Coronavirus (COVID-19): ARIMA-based Time-series Analysis to Forecast near Future and the Effect of School Reopening in India

Hiteshi Tandon¹, Prabhat Ranjan², Tanmoy Chakraborty³ and Vandana Suhag⁴

Abstract

COVID-19, a novel coronavirus, is currently a major worldwide threat. It has infected more than a million people globally leading to hundred-thousands of deaths. In such grave circumstances, it is very important to predict future scenario to support prevention and recurrence of the disease, aid in healthcare service preparation and help in decision making process. Following that notion, a model has been developed for forecasting future COVID-19 cases in India. The time series analysis indicates that the cases will keep on increasing in India in the coming month as the peak has not been attained until now. A statistical analysis based on the effect of reopening of schools has also been performed. It is revealed that there will be a minor increase in the active cases when pre-/primary schools are opened. The present prediction models will assist the government and medical personnel in gaining insight and planning for forthcoming conditions.

Keywords

COVID-19, ARIMA, epidemic, forecasting, school, India

Introduction

The pandemic of 2019-nCov commenced from December 2019 in Wuhan, China and has caused extreme havoc in almost the whole world (World Health Organization [WHO], 2020a; Zhu et al., 2020). 2019-nCoV or COVID-19, commonly known as coronavirus, is a novel highly contagious virus belonging to Coronaviridae family that has been suspected to be transmitted to humans from animals. This virus causes mild to severe respiratory illness and death (Paules et al., 2020). This pandemic has engulfed 188 countries/regions in 9 months infecting 33,423,249 people and taking the death toll to 1,003,008 (Dong et al., 2020). However, the premature cases show the infection is less severe as compared to other coronaviruses like severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), the cases of rapid human-to-human transmission signify that 2019-nCoV is highly infectious than others (Team, 2020). Although a local seafood market in Wuhan is believed to be the source of exposure (Huang et al., 2020), the scope of occurrence of this disease is not clear since its occurrence at present is so dynamic (Paules et al., 2020). Recently, a review on this novel disease has also been published (Singhal, 2020). An apparent variation is present in epidemiological examinations and detection abilities performed by different countries for detecting infected cases (Niehus et al., 2020). Presently, the highest cases of 2019-nCoV infections have been reported in USA, however, the cases have been abruptly rising in India, Brazil, Russia and Colombia (JHU & CSSE, 2020). China, the place of origin of the disease, is now receiving very few cases (Dong et al., 2020). The first case of coronavirus infection in India was reported on 30 January 2020 in Kerala, which was an imported case from Wuhan city of China (Unnithan, 2020). In the initial phase, the spread was extremely slow and only 3 people were positive for more than a month. However, the numbers started rising exponentially after 1 month and continue to do so. The numbers in India have reached up to 6,145,291 for confirmed COVID-19 infected cases with 96,318 deaths and 5,101,397 recoveries as reported on 28 September 2020 (Dong et al., 2020). In spite of rigorous

¹ Department of Chemistry, Manipal University Jaipur, Jaipur, Rajasthan, India
² Department of Mechatronics Engineering, Manipal University Jaipur, Jaipur, Rajasthan, India
³ Department of Chemistry and Biochemistry, School of Basic Sciences and Research, Sharda University, Greater Noida, Uttar Pradesh, India
⁴ Department of Applied Sciences, BML Munjal University, Gurugram, Haryana, India

Corresponding authors:
Hiteshi Tandon, Department of Chemistry, Manipal University Jaipur, Jaipur, Rajasthan 303007, India.
E-mail: hitseshitandon@yahoo.co.in
Tanmoy Chakraborty, Department of Chemistry and Biochemistry, School of Basic Sciences and Research, Sharda University, Greater Noida, Uttar Pradesh 201310, India.
E-mails: tanmoychem@gmail.com; tanmoy.chakraborty@sharda.ac.in
researches being conducted in the almost every country, there
is neither a definite treatment nor a vaccination for the
COVID-19 infection until now. At present, it is a major health
crisis around the world and it would not be wrong to say that
it is ‘an enemy to humanity’. In this circumstance, the only
option is preventing the occurrence of infection and preparing
our healthcare and financial systems for the probable up-comings.

In that reference, it is extremely crucial to construct models
that are computationally competent and realistic so that they
can help policy makers, medical personnel and also general
public. Modelling the disease and providing future forecast of
possible number of daily cases can assist the medical system
in getting prepared for the new patients. The government can
also get an idea of the forthcoming situation according to
which they can plan and make decisions efficiently. The
statistical prediction models are useful in forecasting and
controlling the global epidemic threat.

Recently, a machine learning approach has been employed
to predict the COVID-19 pandemic in India based on China’s
infection spread pattern (Tiwari et al., 2020). A short-term
COVID-19 spread forecast has also been presented for the
two most affected states of India, namely Maharashtra,
Tamil Nadu and Delhi. In order to mitigate the transmission
of the disease, a mathematical model has been developed
incorporating a quarantine class and government intervention
mechanics (Mandal et al., 2020a). The examination of real-
time COVID-19 infection data of as many as nine countries
suggests that the shift of the data from an exponential form to
a power-law form indicates the curve is being flattened
(Verma et al., 2020). A model has been specifically developed
for planning intervention strategies mainly focussing on the
four metro cities of India, viz. Delhi, Mumbai, Kolkata and
Bengaluru, that have intercity connectivitie (Mandal et al.,
2020b). Further, for analysing and predicting the number of
COVID-19 spread in India, genetic evolutionary programming
(GEP) has been utilized (Salgotra et al., 2020). Another
COVID-19 forecasting model was developed using machine
learning approach to study the spread in India (Sujath et al.,
2020). LSTM technique has also been useful to form a data-
driven COVID-19 model which examines the consequence of
social isolation, lockdown and effect of transformation ratio.
It also presents the forecast of the upcoming probable
condition (Tomar & Gupta, 2020).

As compared to other prediction models, auto-regressive
integrated moving average (ARIMA) models are more
able in the prediction of natural adversities (Zhang et al.,
2019). The prevalence of COVID-19 in France, Italy and
Spain has been successfully forecasted employing the ARIMA
model (Ceylan, 2020). For estimating the confirmed COVID-
19 cases in Brazil, short-term forecasting has been carried out
using ARIMA modelling (Ribeiro et al., 2020). Further, a
hybrid ARIMA-WBF (wavelet-based forecasting) model has
also been formed for COVID-19 case forecasting (Chakraborty
& Ghosh, 2020). In view of this fact, in the present effort, we
have employed ARIMA model for predicting the incidence of
2019-nCov disease. For our study, we have identified the best
ARIMA model and then predicted the number the cases for
the next 30 days. This model can be tuned for making further
forecasts incorporating the real-time data.

Worldwide countries are gradually introducing policies
and measures to ease their lockdown limitations so as to
support their economy and alleviate societal pressure along
with managing the spread of COVID-19. Among many
measures, reopening of schools is one of the most important
aspects which needs urgent attention. An extended closure of
schools may lead to grave social, health and economic
imbalances, especially in countries with limited income.
During Ebola outburst from 2014 to 2016 in some countries,
shutting schools has been linked to augment child labour,
socioeconomic issues and violence (Bayham & Fenichel,
2020). In various nations, distance learning strategies are
being employed using digital technologies as a substitute for
conventional school. However it seems complicated to
implement such learning alternatives even in the developed
countries. In a National Institute of Statistics 2015 survey
conducted in Italy (Italian National Institute of Statistics,
2015), it has been observed that 41% of the families residing
in the poorest parts of the country did not own a personal
computer or a tablet. Further, among the households having a
minimum of one child, distance learning was guaranteed by
merely 14.3% families. These statistics clearly indicate that a
considerable section of children would remain greatly
excluded from education and from any kind of socialization
with peers and the world (Esposito & Principi, 2020),
especially in a low-income developing country like India.
Thus, it is vital to assess the possibility of infection spread
through children on school reopening in order to understand
the outcomes and plan the opening of schools accordingly.

Estimating the danger that reopening of schools may have
in the diffusion of the virus poses a critical challenge, since
the part children play in the spread of this virus has not been
understood properly until now. Nevertheless it has been
observed that children having age below 10 years present a
lower COVID-19 prevalence in comparison to the young and
adults (Gudbjartsson et al., 2020; Lavezzo et al., 2020). The
risk of infection spread among children in a classroom has
been proved to be extremely low in Taiwan having an $R_0 < 1$,
which evidently stresses that closing of schools may perhaps
be just marginally useful (Wang et al., 2020). In a COVID-19
modelling study for UK conducted utilizing data from the
Wuhan province of China, it has been estimated that merely
school closings would avert only 2%–4% of deaths which is
insignificant when compared to interventions like social
distancing (Ferguson et al., 2020). An exceptionally small
percentage of children, less than 5%, have been confirmed to
be COVID-19 infected worldwide; they present milder
symptoms and a better prognosis than adults (Ludvigsson,
So far epidemic data does not present any commonness with the preceding influenza outbreaks that contributed towards the infection spread in the community (Centers for Disease Control and Prevention [CDC], 2009; Guinard et al., 2009; Huai et al., 2010; Smith et al., 2009). Though it is possible that such spread may not have been noticed due to impulsive school closures or asymptomatic infections in children or testing restricted only to symptomatic cases in the initial phase.

Sweden did not close its schools while Taiwan has reopened its schools implementing preventive measures (The Local, 2020). Two countries, Denmark and Norway, have also opened their primary schools again (BBC News, 2020; Charlton et al., 2020). In India, reopening of schools has been suggested to occur in October, although it may be postponed based on the progression of the pandemic. Thus, the two main objectives of the study are to find the best predictive model and to apply it in forecasting future incidence of COVID-19 cases and to estimate the COVID-19 infection transmission due to reopening of schools and its effect on the control of pandemic in the approaching months in India. As per our knowledge, this is the first attempt to estimate the spread of infection due to school reopening and study its effect in the evolution of the disease in India.

**Methodology**

**ARIMA Modeling for COVID-19 Spread Prediction**

**Dataset**

Confirmed, recovered and death cases of COVID-19 infection have been collected for countries with highest confirmed infection, namely USA, India, Brazil, Russia and Columbia, and countries in South-East Asia region (India, Bangladesh, Indonesia, Maldives, Nepal, Thailand, Sri Lanka, Bhutan and Timor-Leste), as per WHO region classification, from the COVID-19 Data Repository by the Centre for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU CSSE COVID-19 Data) (https://github.com/CSSEGISandData/COVID-19) from 22 January 2020 to 28 September 2020. This data has been used to build predictive models.

**Model Development**

For forecasting a time series, ARIMA modelling is one of the best and classical modelling techniques. ARIMA models are always represented with the help of some parameters and the model is expressed as ARIMA \((p, d, q)\). Here, \(p\) stands for the order of auto-regression, \(d\) signifies the degree of trend difference while \(q\) is the order of moving average. We have applied an ARIMA model to the time series data of confirmed COVID-19 cases in India. Autocorrelation function (ACF) graph and partial autocorrelation (PACF) graph is used to find the initial number of ARIMA models. These ARIMA models are then tested for variance in normality and stationary. Next, they are checked for accuracy by observing their mean absolute percentage error (MAPE), mean absolute deviation (MAD) and multiple seasonal decomposition (MSD) values to determine the finest model to forecast. In addition, the best fit ARIMA model is compared with linear trend, quadratic trend, moving average and single exponential models using an output of measure of accuracy, namely MAPE, MAD and MSD, so as to select the finest model to forecast. The finest model is the one which has the lowest value for all the measures. After fitting the model, its parameters are estimated followed by verification of the model. The built model is employed to forecast confirmed COVID-19 cases for the next 30 days, that is, 29 September 2020 to 28 October 2020. ARIMA model is generally represented as,

\[
ARIMA(p, d, q): X_t = C + \alpha_t X_{t-1} + \alpha_t X_{t-2} + \cdots + \alpha_t X_{t-p} + \beta_t Z_{t-1} + \beta_t Z_{t-2} + \cdots + \beta_t Z_{t-q} + Z_t
\]

where,

\[
Z_t = X_t - \alpha_t X_{t-1} - \beta_t Z_{t-1}
\]

Here, \(X_t\) is the predicted number of confirmed COVID-19 cases at \(t^{th}\) day, \(C\) is the constant, \(\alpha\) and \(\beta\) are parameters whereas \(Z_t\) is the residual term for \(t^{th}\) day. The trend of forthcoming incidences can be estimated from the previous cases and a time series analysis is performed for this purpose. Time series forecasting refers to the employment of a model to forecast future data based on previously observed data (Box & Jenkins, 1970). In the present study, time series analysis is used to recognize the trends in confirmed COVID-19 cases in India over the period of 30 January 2020 to 28 September 2020 and to predict future cases from 29 September 2020 till 28 October 2020. The level of statistical significance is set at 0.05. A graph is plotted for actual confirmed cases and predicted confirmed cases with respect to time to verify the efficiency of the model. To get an idea of the recovery and death trends in India, a graph is plotted with respect to time.

A comparative study is also performed to examine the status of confirmed COVID-19 cases of India with respect to those of other highly infected countries. A similar comparison is made with the countries of South-East Asia region as well. All the model developments, computations and comparisons have been performed using Minitab 17 Statistical Software (Minitab, 2010).

**Effect of Reopening of Schools and Day Cares on COVID-19 Spread**

**Dataset**

Population of India has been collected from the population pyramid (2020) differentiated into several age groups for the
year 2020. We have specifically used the data of population between 0–19 years (children and youth) and 20–64 years (adults) for our analysis. Again, the confirmed, recovered and death cases of COVID-19 infection in India have been gathered from the JHU CSSE COVID-19 Data repository (https://github.com/CSSEGISandData/COVID-19) from 30 January 2020 to 28 September 2020. These data have been used to estimate and study the effect of school reopening on the spread of COVID-19.

**Method of Analysis**

As discussed in the introduction section, there are numerous indications which imply that the role of younger children in transmission of COVID-19 is weaker than adolescents and adults. However in view of the uncertainties this pandemic surrounds, in our study, we have considered children and adults to be equally vulnerable as a precautionary measure. As observed from the previous studies (Ludvigsson, 2020), it is clear that merely 5% of the total infections have been contributed by the children and the rest are due to adults. Moreover, the cases of deaths in India have been negligible in case of children. Based on this fact, it can be suggested that the 5% of total confirmed COVID-19 cases of India are either recovered or active cases of children (0–19 years) which have become immune or will become immune.

Further, the remaining 95% cases constitute infection in adults. In general, majority of the deaths have been noticed in individuals who are equal to or above the age of 65 along with some cases having severe medical conditions. Thus, we can assume that all the deaths have been more or less in older people. Accordingly, we have reduced the number of death cases from the number of confirmed infection cases in adults (remaining 95% of the total confirmed cases), which provides us the number of active/recovered cases in adults (20–64 years). These adults can be considered to be immune to COVID-19.

The total population of children in India within the age 0–19 years is 487,063,151 (Population Pyramid, 2020). On subtracting the number of immune children from this figure, we have obtained the number of children who are at risk of contracting the infection. Similarly, the total population of individuals between the age 20 and 64 years is 802,221,281 (Population Pyramid, 2020). We have reduced the number of adults immune to COVID-19 from this figure to get the number of people who are still susceptible to infection. It is understandable that not every child would attend school at once when reopened, primarily due to the family’s fear of getting infected. Further, the population of vulnerable children also includes those who never went to school due to their low economic status. Taking into account such circumstances, we have considered that only 80% of the susceptible children population will be actually at the risk of infection when the schools are reopened. In addition, only people up to the age of 64 years have been recommended to leave their house for any work. It means the population working at different places in these times ranges between 20 and 64 years. Among the total adults who are prone to risk, about 80% of them have started working again since 1 June 2020 while some are still at home due to certain activity closures, telework, taking care of children and other reasons. Accordingly, it can be concluded that the recent variations in the infection spread are possibly due to these workers.

Number of COVID-19 confirmed cases comprise of active, recovered and deceased cases. When the number of recovered and deceased cases is eliminated from the confirmed cases, we are left with the number of active cases. Next, on differencing the previous day total active cases from the current day total active cases, we obtain the active cases per day. Our approach is based on the estimation of the effect of school reopening on COVID-19 spread through analysing the effect of recent workplace reopening. We have calculated the active cases percentage per day starting from 1 June 2020 to 28 September 2020 assuming that the cases during this time are a result of transmission in workers. This may be represented as,

\[
\text{Percent active cases per day} = \frac{\text{Number of active cases per day due to reopening of workplaces}}{\text{Actual susceptible working population}} \times 100
\]  

(3)

As mentioned earlier, we have assumed in our study that children are capable of transmitting the disease as much as adults and when the schools are opened, the infection pattern will be similar to that of opening of workplaces. Thus, using the percent active cases per day of working population for school going children, we have estimated the probable active cases per day due to school reopening. This can be expressed as:

\[
\text{Number of active cases per day due to reopening of schools} = \frac{\text{Actual susceptible children population} \times \text{Percent active cases per day}}{100}
\]  

(4)

We have calculated the probable active cases per day caused by school opening for 120 days along with a 95% confidence interval (CI). All the computations have been performed on the Minitab 17 Statistical Software (Minitab, 2010).
Result and Discussion

ARIMA Model

The present work encompasses development of a model to forecast COVID-19 incidences in the coming days. The results for measure of model accuracy for ARIMA, linear trend, quadratic trend, moving average and single exponential model are presented in Table 1. A look at the MAPE, MAD and MSD values suggests that ARIMA (1, 1, 1) model is the most accurate of all for forecasting future incidences as it possesses lower value for all the measures.

Thus, parameters are estimated for the ARIMA (1, 1, 1) model which are displayed in Table 2. It is observed that AR (1) and MA (1) parameters have a p-value of 0.000, thus implying that the parameters are significant in the model.

Figure 1 displays the residual plots for confirmed COVID-19 cases in India from 30 January 2020 to 28 September 2020. A minor deviation of residuals from the straight line can be observed from the plot. This indicates that the errors are somewhat near to normal although with a few outliers. Therefore, the normality assumption can be assumed to be followed. The residual histogram backs up this assumption.

Table 1. Measures of Model Accuracy.

| Models                          | MAPE | MAD    | MSD            |
|---------------------------------|------|--------|----------------|
| ARIMA (1, 1, 1)                 | 5    | 25073  | 161152686      |
| Single exponential method       | 43   | 13536  | 472399510      |
| Moving average (MA)             | 6    | 25394  | 1643397687     |
| Quadratic trend model           | 1963300 | 287057 | 108119000000   |
| Linear trend model              | 5531990 | 794465 | 901508000000   |

Source: The authors.

Table 2. Parameters Estimates of the ARIMA Model.

| Type  | Coef   | SE Coef | t      | p-Value |
|-------|--------|---------|--------|---------|
| AR (1) | 1.0065 | 0.0032  | 313.81 | 0.000   |
| MA (1) | 0.4709 | 0.0605  | 7.79   | 0.000   |

Source: The authors.

Figure 1. Residual Plots for Confirmed COVID-19 Cases in India from 30 January 2020 to 28 September 2020.

Source: JHU CSSE COVID-19 Data repository (https://github.com/CSSEGISandData/COVID-19).
The graph between residuals and the fitted values displays a very little dispersion. This implies that the assumption of constant variance is also satisfied by the model. The non-correlation of residuals is clear from the plot of residuals versus the order of the data. The suitability of the ARIMA (1, 1, 1) model is indicated by the significance of \( p \)-value and other statistics.

The workable model obtained after the substitution of estimated parameters is represented as,

\[
X_t = 1.0065 X_{t-1} - 0.4709 Z_{t-1} + Z_t
\]

ARIMA (1, 1, 1) model (Equation (5)) is used to forecast confirmed COVID-19 cases in India for the next 30 days, that is, 29 September 2020 to 28 October 2020. The forecast for cases is presented in Table 3 with 95% CI. According to the forecast, the number of confirmed COVID-19 cases has not attained the peak in India yet and it will keep on increasing in the coming months as expected from the current circumstances. Similarly, time series analysis presents the meaningful statistics for confirmed COVID-19 data. Figure 2 presents the time series graph of the confirmed COVID-19 cases from 30 January 2020 to 28 October 2020. It is clear from the plot that

Table 3. Figures for Forecasted Confirmed COVID-19 Cases and their Lower and Upper Limits for 30 Days (29 September 2020 to 28 October 2020) with 95% CI.

| Date               | Forecast   | Lower Limit | Upper Limit |
|--------------------|------------|-------------|-------------|
| 29 September 2020  | 6,223,374  | 6,216,087   | 6,230,660   |
| 30 September 2020  | 6,301,964  | 6,288,612   | 6,315,316   |
| 01 October 2020    | 6,381,065  | 6,360,896   | 6,401,234   |
| 02 October 2020    | 6,460,681  | 6,432,924   | 6,488,437   |
| 03 October 2020    | 6,540,814  | 6,504,736   | 6,576,893   |
| 04 October 2020    | 6,621,469  | 6,576,374   | 6,666,563   |
| 05 October 2020    | 6,702,648  | 6,647,878   | 6,757,418   |
| 06 October 2020    | 6,784,355  | 6,719,281   | 6,849,429   |
| 07 October 2020    | 6,866,593  | 6,790,612   | 6,942,574   |
| 08 October 2020    | 6,949,366  | 6,861,896   | 7,036,836   |
| 09 October 2020    | 7,032,677  | 6,933,156   | 7,132,199   |
| 10 October 2020    | 7,116,530  | 7,004,411   | 7,228,649   |
| 11 October 2020    | 7,200,929  | 7,075,680   | 7,326,177   |
| 12 October 2020    | 7,285,876  | 7,146,979   | 7,424,772   |
| 13 October 2020    | 7,371,375  | 7,218,322   | 7,524,428   |
| 14 October 2020    | 7,457,431  | 7,289,723   | 7,625,139   |
| 15 October 2020    | 7,544,046  | 7,361,194   | 7,726,898   |
| 16 October 2020    | 7,631,224  | 7,432,747   | 7,829,702   |
| 17 October 2020    | 7,718,969  | 7,504,392   | 7,933,546   |
| 18 October 2020    | 7,807,285  | 7,576,141   | 8,038,430   |
| 19 October 2020    | 7,896,175  | 7,648,002   | 8,144,349   |
| 20 October 2020    | 7,985,644  | 7,719,984   | 8,251,304   |
| 21 October 2020    | 8,075,694  | 7,792,096   | 8,359,292   |
| 22 October 2020    | 8,166,329  | 7,864,346   | 8,468,313   |
| 23 October 2020    | 8,257,554  | 7,936,741   | 8,578,367   |
| 24 October 2020    | 8,349,372  | 8,009,290   | 8,689,454   |
| 25 October 2020    | 8,441,788  | 8,081,999   | 8,801,576   |
| 26 October 2020    | 8,534,804  | 8,154,876   | 8,914,732   |
| 27 October 2020    | 8,628,425  | 8,227,926   | 9,028,924   |
| 28 October 2020    | 8,722,655  | 8,301,156   | 9,144,154   |

Source: The authors.
Figure 2. Times Series Plot for Confirmed COVID-19 Infections in India from 30 January 2020 to 28 October 2020 (blue line represents actual confirmed cases and yellow lines represent case forecasts).

Source: JHU CSSE COVID-19 Data repository (https://github.com/CSSEGISandData/COVID-19) and The authors.

Figure 3. A Comparative Times Series Plot for Actual Confirmed and Forecasted COVID-19 Cases from 30 January 2020 to 28 September 2020.

Source: JHU CSSE COVID-19 Data repository (https://github.com/CSSEGISandData/COVID-19) and The authors.
the time series is not stationary. An increasing trend is displayed by the time series suggesting further rise in COVID-19 cases.

For testing the efficiency of our forecast, a time series graph is plotted for the actual and forecasted confirmed COVID-19 cases starting from 30 January 2020 till 28 September 2020. The plot is represented by Figure 3. The similarity of forecasted data with actual data is clear from these plots. This comparison reveals the precision of the model in forecasting.

Trend for the number of recovered and deceased cases with respect to time due to COVID-19 infections in India depicted in Figure 4. It is observed that the number of recoveries drastically increase with time while the rate of deceased cases is very low. Thus, a very low mortality rate is expected from the disease.

In the initial phase of infection, the cases were mainly due to the infected people returning from abroad. However, some of those people contributed in the spreading of infection by not following the 14-day quarantine protocol. This was the time when imposing a lockdown could have easily controlled the infection spread and thus, the first lockdown in the country was imposed on 24 March 2020. Just before the country went under lockdown, an unexpected turn of events occurred which adversely affected the pandemic control planning resulting in a sudden surge in infection. Even after imposing a lockdown, India observed a continuous rise in cases (Editorial, 2020; Pulla, 2020). The cause of this rise is linked to the people involved in a large community gathering. Many of them have been tested positive while a number of them are still untraceable. Looking at the circumstances, the lockdown was extended till 3 May 2020 in order to cope up with the situation. It can be seen from the Figure 2 that the cases still kept rising at a hasty pace. Another huge matter of concern was the migration of workers which occurred due to misleading and inaccurate information. This further increased the number of transmissions and infections countrywide.

It is concluded that the negligence on the part of a few people who did not follow the suggested 14-day isolation after returning from abroad and asymptomatic cases with/without a travel history have led to the infection spread although the major turning point of India’s current situation is the mass community gathering, followed by migrating workers. These people have caused extreme transmissions and it is highly suspected that there are many asymptomatic cases which have still not been tested. Apart from that, until now, it has not been confirmed whether a recovered person can act as a carrier of the virus or not. Further, if it is possible, then for how long. To some extent, social media has also contributed towards some cases owing to the fake information being spread through the platform. Thus, it is very important to control such activities and movements if we need to flatten the infection curve.

**Figure 4.** A Comparative Trend for the Number of Recovered and Deceased Cases Due to COVID-19 Infections in India from 30 January 2020 to 28 September 2020.

**Source:** JHU CSSE COVID-19 Data repository (https://github.com/CSSEGISandData/COVID-19).
According to the circumstances, the lockdown was extended further two more times, that is, from 4 May 2020 to 17 May 2020 and from 18 May 2020 to 31 May 2020. Although, both these lockdowns were imposed incorporating some modifications according to the extent of spread in different parts of the country. Such a long shutting down of the country has hampered the economy but it might have helped in slowing down the growth rate of infection as well which would have otherwise multiplied severely.

Looking at the extreme number of current and predicted infected cases, it can be easily concluded that the virus cannot be completely eliminated from our life at least for some time. Currently, there is no vaccine and further increase in lockdown cannot help in containing the virus, in fact it will worsen the economic status. Thus, it is very crucial to plan and implement certain rules and policies strategically which can support the growth of India’s economy and help people in getting their life back to track, simultaneously developing healthcare systems as well. It is suggested that all places must be reopened and all activities should be resumed to normal although imposing certain preventive measures. A responsible strategy preparation is highly crucial for an efficient exit (Gilbert et al., 2020). All public places are required be sanitized at regular intervals. Using of masks, sanitizer liquids or sprays and following social distancing rules needs to be continued for the coming months responsibly. House-to-house or area wise testing for COVID-19 is important to be conducted every month. Following concepts like opening of alternative shops or work from home and so on can also help reduce the risk of exposure.

It has been observed that people are unaware of when and where to go for COVID-19 testing; they lack information and are scared. In fact, hospitals are not accepting patients who may be/are COVID-19 positive. There is no information regarding the COVID-infrastructure and no facility to know if or where beds or ventilators are available. Moreover, some labs are constantly reporting false negatives. This is a very huge issue needs to be tackled quickly and appropriately. The correction of false negative reports may be done through serological testing. It might also help in the detection of potentially protected and non-infectious individuals and in quantifying the part of population which contributes towards ‘herd immunity’ (Dewatripont et al., 2020). Testing must be prioritized for people involved in essential services followed by other people and ultimately the low-risk individuals like children. Advancing in this manner would help in gradually building up ‘herd immunity’. This is crucial in order to prevent or decrease the intensity of upcoming pandemic waves (Dewatripont et al., 2020). It is suggested that public awareness programs must be started to help people understand what should be done in the situation. Public bulletins must be issued and ad campaigns must be started to generate knowledge among masses. Contact tracing must be implemented at a large scale (Dewatripont et al., 2020).

Figure 5. A Comparative Plot of Confirmed COVID-19 Cases in USA, India, Brazil, Russia and Colombia from 22 January 2020 to 28 September 2020.

Source: JHU CSSE COVID-19 Data repository (https://github.com/CSSEGISandData/COVID-19).
Number of COVID-19 testing labs must be increased and every hospital should be obligated to have a separate ward for such cases. Labs providing constantly incorrect results must be removed from the COVID-19 testing centres list. Further, software can be developed which integrates information regarding all the hospitals in every city and their available facilities, including equipments like ventilators. Medical infrastructure must be strengthened and adequate funding should be provided to carry out research based on COVID-19. If only people are made aware of the situation and all these measures are imposed strictly, we can expect to come out of this situation in a short time together with maintaining our well-being. Only if stringent actions are taken, the rate of new infected cases could be expected to begin declining in the coming months.

Figure 5 shows a comparative study of confirmed COVID-19 infection cases of top five highly infected countries of the world. According to the plot, USA is the most infected among all followed by India. Due to enormous population and high infection rate, it is probable that India may soon become the top infected country around the globe. Next in the order are Brazil, Russia and Columbia. Brazil appears to be moving at a high pace of infection while Russia seems to be approaching curve flattening. Infection cases have started to rise in Columbia and they may surpass Russia in the coming months. As very well-known, China has been able to control the pandemic and is now presenting a very few new cases. Looking at the huge population of India, it is extremely crucial that strict prevention measures are continued so that the situation could be controlled in a short period of time. Otherwise, it will become severely difficult to control this hasty infection spread.

A similar comparison is performed for the countries of South-East Asia region and shown in Figure 6. A look at the Figure 6 suggests India to be the most infected among the South-East Asian countries followed by Bangladesh, Indonesia and Maldives. All the four countries are presenting continuous rise in confirmed COVID-19 infections particularly India. Maldives and Nepal have also been observing a surge in their infection rate lately as evident from Figure 6. Thailand, Sri Lanka and Bhutan have been very capable in controlling the spread and are now presenting negligible cases. Another country to be noted is Timor-Leste which was affected the last among all South-East Asian countries and now has zero active cases.

It is clear that measures such as using protective mask, sanitization and social distancing can reduce the chances of exposure to virus and help in containing this pandemic. Thus, these measures should be stringently imposed and followed in India and strict actions must be taken against those people who violate the rules and do not consider the severity of the situation. This prediction model could be valuable in anticipating future cases of infection, if the pattern of virus spread did not change abnormally. It is obvious that this virus
Table 4. Figures for Predicted COVID-19 Active Cases Owing to School Reopening Alone and their Lower and Upper Limits for 120 Days (1 June 2020 to 28 September 2020) with 95% CI.

| Date       | Predicted Active Cases | Lower Limit | Upper Limit |
|------------|------------------------|-------------|-------------|
| 01 June 2020 | 2,222                  | 2,111       | 2,333       |
| 02 June 2020 | 2,471                  | 2,347       | 2,595       |
| 03 June 2020 | 3,393                  | 3,224       | 3,563       |
| 04 June 2020 | 3,179                  | 3,020       | 3,338       |
| 05 June 2020 | 2,673                  | 2,540       | 2,807       |
| 06 June 2020 | 2,842                  | 2,699       | 2,984       |
| 07 June 2020 | 3,310                  | 3,144       | 3,475       |
| 08 June 2020 | 1,779                  | 1,690       | 1,868       |
| 09 June 2020 | 2,652                  | 2,519       | 2,784       |
| 10 June 2020 | −57                    | −54         | −60         |
| 11 June 2020 | 4,986                  | 4,737       | 5,236       |
| 12 June 2020 | 2,391                  | 2,272       | 2,511       |
| 13 June 2020 | 2,168                  | 2,059       | 2,276       |
| 14 June 2020 | 2,282                  | 2,168       | 2,397       |
| 15 June 2020 | 44                     | 42          | 46          |
| 16 June 2020 | 1,244                  | 1,182       | 1,307       |
| 17 June 2020 | 3,132                  | 2,976       | 3,289       |
| 18 June 2020 | 1,740                  | 1,653       | 1,827       |
| 19 June 2020 | 3,050                  | 2,897       | 3,202       |
| 20 June 2020 | 729                    | 692         | 765         |
| 21 June 2020 | 2,987                  | 2,838       | 3,137       |
| 22 June 2020 | 2,203                  | 2,093       | 2,313       |
| 23 June 2020 | 3,042                  | 2,890       | 3,194       |
| 24 June 2020 | 2,121                  | 2,015       | 2,227       |
| 25 June 2020 | 1,791                  | 1,702       | 1,881       |
| 26 June 2020 | 4,814                  | 4,573       | 5,054       |
| 27 June 2020 | 3,441                  | 3,269       | 3,613       |
| 28 June 2020 | 4,294                  | 4,080       | 4,509       |
| 29 June 2020 | 3,041                  | 2,889       | 3,193       |
| 30 June 2020 | 3,064                  | 2,911       | 3,218       |
| 01 July 2020 | 4,118                  | 3,912       | 4,324       |
| 02 July 2020 | 299                    | 284         | 314         |
| 03 July 2020 | 4,857                  | 4,614       | 5,100       |
| 04 July 2020 | 5,700                  | 5,415       | 5,985       |
| 05 July 2020 | 5,148                  | 4,891       | 5,405       |
| 06 July 2020 | 3,818                  | 3,627       | 4,009       |
| 07 July 2020 | 3,265                  | 3,101       | 3,428       |
| 08 July 2020 | 2,944                  | 2,797       | 3,091       |
| 09 July 2020 | 4,190                  | 3,981       | 4,400       |
| 10 July 2020 | 4,085                  | 3,880       | 4,289       |
| 11 July 2020 | 5,362                  | 5,094       | 5,630       |
| 12 July 2020 | 5,699                  | 5,414       | 5,984       |
| Date         | Predicted Active Cases | Lower Limit | Upper Limit |
|--------------|------------------------|-------------|-------------|
| 13 July 2020 | 6,050                  | 5,748       | 6,353       |
| 14 July 2020 | 5,029                  | 4,778       | 5,281       |
| 15 July 2020 | 6,767                  | 6,429       | 7,106       |
| 16 July 2020 | 6,988                  | 6,639       | 7,338       |
| 17 July 2020 | 10,082                 | 9,578       | 10,586      |
| 18 July 2020 | 8,803                  | 8,362       | 9,243       |
| 19 July 2020 | 10,382                 | 9,863       | 10,901      |
| 20 July 2020 | 7,329                  | 6,962       | 7,695       |
| 21 July 2020 | 5,239                  | 4,977       | 5,501       |
| 22 July 2020 | 9,140                  | 8,683       | 9,597       |
| 23 July 2020 | 8,492                  | 8,067       | 8,917       |
| 24 July 2020 | 9,689                  | 9,205       | 10,174      |
| 25 July 2020 | 7,155                  | 6,798       | 7,513       |
| 26 July 2020 | 10,504                 | 9,979       | 11,029      |
| 27 July 2020 | 6,216                  | 5,905       | 6,527       |
| 28 July 2020 | 8,482                  | 8,058       | 8,906       |
| 29 July 2020 | 10,841                 | 10,299      | 11,383      |
| 30 July 2020 | 9,979                  | 9,480       | 10,478      |
| 31 July 2020 | 13,573                 | 12,894      | 14,251      |
| 01 August 2020 | 1,055              | 1,003       | 1,108       |
| 02 August 2020 | 7,073              | 6,719       | 7,426       |
| 03 August 2020 | 4,222              | 4,011       | 4,434       |
| 04 August 2020 | −33                 | −31         | −34         |
| 05 August 2020 | 5,614              | 5,333       | 5,894       |
| 06 August 2020 | 7,248              | 6,886       | 7,611       |
| 07 August 2020 | 7,122              | 6,766       | 7,478       |
| 08 August 2020 | 5,878              | 5,584       | 6,172       |
| 09 August 2020 | 3,772              | 3,583       | 3,960       |
| 10 August 2020 | 4,859              | 4,616       | 5,102       |
| 11 August 2020 | 620               | 589         | 651         |
| 12 August 2020 | 5,888              | 5,594       | 6,183       |
| 13 August 2020 | 4,853              | 4,611       | 5,096       |
| 14 August 2020 | 3,869              | 3,675       | 4,062       |
| 15 August 2020 | 5,944              | 5,647       | 6,242       |
| 16 August 2020 | −496             | −471        | −520        |
| 17 August 2020 | −2,245            | −2,133      | −2,357      |
| 18 August 2020 | 2,031             | 1,930       | 2,133       |
| 19 August 2020 | 5,996             | 5,696       | 6,296       |
| 20 August 2020 | 3,432             | 3,260       | 3,603       |
| 21 August 2020 | 3,228             | 3,067       | 3,389       |
| 22 August 2020 | 6,297             | 5,982       | 6,612       |
| 23 August 2020 | 1,890             | 1,796       | 1,985       |
| 24 August 2020 | −3,913            | −3,717      | −4,108      |

(Table 4 continued)
Date Predicted Active Cases Lower Limit Upper Limit
25 August 2020 1,525 1,449 1,601
26 August 2020 11,662 11,079 12,245
27 August 2020 9,769 9,280 10,257
28 August 2020 6,338 6,021 6,655
29 August 2020 7,848 7,456 8,241
30 August 2020 10,162 9,654 10,670
31 August 2020 2,451 2,328 2,573
01 September 2020 9,318 8,852 9,784
02 September 2020 8,691 8,256 9,125
03 September 2020 9,503 9,027 9,978
04 September 2020 9,311 8,846 9,777
05 September 2020 9,711 9,226 10,197
06 September 2020 12,333 11,716 12,949
07 September 2020 704 669 740
08 September 2020 8,355 7,937 8,773
09 September 2020 13,191 12,532 13,851
10 September 2020 14,924 14,178 15,670
11 September 2020 9,053 8,600 9,506
12 September 2020 9,067 8,613 9,520
13 September 2020 8,192 7,782 8,601
14 September 2020 2,114 2,008 2,219
15 September 2020 3,584 3,405 3,763
16 September 2020 8,573 8,144 9,001
17 September 2020 4,749 4,511 4,986
18 September 2020 −2,314 −2,198 −2,430
19 September 2020 −1,917 −1,822 −2,013
20 September 2020 −4,596 −4,366 −4,825
21 September 2020 −16,758 −15,920 −17,596
22 September 2020 −4,571 −4,343 −4,800
23 September 2020 −1,219 −1,158 −1,280
24 September 2020 2,281 2,167 2,395
25 September 2020 −5,589 −5,309 −5,868
26 September 2020 −2,791 −2,651 −2,930
27 September 2020 3,812 3,621 4,003
28 September 2020 −9,206 −8,746 −9,666

Source: The authors.
is new and has the capability to be transmitted intensely. Hence, it may have an influence on the predictions, however as per our knowledge, in the present situation this model is the finest.

School Reopening and Variation in COVID-19 Spread

In the second part of our study, we have estimated the probable number of active cases per day as a result of opening of schools in addition to the daily active cases. Our study is based on the supposition that children are as effective as adults in transmitting the disease. Thus both can spread the infection at a same rate if schools are also opened along with the workplaces. This has been done to avoid any underestimation since the role of children is not yet exactly clear. A look at the Table 4 suggests that reopening of schools will certainly add to the active number of COVID-19 cases per day, as expected, although the number of transmissions would not be very high. These results have been obtained assuming maximum possible transmission rate similar to adults. However, practically the number of active cases due to school going children will be less. Since the children in their early years require a lot of supervision, it is important to commence their learning and development by reopening schools (United Nations International Children Emergency Fund [UNICEF], 2020). Schools may be gradually opened based on either partial and/or differentiated approach.

In the partial reopening, a maximum of 50% students in every class can return which will be followed by the return of other 50% in an alternate day pattern. Differentiated approach includes return of particular sections of the school in different sets. In the initial phase, it is suggested that both approaches should be combined and pre-schools should be started with 50% strength from October. As per the prevailing scenario, this approach can then be extended to primary followed by middle school. The last to be opened should be high school due to their higher infection spread ability than the others. WHO has released a report on the preventive measures and controls that must be adopted in schools (WHO, 2020b) and these must be kept in mind while preparing for the openings. Before entering the premises, facilities must be there for rigorous tracing, testing and isolation of cases at a large-scale. Social distancing must be followed as strictly as possible. Masks must be made compulsory for middle and high school students. Sanitization of buildings must be carried out alternately or at least once in a week as appropriate. There is an urgent requirement to further perform virological and epidemiological examinations so that the role of children in the communication of the disease can be minutely understood among different age groups in schools and in the community.

Conclusion

The novel coronavirus disease (COVID-19) has been declared as pandemic by WHO and is currently a major global threat. In order to support the prevention and control of the disease, we have conducted this study to examine the finest model for the prediction of confirmed COVID-19 infection cases and to employ that model for forecasting future COVID-19 infection cases in India for the coming month. As per the model forecast, the confirmed cases will keep rising in the coming months. The time series analysis shows an exponential enhancement in the infected cases. It is anticipated that the efforts made by government in order to deal with the situation, effective functioning of healthcare systems and undertaking rigorous research would contribute in declining the spread. A comparative study with highly infected countries and countries in South-East Asia region indicates that India may take some time but will be able to control the situation if prevention measures such as city sanitization, house-to-house testing, usage of preventive gears and social distancing are strictly followed. An analysis based on the effect of resuming schools on the spread of COVID-19 has also been performed. It is predicted that the opening of schools can contribute towards the transmission of the spread although the increase would be minor, especially when only pre-schools and primary schools are opened. These prediction models and estimations are believed to help the government in their decision making process and medical workforce to be prepared for the upcoming situations and have more readiness in healthcare systems.

Acknowledgements

Dr. Tanmoy Chakraborty is thankful to Sharda University and Dr. Hiteshi Tandon is thankful to Manipal University Jaipur for providing research facility.

Author Contributions

Hiteshi Tandon: Conceptualization, methodology, software, formal analysis, validation, investigation, writing–original draft, visualization.

Prabhat Ranjan: Resources.

Tanmoy Chakraborty: Conceptualization, supervision, writing–review and editing.

Vandana Suhag: Supervision, writing–review and editing.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.
Sujath, R., Chatterjee, J. M., & Hassanien, A. E. (2020). A machine learning forecasting model for COVID-19 pandemic in India. Stochastic Environmental Research and Risk Assessment, 34, 959–972.

Team, N. C. P. E. R. E. (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) – China, 2020. China CDC Weekly, 41, 145.

Tiwari, S., Kumar, S., & Guleria, K. (2020). Outbreak trends of Corona Virus (COVID-19) in India: A prediction. Disaster Medicine and Public Health Preparedness, 14(5), e33–e38. https://pubmed.ncbi.nlm.nih.gov/32317044/

Tomar, A., & Gupta, N. (2020). Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. Science of the Total Environment, 728, 138762.

United Nations International Children Emergency Fund. (2020). Framework for reopening schools. https://www.unicef.org/media/68366/file/Framework-for-reopening-schools-2020.pdf

Unnithan, P. S. G. (2020). Kerala confirmed first novel coronavirus case in India. https://www.indiatoday.in/india/story/kerala-reports-first-confirmed-novel-coronavirus-case-in-india-1641593-2020-01-30

Verma, M. K., Asad, A., & Chatterjee, S. (2020). COVID-19 pandemic: Power law spread and flattening of the curve. Transactions of the Indian National Academy of Engineering, 5, 103–108.

Wang, C. J., Ng, C. Y., & Brook, R. H. (2020). Response to COVID-19 in Taiwan: Big data analytics, new technology, and proactive testing. JAMA, 323, 1341–1342.

World Health Organization (WHO). (2020a). Coronavirus disease (COVID-19) outbreak. https://www.who.int/emergencies/diseases/novel-coronavirus-2019

World Health Organization (WHO). (2020b). Key messages and actions for COVID-19 prevention and control in schools. https://www.who.int/docs/default-source/coronaviruse/key-messages-and-actions-for-covid-19-prevention-and-control-in-schools-march-2020.pdf?sfvrsn=baf81d52_4

Zhang, Y., Yang, H., Cui, H., & Chen, Q. (2019). Comparison of the ability of ARIMA, WNN and SVM models for drought forecasting in the Sanjiang Plain, China. Natural Resources Research, 29, 1447–1464.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., & Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. New England Journal of Medicine, 382, 727–733.