parameters are globally acknowledged and can be used for

Fig. 10 Plot diagrams for the prediction of cumulative positive cases



other countries as well given the parameter values are determined in their own context. In our experiments, the CDR (case, death, recovery)-only data did not show any promising result; hence, we suggest considering the dynamics we mentioned in their research settings.

Acknowledgments The authors are grateful to Dr. Carmen Zeta Lamagna, Vice Chancellor, American International University - Bangladesh (AIUB), who has inspired us to work on COVID-19 and has patronized us with all sort of cares to make project outcomes publicly available. The authors are also thankful to Md. Manirul Islam, Associate Professor and Director, AIUB Institute of Continuing Education for his untiring support.

Author Contributions This work was carried out in close collaboration between all the authors. Firstly, the research theme was defined by all. Then, they contributed in developing a design of the system which was further refined after initial implementation. All of them had their own contributions in the writing of the paper as well. Finally, all the authors checked and approved the final manuscript.

Compliance with Ethical Standards

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of Interest The authors declare that they have no conflict of interest.

Informed Consent Informed consent was obtained from all individual participants included in the study.


References

- Hays JN, Hays JN. Epidemics and pandemics their impacts on human history. ABC-CLIO, Inc. 2005.
- Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. *J Travel Med* 27(3). 2020.
- Abouk R, Heydari B. The immediate effect of COVID-19 policies on social distancing behavior in the United States. *SSRN Electronic Journal*. 2020.
- Mahmud MS, Kamrujjaman M, Jubyrea J, Islam MS, Islam MS. Quarantine vs social consciousness: a prediction to control COVID-19 infection. *J Appl Life Sci Int* 20–27. Available from: <http://www.journaljalsi.com/index.php/JALSI/article/view/30150>. 2020.
- Scherer CW, Cho H. A social network contagion theory of risk perception. *Risk Anal*. 2003;23(2):261–67.
- Viscusi WK. Alarmist decisions with divergent risk information. *Econ J*. 1997;107(445):1657–70. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1994221 <https://academic.oup.com/ej/article/107/445/1657/5063987>.
- Lerner JS, Gonzalez RM, Small DA, Fischhoff B. Effects of fear and anger on perceived risks of terrorism: a national field experiment. *Psychol Sci*. 2003;14(2):144–50.
- Rudisill C, Costa-Font J, Mossialos E. Behavioral adjustment to avian flu in Europe during spring 2006: the roles of knowledge and proximity to risk. *Soc Sci Med*. 2012;75(8):1362–71.
- Weinbe J, Keaveny P. (eds.) Coronavirus lockdown: fresh data on compliance and public opinion. The Conversation. Available from: <https://theconversation.com/coronavirus-lockdown-fresh-data-on-compliance-and-public-opinion-135872>. 2020.
- Levita L. Initial research findings on the impact of COVID-19 on the well-being of young people aged 13 to 24 in the UK. University of Sheffield, Department of Psychology. Available from: <https://drive.google.com/file/d/1AOc0wCPqv2gfFSQ-DVmw12vrqQK01z0V/view>. 2020.
- Merreffield C. New surveys: support for lockdown hinges on politics and trust in media. *Journalist's Resource*. Available from: <https://journalistsresource.org/studies/politics/polarization/media-trust-politics-lockdown-support/>. 2020.
- Mahfujul S, Marjan H, Rahman S. Fragile democracy, indications of failed state and lack of good governance: Perspective Bangladesh. *J Mass Communicat Journalism*. 2014;4:3.
- Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc Royal Soc London Ser A*. 1927;115(772):700–21. Available from: <https://ui.adsabs.harvard.edu/abs/1927RSPSA.115..700K/abstract>.
- You C, Deng Y, Hu W, Sun J, Lin Q, Zhou F, et al. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *Int J Hygiene Environ Health*. 2020;228:113555. Available from: <https://www.sciencedirect.com/science/article/pii/S1438463920302133> <https://linkinghub.elsevier.com/retrieve/pii/S1438463920302133>.
- Huppert A, Katriel G. Mathematical modelling and prediction in infectious disease epidemiology. *Clinic Microbiol Infect*. 2013;19(11):999–1005. Available from: <http://www.sciencedirect.com/science/article/pii/S1198743X14630019>.
- Li J, Cui N. Dynamic analysis of an SEIR model with distinct incidence for exposed and infectives. *Scientif World J*. 2013;2013:1–5. Available from: <https://www.hindawi.com/journals/tswj/2013/871393/>.
- Bassett MT. The Coronavirus isn't worse in cities because of density. *The New York Times*. Available from: <https://www.nytimes.com/2020/05/15/opinion/sunday/coronavirus-cities-density.html>. 2020.
- Boundless Microbiology. Classifying immunities. Lumencandela. [Access 17 Jun 2020]. Available from: <https://courses.lumenlearning.com/boundless-microbiology/chapter/classifying-immunities/>. 2020.
- Soriano JMM, González A, Soler C, Catalá-Gregori AII, Merino-Torres JFF. Body mass index (BMI) and coronavirus disease 2019 (COVID-19): a living systematic review. *Preprints*. 2020;2019(May):1–15.
- Van Doremalen N, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. 2020.
- Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Internal Med*. 2020;172(9):577–82. Available from: <https://www.acpjournals.org/doi/10.7326/M20-0504>.
- Simpson F. Coronavirus: a personal risk assessment. *Medium*. [Access 17 Jun 2020]. Available from: <https://medium.com/@fergus2/coronavirus-a-personal-risk-assessment-50003c952c03>. 2020.
- Pequeno P, Mendel B, Rosa C, Bosholn M, Souza JL, Baccaro F, et al. Air transportation, population density and temperature predict the spread of COVID-19 in Brazil. *PeerJ*. 2020;8:e9322. Available from: <https://peerj.com/articles/9322>.

24. Qiu Y, Chen X, Shi W. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *J Population Econ* 1–46. Available from: <https://doi.org/10.1007/s00148-020-00778-2>. 2020.
25. Stojkoski V, Utkovski Z, Jolakoski P, Tevdovski D, Kocarev L. The socio-economic determinants of the coronavirus disease (COVID-19) pandemic. *SSRN Electron J*. 2020.
26. Squalli J. Evaluating the determinants of COVID-19 mortality: a cross-country study. *medRxiv*. 20099093. Available from: <https://doi.org/10.1101/2020.05.12.20099093v1>. 2020.
27. Lakshmi Priyadarsini S, Suresh M. Factors influencing the epidemiological characteristics of pandemic COVID 19: a TISM approach. *Int J Healthcare Manag*. 2020;13(2):89–98. Available from: <https://doi.org/10.1080/20479700.2020.1755804>.
28. Ujjie M, Tsuzuki S, Ohmagari N. Effect of temperature on the infectivity of COVID-19. *Int J Infectious Diseases*. 2020;95:301–03.
29. Yao Y, Pan J, Liu Z, Meng X, Wang WW, Kan H, et al. No association of COVID-19 transmission with temperature or UV radiation in Chinese cities. *Europ Respiratory J*. 2020;55(5):2000517. Available from: <https://doi.org/10.1183/13993003.00517-2020>.
30. Ramachandran R. Temperature not a factor in COVID transmission - Frontline. *Frontline*. Available from: <https://frontline.thehindu.com/cover-story/article31403879.ece>. 2020.
31. Jin JM, Bai P, He W, Wu F, Liu XF, Han DM, et al. Gender differences in patients with COVID-19: focus on severity and mortality. *Front Public Health*. 2020;8:152. Available from: <https://www.frontiersin.org/article/10.3389/fpubh.2020.00152/full>.
32. Canning D, Karra M, Dayalu R, Guo M, Bloom DE. The association between age, COVID-19 symptoms, and social distancing behavior in the United States. Available from: <http://medrxiv.org/content/early/2020/04/23/2020.04.19.20065219.abstract>. 2020.
33. Anwar S, Nasrullah M, Hosen MJ. COVID-19 and Bangladesh: challenges and how to address them. *Front Public Health*. 2020;8:154.
34. Kalam A, Hussain AM. Modeling and analysis of the early-growth dynamics of COVID-19 transmission. *Preprints* 1–26. 2020.
35. Bin S, Sun G, Chen CC. Spread of infectious disease modeling and analysis of different factors on spread of infectious disease based on cellular automata. *Int J Environ Res Public Health* 16(23). 2019.
36. Honjo K. Social epidemiology: definition, history, and research examples. *Environ Health Prevent Med*. 2004;9(5):193–99. Available from: <https://doi.org/10.1007/BF02898100>.
37. Fang Y, Nie Y, Penny M. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: a data-driven analysis. *J Med Virol*. 2020;92(6):645–59. Available from: <https://doi.org/10.1002/jmv.25750>.
38. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med*. 2020;18(1):124. Available from: <https://doi.org/10.1186/s12916-020-01597-8>.
39. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci USA*. 2020;117(17):9241–43.
40. Abdullah M. Bangladeshi researchers decode the Sars-Cov-2 genome sequence. Dhaka: Dhaka Tribune. [Access 15 Jul 2020]. Available from: <https://www.dhakatribune.com/health/coronavirus/2020/05/13/bangladeshi-researchers-decode-the-sars-cov2-genome-sequence>. 2020.
41. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*. 2020;368(6493):860–68. Available from: <https://science.sciencemag.org/content/368/6493/860.abstract>.
42. Prompetchara E, Ketloy C, Palaga T. Immune responses in COVID-19 and potential vaccines: lessons learned from SARS and MERS epidemic. *Asian Pac J Allergy Immunol*. 2020;38(1):1–9.
43. Lakdawala S. What we do and do not know about COVID-19's infectious dose and viral load. *Discover*. Accessed 15 Jul 2020. Available from: <https://theconversation.com/what-we-do-and-do-not-know-about-covid-19s-infectious-dose-and-viral-load-135991>. 2020.
44. Hogan A. How much of the coronavirus does it take to make you sick? *Stat News*. [Accessed 15 Jul 2020]. Available from: <https://www.statnews.com/2020/04/14/how-much-of-the-coronavirus-does-it-take-to-make-you-sick/>. 2020.
45. MacLachlan J, Cowie B. The positives and negatives of mass testing for coronavirus. *Medical Press*. [Accessed 15 Jul 2020]. Available from: <https://medicalxpress.com/news/2020-05-positives-negatives-mass-coronavirus.html>. 2020.
46. Organization WH. Coronavirus disease (COVID-19) Situation Report-111. World Health Organization. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>. 2020.
47. Ritchie H, et al. Coronavirus pandemic (COVID-19) - our world in data. *Statistics and Research*. [Accessed 15 Jul 2020]. Available from: <https://ourworldindata.org/coronavirus>. 2020.
48. Islam MS, Irana Ira J, Ariful Kabir KM, Kamrujjaman M. COVID-19 epidemic compartments model and Bangladesh. *Preprint*. Available from: www.preprints.org. 2020.
49. Truelove S, Abraham O, Altare C, Lauer SA, Grantz KH, Azman AS, et al. The potential impact of COVID-19 in refugee camps in Bangladesh and beyond: a modeling study. *PLOS Med*. 2020;17(6):e1003144. Available from: <https://doi.org/10.1371/journal.pmed.1003144>.
50. University JH. Timeline of COVID-19 policies, cases, and deaths in your state - Johns Hopkins Coronavirus Resource Center. Johns Hopkins University & Medicine. [Accessed 29 Jun 2020]. Available from: <https://coronavirus.jhu.edu/data/state-timeline/new-confirmed-cases/alabama/28>. 2020.
51. Xiaofeng Y, Lin L, Yalin W. Nonlinear dynamic soft sensor modeling with supervised long short-term memory network. *IEEE Trans Indust Inform*. 2020;16(5):3168–76. Available from: <https://ieeexplore.ieee.org/document/8654687>.

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

Affiliations

Khandaker Tabin Hasan¹  · **M. Mostafizur Rahman¹** · **Md. Mortuza Ahmmed¹** · **Anjir Ahmed Chowdhury¹** · **Mohammad Khairul Islam¹**

M. Mostafizur Rahman
mostafiz.math@aiub.edu

Md. Mortuza Ahmmed
mortuza@aiub.edu

Anjir Ahmed Chowdhury
anzira431@gmail.com

Mohammad Khairul Islam
rulangbd@gmail.com

¹ American International University-Bangladesh,
Dhaka, Bangladesh