

Article

Data Analytics and Mathematical Modeling for Simulating the Dynamics of COVID-19 Epidemic—A Case Study of India

Himanshu Gupta ¹, Saurav Kumar ¹, Drishti Yadav ¹, Om Prakash Verma ¹, Tarun Kumar Sharma ²,
Chang Wook Ahn ^{3,*} and Jong-Hyun Lee ^{4,*}

¹ Department of Instrumentation and Control Engineering, Dr B R Ambedkar National Institute of Technology Jalandhar, Jalandhar 144011, India; guptah.nitj@gmail.com (H.G.); srvnitj@gmail.com (S.K.); drish131196@gmail.com (D.Y.); vermaop@nitj.ac.in (O.P.V.)

² Department of Computer Science and Engineering, Shobhit University Gangoh, Saharanpur 247341, India; taruniitr1@gmail.com

³ AI Graduate School, Gwangju Institute of Science and Technology, Gwangju 61005, Korea

⁴ Research Center for Convergence, Sungkyunkwan University, Suwon 16419, Korea

* Correspondence: cwan@gist.ac.kr (C.W.A.); ljh08375@skku.edu (J.-H.L.)

Abstract: The global explosion of the COVID-19 pandemic has created worldwide unprecedented health and economic challenges which stimulated one of the biggest annual migrations globally. In the Indian context, even after proactive decisions taken by the Government, the continual growth of COVID-19 raises questions regarding its extent and severity. The present work utilizes the susceptible-infected-recovered-death (SIRD) compartment model for parameter estimation and fruitful prediction of COVID-19. Further, various optimization techniques such as particle swarm optimization (PSO), gradient (G), pattern search (PS) and their hybrid are employed to solve the considered model. The simulation study endorse the efficiency of PSO (with or without G) and G+PS+G over other techniques for ongoing pandemic assessment. The key parametric values including characteristic time of infection and death and reproduction number have been estimated as 60 days, 67 days and 4.78 respectively by utilizing the optimum results. The model assessed that India has passed its peak duration of COVID-19 with more than 81% recovery and only a 1.59% death rate. The short duration analysis (15 days) of obtained results against reported data validates the effectiveness of the developed models for ongoing pandemic assessment.

Keywords: COVID-19; compartment modeling; epidemiology; predictive modeling; optimization; particle swarm optimization



Citation: Gupta, H.; Kumar, S.; Yadav, D.; Verma, O.P.; Sharma, T.K.; Ahn, C.W.; Lee, J.-H. Data Analytics and Mathematical Modeling for Simulating the Dynamics of COVID-19 Epidemic—A Case Study of India. *Electronics* **2021**, *10*, 127. <https://doi.org/10.3390/electronics10020127>

Received: 5 December 2020

Accepted: 4 January 2021

Published: 8 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The novel coronavirus disease 2019 (COVID-19) has created one of the biggest challenges in the history of mankind. With over 63.89 million positive cases and 1.48 million deaths as of 1 December 2020 [1], the COVID-19 continues to its deadly killing spree in all corners of the world. The disease caused by SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) is primarily spread between people during close contact mainly via small droplets produced by coughing, sneezing and talking [2,3]. It is the third virus of corona family and other two are SARS-CoV (Severe Acute Respiratory Syndrome Coronavirus) and MERS-CoV (Middle East Respiratory Syndrome Coronavirus) [4]. Compared to the other two viruses, SARS-Cov-2 not only displays unusual epidemiological traits but the number of deaths associated with it also exceeds greatly. This poses a colossal threat to the global public health and economy [5].

The first case was reported in late December 2019 in Wuhan, Hubei, China and after that despite the radical measures taken by various countries to contain it; the number of COVID-19 patients grew exponentially across the globe. As a result of this, World Health Organization (WHO) declared this pneumonia a pandemic on 11 March 2020, a first of its

kind since World War II [6,7]. The symptoms of COVID-19 are around 80% mild (which include fever, cough and shortness of breath with the mean incubation period of 5–6 days), 13.8% severe and 6.2% critical (which include respiratory failure, septic shock, and/or multiple organ failure) [8]. The median time for mild and severe/critical patients from the onset to clinical recovery is approximately 2 weeks and 3–6 weeks respectively [8]. Further, it has been found that the deadly effect of COVID-19 is also associated with demographic conditions such as population, population density, age structure and so forth and pre-existing medical condition of an individual [9,10].

The COVID-19 has triggered one of the largest annual migrations in the world which resulted in a rapid global spread of the virus. Because of this, the epidemic center was shifted to Europe and now to USA which has already more than 117,834 deaths by 14 June 2020 [1]. India has also felt the impact of this global pandemic with the first reported case on 30 January 2020 which gradually increased to 519 positive cases and 40 deaths by 24 March 2020 [11]. In the absence of any vaccine, Indian policymakers, like other countries, have implemented a nation-wide lockdown which after some relaxations still continues as of the submission date of this study on 3 December 2020. The situation in overpopulated countries like India is very threatening as even after 60 days of lockdown exponential growth in the number of infected people is observed. Currently, the ever-infected people in India have already crossed 9.48 million. Moreover, with significantly increased testing rate per million population (which is still much less as compared to others), more number of cases are expected to come. The rate of coronavirus tests performed per million in most impacted countries as of 26 November 2020 is shown in Figure 1. As impact of health on severe economic measures are well known, policymakers in India, like others, are facing a difficult situation when trying to balance between draconian public health actions and keeping the economy alive [12].

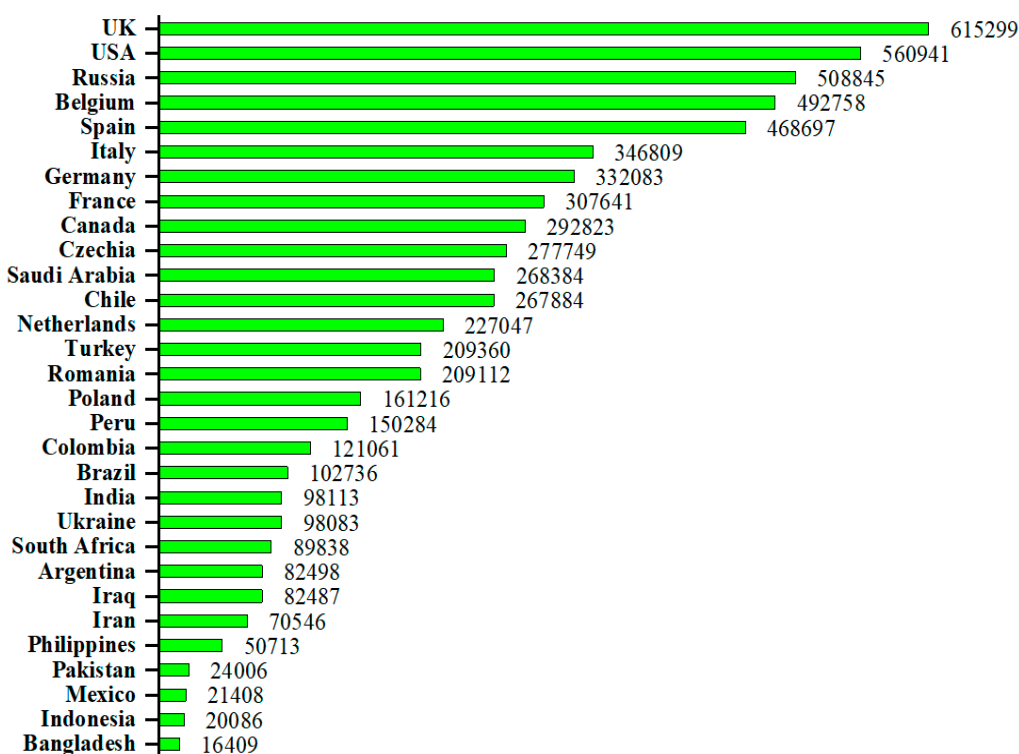


Figure 1. Rate of coronavirus testing performed (per million) in the most impacted countries as of 26 November 2020 [13].

Due to the lack of previous exposure and still unknown behavior of the novel coronavirus, future prediction of its spread and acquired herd-immunity cannot be properly anticipated. Nevertheless, the epidemiological models could help in understanding the

dynamics of COVID-19. Moreover, this will help the authorities in optimum resource sharing including medical infrastructure and administrative service [14]. A number of heuristic models have been proposed to predict COVID-19 outbreak which may be broadly categorized into time series and compartment models. Long term prediction, especially for the current pandemic using time series models (which are mainly based on exponential [14] or logistic [15] curve fitting and are used to describe infection data using scarce (or none) physical meaning parameters), should be avoided. On the other hand, compartment models provide right trade-off between the ease of solution, need of having physical-based parameters and the limited available data of the ongoing epidemic [16,17]. The study of spread of infectious disease using compartment model started in 1927 when Kermack–McKendrick gave their theory [18], leading to the introduction of SIR (Susceptible–Infected–Recovered) model and their successors. In this, the total population of interest, based on the infection status, is subdivided into small compartments (susceptible, exposed, infected, recovered etc.) and the flow between these compartments is governed by ordinary differential equations.

The present investigation utilizes the SIRD model, one of the basic compartment models, to evaluate and forecast the COVID-19 outbreak in India. In order to incorporate some of the proactive actions taken by India, various parameters including dynamic behavior of infection, recovery and death rate have been considered. Although the key parametric values required for modeling are published by WHO from time to time but they represent the global average. Therefore, in this study, these parameters have been estimated using various optimization techniques. Based on the optimum fitting to Indian outbreak, best parametric values have been chosen which are further utilized by the model for predicting the COVID-19 pandemic in India.

The key contributions of this paper are highlighted below:

- An efficient forecasting model to forecast the infected, recovered and death cases of the COVID-19 in India based on available epidemiological data.
- A maiden attempt to evaluate the effectiveness of PSO, G and PS with their hybrid combinations for prediction in Indian context, to the best knowledge and belief of authors.
- Estimation of key parametric values for SIRD modeling in context of India rather than utilizing the median values published by WHO.

The rest of this paper is organized as follows. Section 2 describes the basics of model used in the present study and Section 3 explains the evolutionary kinematic of COVID-19. The stability analysis has been discussed in Section 4. Thereafter, Section 5 enlightens about epidemiological data and its source. In Section 6, the simulation setup and various optimization techniques used have been presented. Section 7 presents a widespread discussion of the findings and at last, the concluding remarks are presented in Section 8.

2. The SIRD Model

The compartment models are based on the assumption that every individual in the compartment must exhibit same characteristics. The main reason of choosing SIRD model in the present analysis is its simplicity and easy implementation among other compartment models along with high robustness in elucidating the evolution of the pandemic. In SIRD model, the total population of interest is subdivided into Susceptible (S) (number of people that might be infected), Infected (I) (number of people already infected), Recovered (R) (number of people that have been recovered) and Death (D) (number of deaths occurred). The schematic of SIRD model has been illustrated in Figure 2.

The model has been developed with constant population ($N = 1.35$ billion). The birth rate and death rate are assumed to be same for the period of interest; hence vital dynamics have no impact on total population. Although circumstantial cases are found in the literature for re-infected people but the reinfection rate appears negligible [19,20]. Hence, the probability rate of re-susceptible once recovered from the infection has not been consid-

ered. The ordinary differential equations describing the evolution of population in each compartment over time for the present SIRD model [6] are reported in Equations (1)–(4).

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \quad (1)$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \alpha I - \mu I \quad (2)$$

$$\frac{dR}{dt} = \alpha I \quad (3)$$

$$\frac{dD}{dt} = \mu I \quad (4)$$

where, β represents the rate of transmission (infection), α represents the rate of recovery and μ represents the death rate. S is the number of people susceptible to disease. I represents the number of people ever infected. R is the number of people who recovered and D represents the number of people who died. The total population under consideration is represented by N .

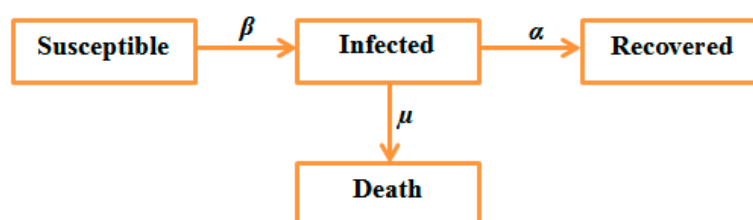


Figure 2. Schematic of SIRD model.

A susceptible person becomes infected by coming into ample contact with an infectious person and the probability of being infected is directly proportional to the product of fraction of infected (I/N) and susceptible population. A fraction of infected people can recover whose recovery is directly proportional to the recovery rate and inversely proportional to the average duration of infection as illustrated in Equation (3). Moreover, a fraction of critically infected population may die which is directly proportional to the case fatality rate or lethality of disease and inversely proportional to the average duration between the infection and death as shown in Equation (4). Equations (1)–(4) are not independent because of the consideration of a closed population for diseases with nation-wide outbreak and low lethality rate (like COVID-19) and therefore, they all must follow Equation (5) at any stage.

$$N = S + I + R + D. \quad (5)$$

Hence, values from any of these five equations can be determined using the other four and their initial values are taken from the available data [11].

3. The Evolutionary Kinematic of COVID-19

Most of the available simple compartment models (SIR [21], SIRD [6] and SEIR [22], etc.) are based on the constant kinematics of β , α and μ which are explained in Equations (1)–(4). Based on these factors, they describe the epidemic growth until a very large population is being infected and achieve herd immunity [23]. However, due to the momentous influence of various suppression strategies (such as social distancing and lockdown) taken by policymakers and new findings on the epidemic, an accurate forecasting based on the growth of its kinematics becomes cumbersome.

3.1. The Infection Rate (β)

The infection or transmission rate (β) is generally used as a fitting parameter to describe the epidemic. However, on account of various aggressive mitigation and suppression

strategies taken by government for the current epidemic, the number of adequate contacts per person per unit time decreases drastically. Hence, in this work, to capture the behavioral changes by mitigation and suppression steps, a time varying β as described in Equation (6) is considered.

$$\beta(t) = \begin{cases} \beta_0, & t < t_{\text{lockdown}} \\ \beta_0 \left(e^{-\frac{t-t_{\text{lockdown}}}{\tau_\beta}} \right)^2 + \beta_1, & t \geq t_{\text{lockdown}} \end{cases} \quad (6)$$

As per the previous findings, β depends upon three characteristic values β_0 , β_1 and τ_β [24]. There is an initial high value β_0 and a final value β_1 , which for all practical purposes is assumed to be zero considering a long enough isolation period. After lockdown, β decays exponentially from β_0 to β_1 . Finally, τ_β represents the settling time and it has been assumed that at $t = 3\tau_\beta$, transmission rate is decreased to 90% of its initial value.

3.2. The Recovery Rate (α)

For COVID-19 being a new disease with inadequate known characteristics, the recovery rate may not be constant. It mainly depends upon the policies adopted by various countries, existing health care system, new clinical findings about the disease, ability of doctors and medical staff to quickly learn new therapeutic and unorthodox procedure for detection and treatment and so forth. Hence, the recovery time gradually decreases from its initial high value to a constant value which is inversely proportional to the mean infection period. In this work, the recovery rate (α) has been modeled with a logistic function [25] which is expressed by Equation (7).

$$\alpha(t) = \alpha_0 + \frac{\alpha_1}{1 + e^{(-t+\tau_\alpha)}}, \quad (7)$$

where, α_0 represents the initial recovery rate and $\alpha_0 + \alpha_1$ represents the final recovery rate after τ_α duration.

3.3. The Death Rate (μ)

Like any other infectious disease, the death rate of any new disease cannot be constant. Usually, it is initially high due to lack of awareness about the disease and hence, only severe cases are detected. However, with increasing awareness and introduction of dedicated treatment methods, the death rate decreases gradually and finally, settles down to a long-term mortality rate. Moreover, with the introduction of various non-pharmaceutical interventions (like social distancing and lockdown), the number of new infected reduces drastically which in turn reduces death rate. Based on this background, a time varying death rate as described by Equation (8) is considered in the present analysis.

$$\mu(t) = \begin{cases} \mu_0, & t < t_{\text{lockdown}} \\ \mu_0 \left(e^{-\frac{t-t_{\text{lockdown}}}{\tau_\mu}} \right) + \mu_1, & t \geq t_{\text{lockdown}} \end{cases} \quad (8)$$

where, μ_0 and μ_1 represent the initial and final death rate respectively. Here, it is assumed that after τ_μ duration, μ reduces by 90% of its initial value.

3.4. The Reproduction Number (R)

The reproduction number, R_0 is a very crucial parameter in any epidemic. Most of the statistical models of COVID-19 are based on R_0 . It basically explains the number of cases each infected case can directly generate considering all individuals susceptible to infection [26]. It is simply a threshold which is used to determine the nature of any disease being an epidemic ($R_0 > 1$) or not ($R_0 < 1$). Generally, the higher the value of R_0 ,

the tougher it is to control the epidemic. For basic compartment model (SIR), R_0 calculated by Equation (9).

$$R_0 = \frac{\text{typical time until removal}}{\text{typical time in contacts}} = \frac{\beta}{\alpha}. \quad (9)$$

The normalized infected population [22] may be calculated by Equation (10).

$$i = \frac{I}{N}, \quad (10)$$

where, i represent the normalized infected population. R_0 for SIRD model can be calculated from Equation (11) which is obtained by putting Equation (10) in Equation (2).

$$\frac{di}{dt} = \frac{\beta Si}{N} - \alpha i - \mu i. \quad (11)$$

On putting Equation (11) equal to zero Equation (12) is formed.

$$\begin{aligned} \frac{di}{dt} &= 0 \\ \Rightarrow \left(\frac{\beta S}{N} - \alpha - \mu \right) i &= 0. \end{aligned} \quad (12)$$

Since, normalized infected population cannot be zero and therefore, Equation (12) may be modified as Equation (13).

$$\frac{\beta S}{(\alpha + \mu)N} = 1. \quad (13)$$

Further, R_0 may be defined by Equation (14).

$$R_0 = \frac{\beta S}{(\alpha + \mu)N}. \quad (14)$$

Moreover, at the initial phase of outbreak, everyone is susceptible ($S \approx N$). Because of very large population with reasonably low total infected, $S \approx N$ holds true at any time, therefore, Equation (13) is equivalent to Equation (15).

$$R_0 = \frac{\beta}{(\alpha + \mu)}. \quad (15)$$

Conclusively, if β is greater than $\alpha + \mu$, then the disease is epidemic; else it will die out. Since, it is usually a mathematical parameter with no physical meaning, it alone cannot describe the true nature of any disease [27].

4. Stability Analysis

To analyze the stability of the disease-free equilibrium (DFE) in terms of the reproduction number, let it is defined by Equation (16).

$$E_0 = (N, 0). \quad (16)$$

Moreover, Equations (1) and (2) yields that $0 \leq S$, $0 \leq I$ and $S + I \leq N$ and the set $\Omega = \{(S, I): S \geq 0; I \geq 0; S + I \leq N\}$ is a positively invariant compact set for the model. On considering the Lyapunov-LaSalle function as $V(S, I) = I$, the globally asymptotically stability has been analyzed using Equation (17).

$$\dot{V} = I = \frac{\beta SI}{N} - (\alpha + \mu)I = I(R_0 - 1)(\alpha + \mu) \leq 0. \quad (17)$$

Furthermore $\dot{V} = 0$ if $I = 0$ or $R_0 = 1$. Therefore, the largest invariant set contained in the set represented by Equation (18) is reduced to DFE and is globally asymptotically stable in Ω .

$$L = \{(S, I) \in \Omega / \dot{V}(S, I) = 0\} \quad (18)$$

5. The Epidemiological Data

The data utilized in the present investigation has been taken from the Ministry of Health and Family Welfare (MOHFW), Government of India [11] and publicly available dashboard of India Today [28]. Although, as stated earlier, the very first case was found on 30 January 2020 but until 2 March 2020, all the infected cases were recovered and no new cases were reported. So, for this analysis, data from 3 March 2020 to 28 November 2020 has been considered. In the early phase of outbreak, all the positive cases are because of the overseas travelers and in response to that, India implemented travel ban on 12 March 2020. Although, some people have been migrated after this date but they are properly tested and quarantined. Moreover, considering high internal mobility due to festive season and very limited international migrations, the whole nation has been considered as a closed population in this analysis.

The data available in the aforementioned source provide details about the total active, cured/discharged, deaths and migrated. The total cumulative count has been estimated by summing all the available data and the number of incident cases (new daily cases) have been estimated by the difference of current and previous day cumulative count. For better visualization of data, the variations in cumulative and daily cases with respective dates have been illustrated in Figure 3a–c for infected, recovered and death respectively. The number of cumulative cases and new daily cases are represented on left side and right side respectively in each figure.

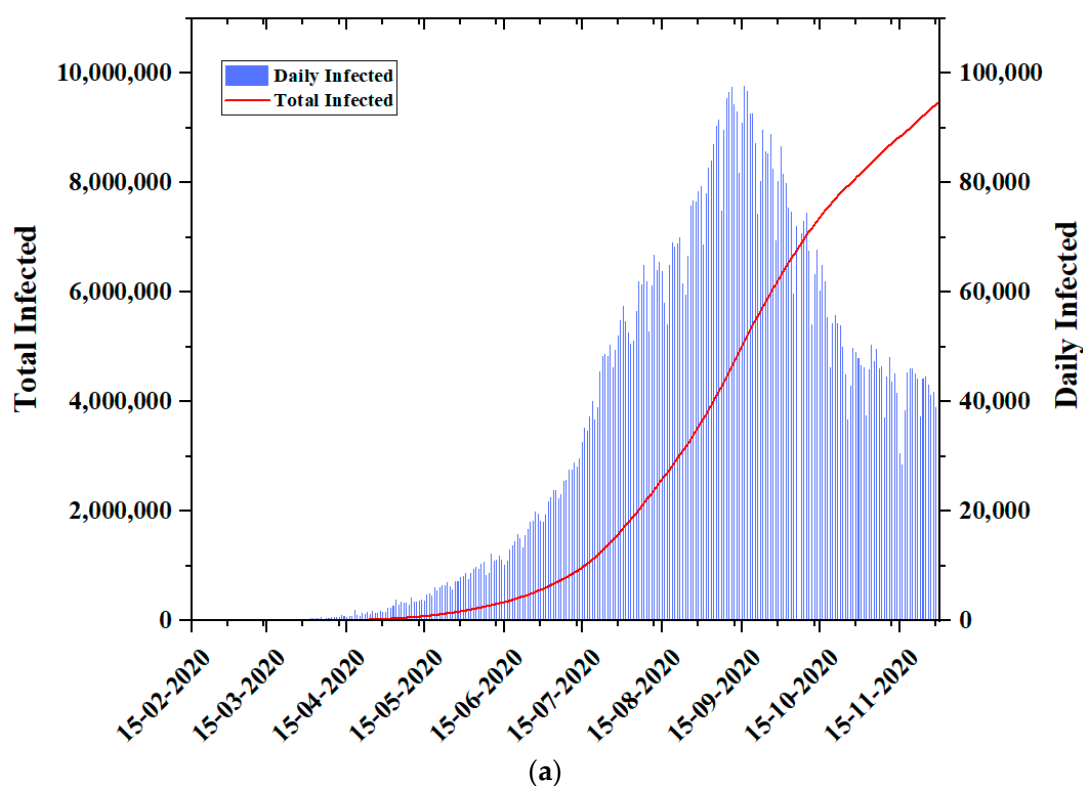


Figure 3. Cont.

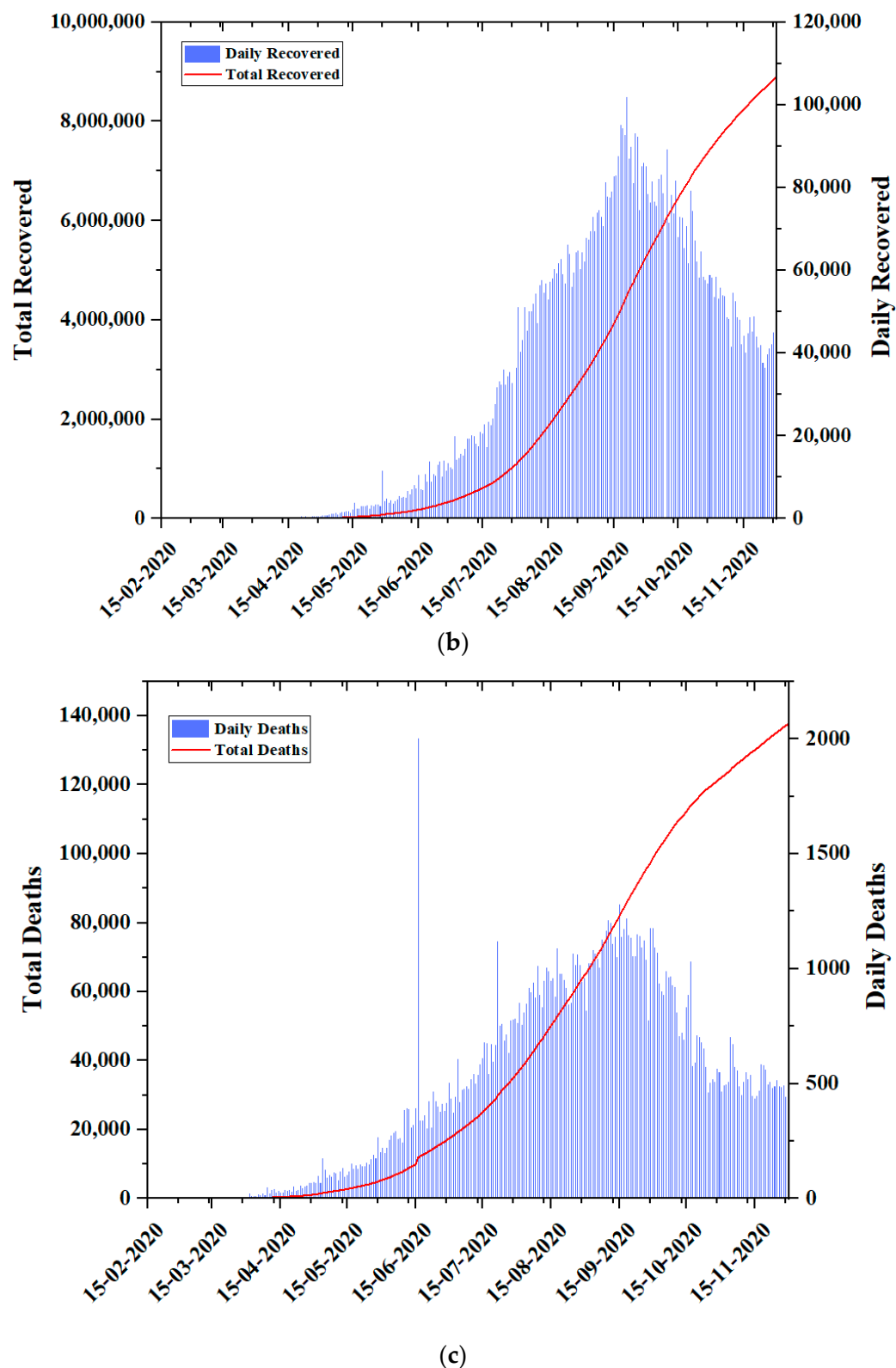


Figure 3. The variations in cumulative and daily cases with number of cumulative cases and new daily cases are represented on left side and right side respectively for (a) Infected (b) Recovered (c) Deaths.

Moreover, the population density of India is 464.1 per square kilometer which is amongst top 30 in the world and the total population of India is more than 1.38 billion [29] (second largest on the globe) out of which more than 31% is urban population [30].

The urban population resides in very dense clusters which creates a significant challenge for authorities to apply strict social distancing norms. To have a better visualization of COVID-19 spread, a surface plot for the top 10 most affected states in India up until 30 November 2020 has been sketched in Figure 4. It has been observed that the spread is asymmetrically distributed and these states carry about 75% cases out of total cases found in India, Maharashtra being the most affected. These states carry maximum proportion of Indian population and have a significant number of economic hubs in India which points towards their high population density than the average value. It validates that people in these states are compelled to live in highly dense clusters and containment of spread by social distancing norms is very difficult. However, the effect of population density and other such parameters towards the suppression of COVID-19 is beyond the scope of this study.

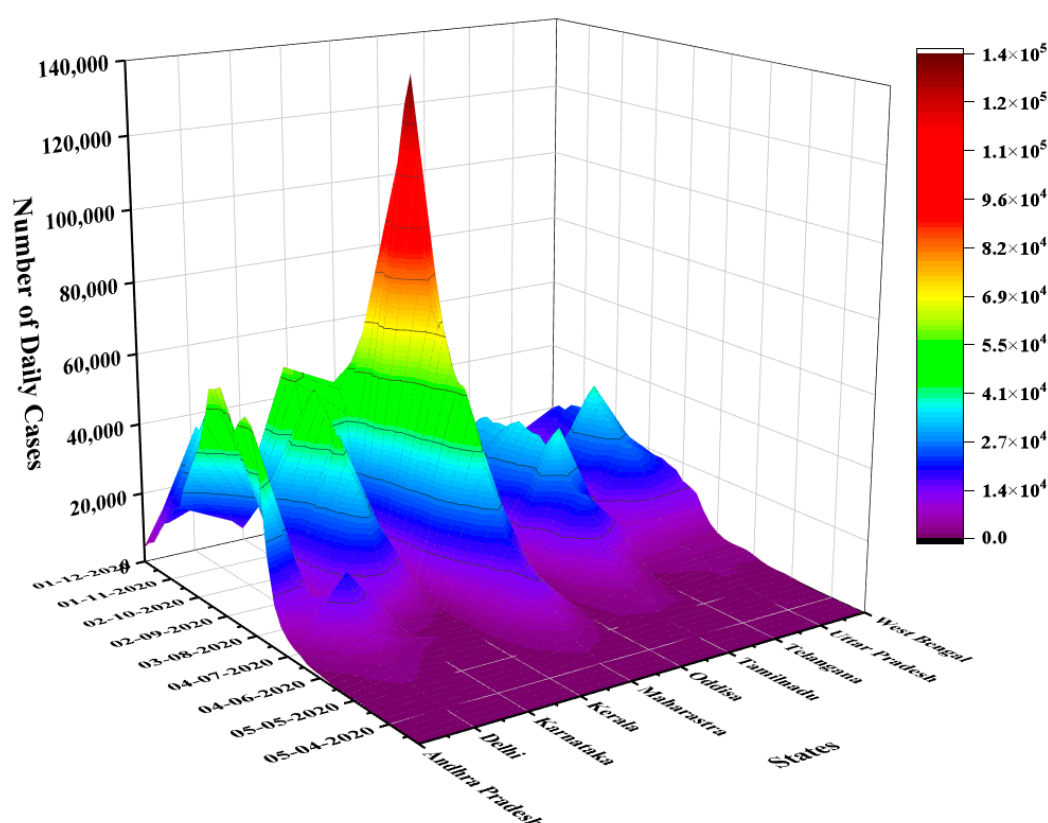


Figure 4. COVID-19 top 10 most affected states in India.

6. Model Simulation and Optimization

In the present investigation, the entire simulation work has been done on MATLAB 2016a programming platform. The optimization of model parameters has been implemented by the minimization of residuals (objective function) estimated between available epidemiologic data and modeling data with default settings. Therefore, if x_i and \hat{x}_i represents the modeling and experimental data at a certain time i respectively, then the residuals will be calculated as $J_i = x_i - \hat{x}_i$. Further, because cumulative data is highly correlated which may produce highly misleading results. Therefore, in this work the daily incident data of infected, recovered and deaths have been utilized. However, the incident data is not only independent but also greatly scattered as a result of which the ordinary least squares may generate inappropriate result. To overcome this problem, the bi-square M-estimator

based regression method has been implemented which recalculates the residuals at a time according to Equation (18) [24].

$$\mathfrak{J}(J_i) = \begin{cases} \frac{k^2}{6} \left[1 - \left\{ 1 - \left(\frac{J_i}{k} \right)^2 \right\}^3 \right] & \text{for } |J_i| \leq k, \\ \frac{k^2}{6} & \text{for } |J_i| > k \end{cases} \quad (19)$$

where, $k = 5.38 \times \text{Mean absolute deviation of the values in the residuals}$. Therefore, the impact of outliers ($|J_i| > k$) has been nullified. Further, in the present investigation three independent data sets (daily new infected, daily new recovered and daily new deaths) have been incorporated which converts the problem in a multi-objective function minimization. The weighted sum of single objective function $\mathfrak{J}(J_i)$ has been employed as the final objective function to estimate the goodness of fit between the model predicted and epidemiological data as described in Equation (19).

$$\text{argmin} \left\| \sum_i \left\{ \frac{k_1(\mathfrak{J}(J_i(I)))^2}{|I|} + \frac{k_2(\mathfrak{J}(J_i(R)))^2}{|R|} + \frac{k_3(\mathfrak{J}(J_i(D)))^2}{|D|} \right\} \right\|, \quad (20)$$

where, $k_1 = 20$, $k_2 = 50$ and $k_3 = 100$ have been empirically chosen to compensate the different order of magnitude between single objective functions. Further, this global objective function has been used to compute the parameters of developed model at global minima using proposed optimization techniques including pattern search (PS), particle swarm (PSO), gradient descent (G) and their hybrid. Nevertheless, the generalized step-by-step pseudo code for the proposed optimization techniques (PS, PSO, and G) has been comprehensively demonstrated in Algorithms 1–3 respectively.

Algorithm 1 (Pseudo code for PS)

```

for each step  $i = 1, \dots, S$  do,
    Initialize the default search step  $\alpha_0$ 
    Initialize the current solution  $\beta_0$ 
     $\alpha = \alpha_0$ 
    while  $i < S$  or error  $\geq$  error bound do:
        for each column  $\beta_i$  in  $B$  do
             $\theta = \{\beta_0 + \alpha \times \beta_i\}$ 
            Evaluate the nearest neighbors in  $\theta$ 
        If  $\theta_{new} > \theta_{previous}$ 
            Update the current solution to the best neighbor in  $\theta$ 
             $\alpha = \alpha_0$ 
        Else  $\alpha = \frac{\alpha_0}{2}$ 

```

Algorithm 2 (Pseudo code for PSO)

```

for each particle  $i = 1, \dots, S$  do,
    Initialize the particle's position with a uniformly distributed random vector :  $x_i \sim U(b_l, b_u)$ 
    Initialize the particle's best known position to its initial position :  $p_i \leftarrow x_i$ 
    If  $f(p_i) < f(g)$  then update the swarm's best known position :  $g \leftarrow p_i$ 
    Initialize the particle's velocity :  $v_i \sim U(\pm|b_u - b_l|)$ 
    while  $i \leq S$  or error  $\geq$  error bound do:
        for each particle  $i = 1, \dots, S$  do
            for each dimension  $d = 1, \dots, n$  do
                Pick random numbers :  $r_p, r_g \sim U(0, 1)$ 
                Update particle's velocity:  $v_{i,d} \leftarrow \rho v_{i,d} + \omega_p r_p (p_{i,d} - x_{i,d}) + \omega_g r_g (g_d - x_{i,d})$ 
                Update the particle's position:  $x_i \leftarrow x_i + v_i$ 
                If  $f(x_i) < f(p_i)$  then update the particle's best known position :  $p_i \leftarrow x_i$ 
                If  $f(p_i) < f(g)$  then update the swarm's best known position :  $g \leftarrow p_i$ 

```

Algorithm 3 (Pseudo code for G)

```

for each step  $i = 1, \dots, S$  do,
    Initialize with  $f$  being a differentiable function ( $\mathbb{R}^n \rightarrow \mathbb{R}$ )
    Initialize with any random solution  $x^0$ 
    while  $i < S$  or error  $\geq$  error bound do:
        for each  $i = 1, \dots, S$  do
             $x^{i+1} \leftarrow x^i - \alpha^i \nabla f(x)$ ;
            with  $\alpha^i = \operatorname{argmin}_{\alpha \in \mathbb{R}} f(x^i - \alpha \nabla f(x))$ 
        Update  $x^i$ 

```

7. Result and Discussion

The SIRD model presented in this study is based upon ten parameters as discussed in Section 3. Most of the reported articles on mathematical modeling for COVID-19 forecasting use some parameters which are based upon the median value given by WHO. However, due to the fact that the origin of the virus is still unknown, the use of these values may produce highly misleading parameters, resulting in wrong prediction. To avoid this pitfall, all the parameters except $t_{lockdown}$ have been estimated using the evolutionary data. In order to validate the robustness of the estimated parameters, PSO, PS and G along with their combinations have been used. It has been found that the values of these estimated parameters greatly vary depending upon the optimization technique and level of hybridization used. The parametric values estimated by G and PS are highly misleading however, PSO estimates the optimum values. Simulated results also reveal that the parameters estimated by PSO with or without G are approximately same. However, hybridization of G with PSO only increases the complexity and simulation time without any fruitful advantage and it may be avoided. On the contrary side, a drastic impact of hybridization has been observed in the estimated parameters of PS and depending upon the level of hybridization, the estimated parameter approaches towards the values estimated by PSO. With level-1 hybridization, G followed by PS (G + PS), a variation of 4.8×10^{-2} (23.76%), 12.48 (20.68%), 0.31×10^{-2} (13.54%), 0.951×10^{-2} (14.68%), 4.17 (9.56%), 5.93×10^{-2} (18.74%), 9.80×10^{-3} (136.11%) and 52.78 (78.32%) in β_0 , τ_β , α_0 , α_1 , τ_α , μ_0 , μ_1 and τ_μ respectively with respect to PSO has been observed. However, with level-2 hybridization, G followed by PS followed by G (G + PS + G), almost same values of parameters as of PSO have been observed with a variation of 1×10^{-4} (0.05%), 0.03 (0.05%), 0, 1×10^{-3} (1.54%), 0.65 (1.49%), 1.9×10^{-4} (5.99%), 1.5×10^{-3} (20.83%) and 3.78 (5.61%) in the respective values of β_0 , τ_β , α_0 , α_1 , τ_α , μ_0 , μ_1 and τ_μ . Therefore, with level-2 hybridization, the parametric values of PS have been greatly improved. The various estimated parameters and their bounds have been presented in Table 1. Also, the outbreak of COVID-19 in India can be viewed as from regional to national level and in response to this, various states applied lockdown and curfew. However, they are limited to a particular region only whereas, the nationwide lockdown was implemented on 25 March 2020. Hence, in this study, $t_{lockdown}$ has been considered as 22 days.

According to the parameters estimated in Table 1, the model values of infected, recovered and deaths have been generated using different optimization techniques and to find the best fit, it has been plotted in Figure 5a–c respectively. It has been found that G and PS based optimization alone produces worst fit to the available epidemiological data whereas, G + PS yield slightly better fitting; however, it is also unable to deliver optimum fit. This may be because of the reason that instead of finding global minima, they got stuck to local minima. However, when the objective function was minimized by PSO (with or without G) or G + PS + G, then global minima has been achieved and optimum fitting to the available data has been obtained. Moreover, although with level-2 hybridization, PS achieves global minima but it also enhances system complexity and simulation time. However, PSO alone can lead towards global minima as evident from Figure 5. Therefore,

it has been assumed that the parameters estimated by PSO are the parameters of the Indian epidemiological data.

Table 1. Estimated model parameters.

Parameter	Range	Optimization Technique						
		G	PS	PSO	G + PS	G + PSO	G + PS + G	G + PSO + G
β_0 (per day)	[0, 1]	3.55×10^{-9}	1.68×10^{-9}	0.202	0.25	0.2019	0.2021	0.2021
τ_β (day)	[10, 100]	10.00	10.00	60.34	47.86	60.38	60.31	60.34
α_0 (per day)	[0.01, 0.1]	0.01	0.01	0.0229	0.0198	0.023	0.0229	0.0227
α_1 (per day)	[0.001, 0.1]	0.09889	0.09759	0.06479	0.0743	0.06497	0.06577	0.06451
τ_α (day)	[5, 80]	78.69	65.43	43.63	39.46	43.69	42.98	43.62
μ_0 (per day)	[0, 0.5]	0.417	0.417	3.17×10^{-3}	6.26×10^{-2}	3.19×10^{-3}	3.36×10^{-3}	3.17×10^{-3}
μ_1 (per day)	[0, 0.1]	0.07	0.06	7.2×10^{-4}	1.7×10^{-3}	7.2×10^{-4}	8.7×10^{-4}	7.5×10^{-4}
τ_μ (day)	[10, 60]	10.00	9.89	67.39	14.61	67.39	63.61	67.39

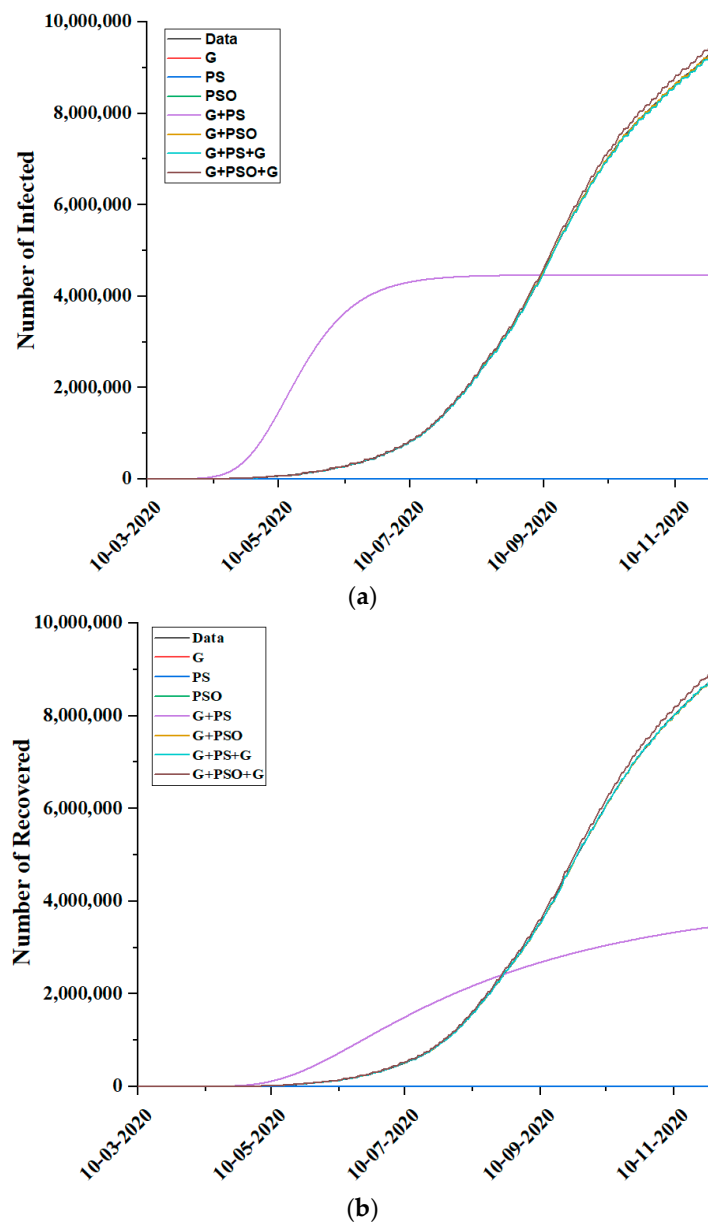


Figure 5. Cont.

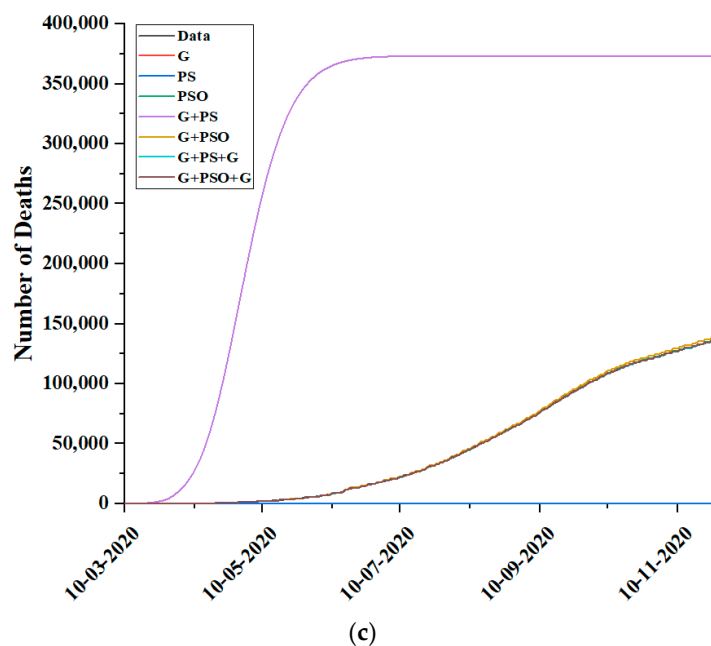


Figure 5. Comparative analysis of real time Indian epidemiologic data and fitted model by SIRD using various optimization techniques and their hybrid (a) Infected (b) Recovered (c) Deaths.

Based on the above findings, the modeling results for kinetic of infection, recovery and death rate obtained for Indian COVID-19 epidemiologic using all the above mentioned optimization techniques (G, PS, PSO, G + PS, G + PSO, G + PS + G and G + PSO + G) have been reported in Figure 6. Here, it is important to analyze the value of acquired modeling parameters, reported in Table 1 and plotted in Figure 6 to have some physical considerations. Assuming constant contact rate before lockdown, β remains constant in that duration and after that, due to isolation, it decreases exponentially with the characteristic time of τ_β equals to 60 days. Therefore, after 180 days ($3\tau_\beta$) from the onset of epidemiology, β reduces to 90%. The recovery parameters (α_0 , α_1 and τ_α) are useful to detect the average number of days required for recovery (T_r). The kinetic of recovery has been found as constant which represents only severe cases have been observed and reported. Therefore, T_r has been calculated as 44 which is within the bounds of very severe infections [8]. The parameters associated with kinetic of death (μ_0 , μ_1 and τ_μ) have been used to acquire the information regarding the average time between symptom and death. Considering the 4% death rate [31], average death time (T_d) gradually increases from 13 days until it reaches seven months from the onset of pandemic and achieves a long-term mortality rate. This can be described as at the start of pandemic; mostly severe cases have been reported. Nevertheless, in spite of weak medical infrastructure, the factors like a very large young population, warmer weather conditions and humidity [32] along with the awareness programs initiated by the government favor India in achieving long term lethality rate.

Another important parameter to analyze the severity of infectious disease is R_0 and it has been regularly calculated using the parameters estimated in Table 1 with the help of Equation (15). At the start of the outbreak, moderate initial value (3.22) has been observed that goes up to 9.78 by first week of May 2020. This is because at the beginning, it has been driven by high initial value of β and very low values of α and μ . However, on the application of social distancing norms, lockdown and other suppression strategies, it started to fall gradually and settle down to less than 1 after five months of outbreak. The average value has been calculated as 4.78 whereas in literature, the R_0 of 2.56 has been reported for India [33]. However, based on the long enough duration of disease and different methods used for calculation, the estimated data can also be considered.

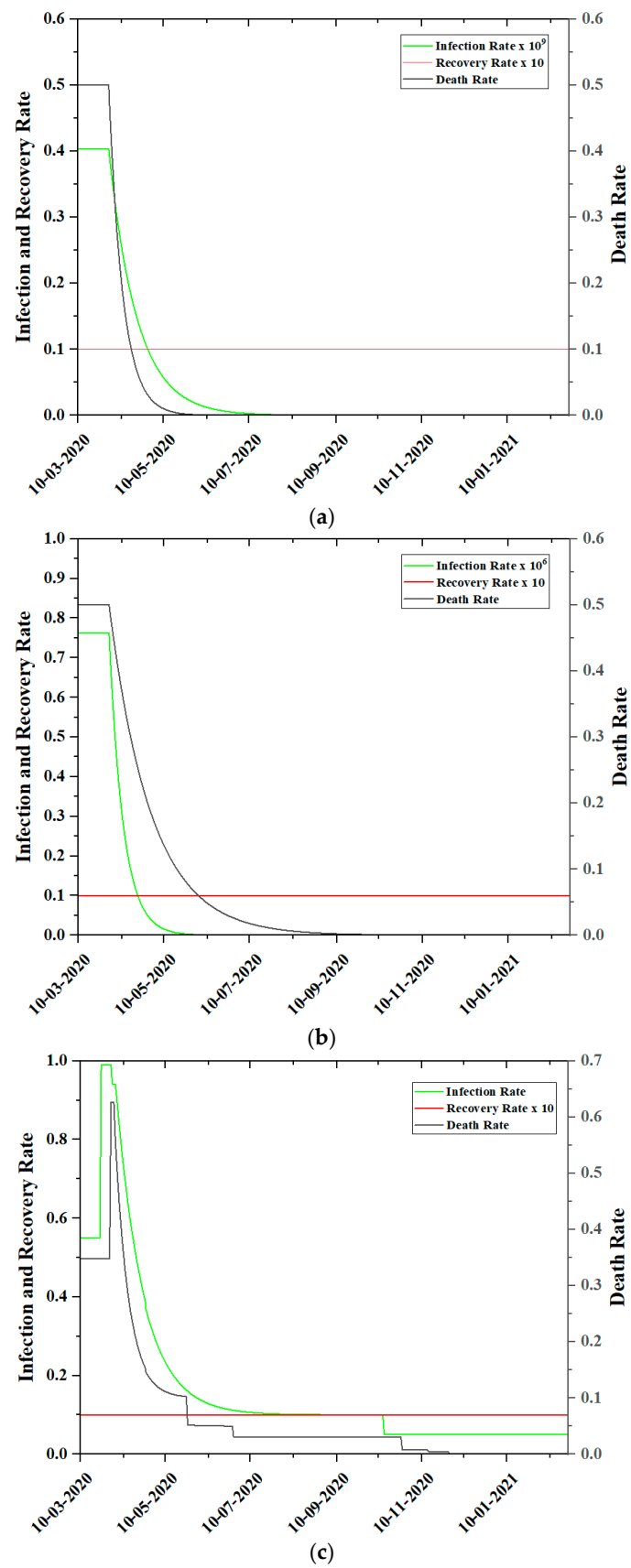


Figure 6. Cont.

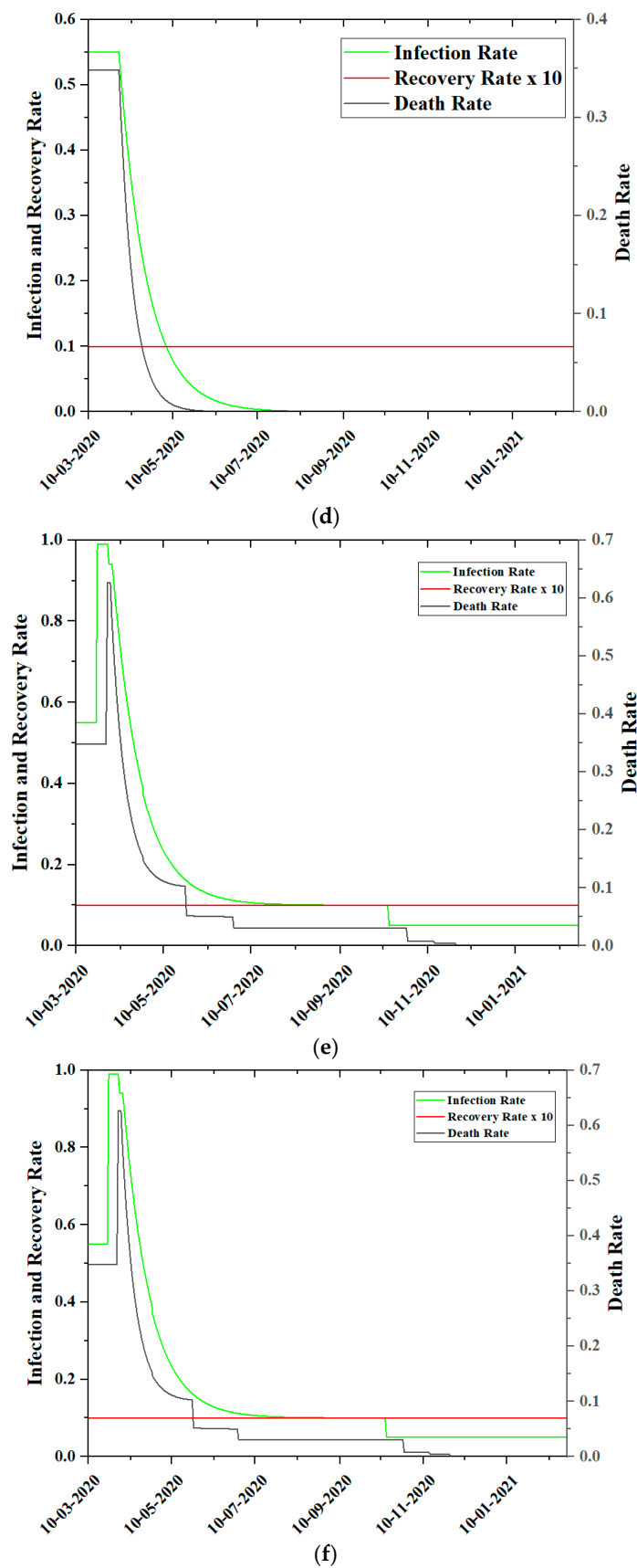


Figure 6. Cont.

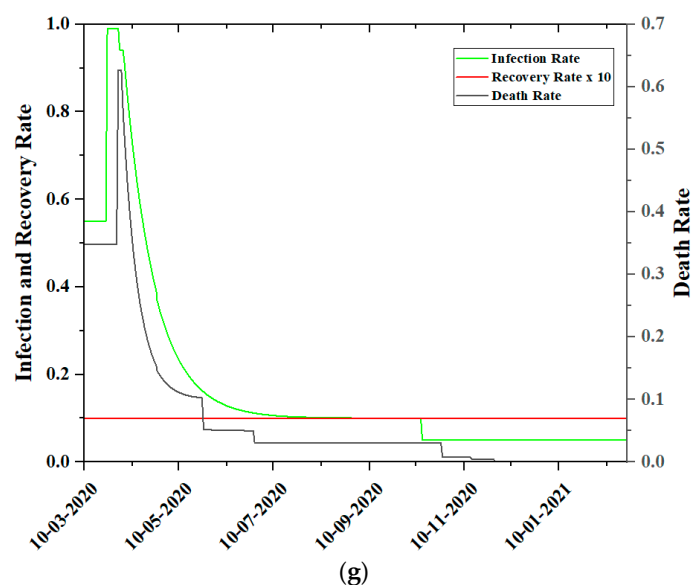


Figure 6. Kinematic of infection, recovery and death for Indian COVID-19 outbreak using (a) G (b) PS (c) PSO (d) G + PS (e) G + PSO (f) G + PS + G (g) G + PSO + G.

Moreover, when compared with the results obtained by PSO, G + PS + G also predicts similar value of T_r . However, the respective values of τ_β , τ_μ , T_d and initial and average R_0 are predicted by G + PS + G as 60.31, 63.61, 12, 3.07 and 4.6 and accordingly a very small variation with respect to PSO of 0.05%, 5.6%, 7.69%, 4.65% and 3.76% has been observed and hence, are also acceptable. The values of τ_β , T_r , τ_μ , T_d and initial and average R_0 obtained by all these techniques have been tabulated in Table 2. The variation in R_0 using G, PS, PSO, G + PS, G + PSO, G + PS + G and G + PSO + G have been demonstrated in Figure 7 respectively for better visualization.

Table 2. Spread of COVID-19 outbreak.

Parameter	Optimization Technique						
	G	PS	PSO	G + PS	G + PSO	G + PS + G	G + PSO + G
τ_β (day)	10.00	46.71	60.34	47.86	60.34	60.31	60.34
T_r (day)	100	46	44	51	44	44	44
τ_μ (day)	10.00	18.28	67.39	14.61	67.39	63.61	67.39
T_d (day)	0.1	2	13	1	13	12	13
Initial R_0	7.91×10^{-11}	1.5×10^{-6}	3.22	1.53	3.22	3.07	3.22
Average R_0	0	0	4.78	2.7	4.78	4.6	4.78

The Indian outbreak prediction curves for COVID-19 using all these techniques for active infected, total ever infected, recovered and death have been demonstrated in Figure 8. It has been found that using PSO, the model predicts last week of September as the peak of Indian outbreak using optimum parameters and thereafter, the number of infected people will gradually decrease. However, at its peak, the number of active infected people predicted by the model is 1,025,407 and towards the end of simulation time (350 days), 273,586 and 10,738,028 active infected and ever infected people respectively, have been estimated. The total number of recovered and death towards the end of simulation of the ongoing pandemic have been predicted as 10,313,876 (96.05%) and 150,566 (1.40%) respectively. Interestingly, the model predicts same time as the peak of Indian outbreak by using G + PS + G also and the total active infected people have been estimated as 1,060,303 at that time. Furthermore, as compared to PSO, a respective difference in ever infected population of 0.21% and 0.08% has been observed at peak and end of simulation time.

The values regarding the forecasting of COVID-19 outbreak in India by all these techniques have been presented in Table 3.

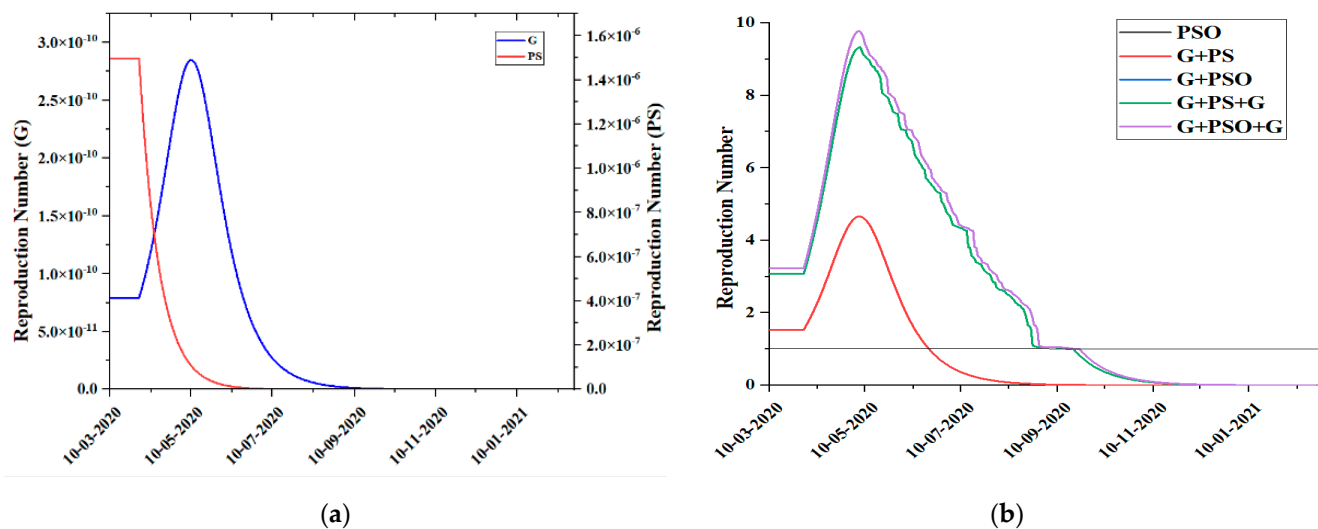


Figure 7. Dynamic behavior of reproduction number (a) G and PS (b) PSO, G + PS, G + PSO, G + PS + G and G + PSO + G.

Table 3. Forecasting of COVID-19 outbreak.

Parameter	Optimization Technique						
	G	PS	PSO	G + PS	G + PSO	G + PS + G	G + PSO + G
P	0	0	200	103	200	198	199
I_P	62	62	1,025,407	2,629,547	1,025,407	1,060,303	1,077,462
R_P	4	4	4,739,742	986,113	4,734,007	4,267,800	4,270,367
D_P	0	0	92,610	370,437	92,610	86,635	85,141
Total ₃₅₀	62	62	10,738,028	4,468,343	10,748,208	10,729,586	10,737,867
I_{350}	1.97×10^{-7}	3.02×10^{-11}	273,586	268,841	249,445	202,423	200,803
R_{350}	56.86	56.86	10,313,876	3,826,073	10,348,197	10,376,514	10,386,498
D_{350}	5.14	5.14	150,566	373,427	150,566	150,649	150,566

Further, the short term effectiveness of the developed models has been compared against actual pandemic values after 15 days (13 December 2020). The obtained results reveal that G, PS and G + PS are incapable to predict the epidemic; however, the models developed by employing any variant PSO and G + PS + G predict the values of I , R and D with significant accuracy (99%) which validates the efficiency of these developed models for accurate prediction of the ongoing pandemic. The obtained results along with reported data have been presented in Table 4.

Table 4. Result validation.

Parameter	Actual	Optimization Technique						
		G	PS	PSO	G + PS	G + PSO	G + PS + G	G + PSO + G
I	353,715	0	0	354,618	610,371	355,219	354,218	354,515
R	9,357,464	5	5	9,431,09	3,484,519	9,398,497	9,428,547	9,407,594
D	143,393	57	57	144,589	373,427	143,796	144,217	143,748

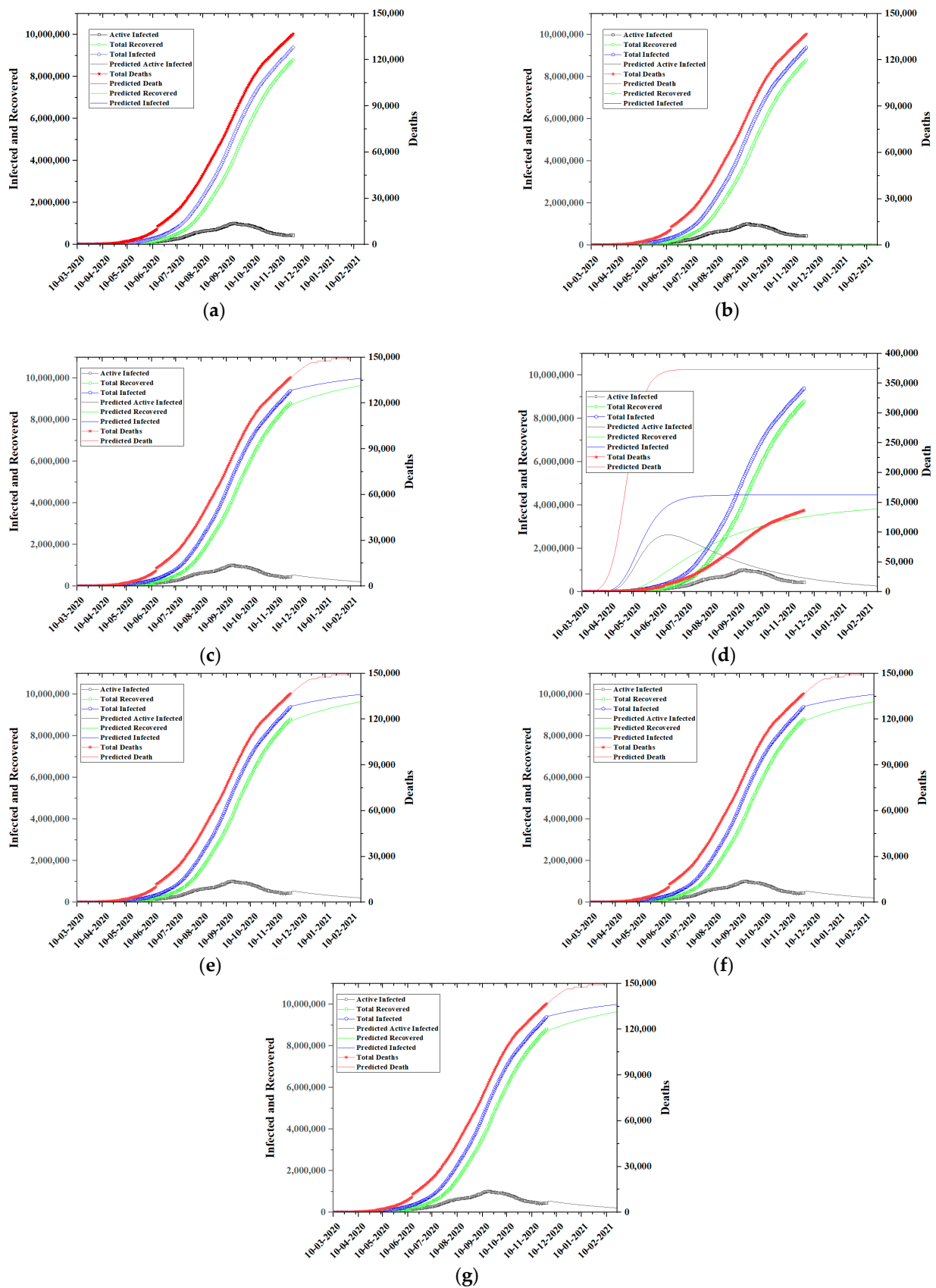


Figure 8. Prediction curve of Indian COVID-19 using (a) G (b) PS (c) PSO (d) G + PS (e) G + PSO (f) G + PS + G (g) G + PSO + G.

8. Conclusions

In the present investigation, the SIRD compartment model has been used for investigating the evolution and prediction of COVID-19 in India. To incorporate behavioral change in key parameters because of lockdown, social distancing and other non-pharmaceutical interventions, dynamic behavior of β , α , μ and hence, R_0 has been considered. The modeling parameters have been optimized using gradient (fmincon), particle swarm, pattern search and their hybrid. Through simulation-based investigation, it has been found that PSO along with any combination of G and G + PS + G produce almost identical results but considering the model complexity and time required for simulation, PSO has assumed to be the best optimization technique and therefore, τ_β of 60 days, τ_μ of 67 days and R_0 of 4.78 have been estimated. However, G, PS and G + PS did not yield optimum fitting, thereby validating their inappropriateness in the current pandemic assessment.

Based on the above parameters, model predicted last week of September as the peak duration of COVID-19 pandemic in India with more than 1,025,407 active infected people at its peak with an accuracy of 97% and even after 350 days from the onset of the pandemic, more than 273,586 people will remain infected with a total of ever infected people crosses 10.7 million. However, by that time, more than 96% people will be recovered and only around 1.4% death has been projected. It also anticipated that even at its peak, around 81% people have been recovered with 1.59% death which is far better than other countries. Nonetheless, it also creates an alarming situation considering the fact that India is among the lowest health workforce density [34] with the ratio of nurse to physicians only 0.6:1 as compared to 3:1 in developed countries [23]. Despite the weak medical infrastructure, comparatively less severe pandemic has been observed which may be because of very high young population accompanied by proactive response by policymakers. However, constant recovery rate indicated that most of the reported cases in India are associated with severe infections. Therefore, the pandemic data could be extremely underestimated in total number of infected as well as recovered.

Notably, the effectiveness of the present investigation extremely lies on the quality of data. The proposed methodology should be used only for quality understanding of the Indian pandemic and crude predictions and not for any change in policy or decision. Moreover, the accuracy of prediction depends on a number of factors such as policy changes (leading to drastic variations in pandemic data), false reported data and modifications in guidelines to report the data, novel findings and introduction of vaccines and so forth. These factors may have a significant influence on the prediction accuracy and are envisioned as the scope of further investigation.

Author Contributions: Conceptualization, H.G., S.K. and O.P.V.; Methodology, H.G., S.K. and T.K.S.; Software, H.G., S.K., and D.Y.; Validation, D.Y., O.P.V. and T.K.S.; Formal analysis, O.P.V., T.K.S., and C.W.A.; Investigation, H.G., O.P.V. and J.-H.L.; Data curation, H.G., S.K., and D.Y.; Writing—original draft preparation, H.G., D.Y. and O.P.V.; Writing—review and editing, T.K.S., C.W.A. and J.-H.L.; Visualization, H.G., C.W.A. and J.-H.L.; Supervision, O.P.V., C.W.A. and J.-H.L.; Funding acquisition, C.W.A. and J.-H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1C1C1009720).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository that does not issue DOIs. Publicly available datasets were analyzed in this study. This data can be found here: reference number [11,28].

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

B	Infection (transmission) rate (per day)
A	Recovery rate (per day)
μ	Death rate (per day)
N	Total population
I	Number of infected
D	Number of deaths
I_p	Number of active infected at peak
D_p	Number of death at peak
T_r	Number of days required for recovery
G	Gradient Descent
τ_β	Characteristic time of infection (day)
τ_α	Characteristic time of recovery (day)
τ_μ	Characteristic time of death (day)
S	Susceptible population
R	Number of recovered
R_0	Reproduction number
R_p	Number of recovered at peak
P	Peak day from the outbreak
T_d	Time required in achieving long term mortality rate
PS	Pattern Search
PSO	Particle Swarm Optimization
$G + PS + G$	Gradient Descent followed by Pattern Search followed by Gradient Descent
$G + PSO + G$	Gradient Descent followed by Particle Swarm Optimization followed by Gradient Descent
$G + PS$	Gradient Descent followed by Pattern Search
$G + PSO$	Gradient Descent followed by Particle Swarm Optimization
$t_{lockdown}$	Number of days after which country wide lockdown implements from 3 March 2020
DFE	Disease Free Equilibrium

References

1. Coronavirus Update (Live): 63,892,165 Cases and 1,480,163 Deaths from COVID-19 Virus Pandemic-Worldometer. Available online: <https://www.worldometers.info/coronavirus/> (accessed on 1 December 2020).
2. Wang, J.; Du, G. COVID-19 may transmit through aerosol. *Ir. J. Med. Sci.* **2020**, *189*, 1143–1144. [CrossRef] [PubMed]
3. Liu, R.; Li, D.; Kaewunruen, S. Role of Railway Transportation in the Spread of the Coronavirus: Evidence From Wuhan-Beijing Railway Corridor. *Front. Built Environ.* **2020**, *6*, 190. [CrossRef]
4. Misra, S.; Jeon, S.; Lee, S.; Managuli, R.; Jang, I.-S.; Kim, C. Multi-Channel Transfer Learning of Chest X-ray Images for Screening of COVID-19. *Electronics* **2020**, *9*, 1388. [CrossRef]
5. Wu, J.T.; Leung, K.; Leung, G.M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet* **2020**, *395*, 689–697. [CrossRef]
6. Chatterjee, S.; Sarkar, A.; Chatterjee, S.; Karmakar, M.; Paul, R. Studying the progress of COVID-19 outbreak in India using SIRD model. *Indian J. Phys.* **2020**, *1*. [CrossRef] [PubMed]
7. Ndairou, F.; Area, I.; Nieto, J.J.; Torres, D.F.M. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solitons Fractals* **2020**, *135*, 109846. [CrossRef] [PubMed]
8. Bedford, J.; Enria, D.; Giesecke, J.; Heymann, D.L.; Ihekweazu, C.; Kobinger, G.; Lane, H.C.; Memish, Z.; don Oh, M.; Sall, A.A.; et al. COVID-19: Towards controlling of a pandemic. *Lancet* **2020**, *395*, 1015–1018. [CrossRef]
9. Jung, Y.; Agulto, R. A Public Platform for Virtual IoT-Based Monitoring and Tracking of COVID-19. *Electronics* **2020**, *10*, 12. [CrossRef]
10. Alamo, T.; Reina, D.; Mammarella, M.; Abella, A. Covid-19: Open-Data Resources for Monitoring, Modeling, and Forecasting the Epidemic. *Electronics* **2020**, *9*, 827. [CrossRef]
11. MoHFW | Home. Available online: <https://www.mohfw.gov.in/> (accessed on 14 December 2020).
12. McKibbin, W.J.; Fernando, R. The Global Macroeconomic Impacts of COVID-19: Seven Scenarios. *SSRN Electron. J.* **2020**. [CrossRef]
13. COVID-19 Testing Rate by Country | Statista. Available online: <https://www.statista.com/statistics/1104645/covid19-testing-rate-select-countries-worldwide/> (accessed on 2 December 2020).
14. Remuzzi, A.; Remuzzi, G. COVID-19 and Italy: What next? *Lancet* **2020**, *395*, 1225–1228. [CrossRef]
15. Batista, M. Estimation of the final size of the second phase of the coronavirus epidemic by the logistic model. *medRxiv* **2020**. [CrossRef]
16. Lalwani, S.; Sahni, G.; Mewara, B.; Kumar, R. Predicting optimal lockdown period with parametric approach using three-phase maturation SIRD model for COVID-19 pandemic. *Chaos Solitons Fractals* **2020**, *138*, 109939. [CrossRef]

17. Reis, R.F.; de Melo Quintela, B.; de Oliveira Campos, J.; Gomes, J.M.; Rocha, B.M.; Lobosco, M.; Weber dos Santos, R. Characterization of the COVID-19 pandemic and the impact of uncertainties, mitigation strategies, and underreporting of cases in South Korea, Italy, and Brazil. *Chaos Solitons Fractals* **2020**, *136*, 109888. [CrossRef] [PubMed]
18. Kermack, W.O.; McKendrick, A.G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. A Contain. Pap. A Math. Phys. Character* **1927**, *115*, 700–721. [CrossRef]
19. Lan, L.; Xu, D.; Ye, G.; Xia, C.; Wang, S.; Li, Y.; Xu, H. Positive RT-PCR Test Results in Patients Recovered from COVID-19. *J. Am. Med. Assoc.* **2020**, *323*, 1502–1503. [CrossRef] [PubMed]
20. Bao, L.; Deng, W.; Gao, H.; Xiao, C.; Liu, J.; Xue, J.; Lv, Q.; Liu, J.; Yu, P.; Xu, Y.; et al. Lack of Reinfection in Rhesus Macaques Infected with SARS-CoV-2. *bioRxiv* **2020**. [CrossRef]
21. Miller, J.C. Mathematical models of SIR disease spread with combined non-sexual and sexual transmission routes. *Infect. Dis. Model.* **2017**, *2*, 35–55. [CrossRef]
22. Jiao, J.; Liu, Z.; Cai, S. Dynamics of an SEIR model with infectivity in incubation period and homestead-isolation on the susceptible. *Appl. Math. Lett.* **2020**, *107*, 106442. [CrossRef]
23. Rai, B.; Shukla, A.; Dwivedi, L.K. COVID-19 in India: Predictions, Reproduction Number and Public Health Preparedness. *medRxiv* **2020**. [CrossRef]
24. Caccavo, D. Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model. *medRxiv* **2020**. [CrossRef]
25. Torrealba-Rodriguez, O.; Conde-Gutiérrez, R.A.; Hernández-Javier, A.L. Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models. *Chaos Solitons Fractals* **2020**, *138*, 109946. [CrossRef] [PubMed]
26. Fraser, C.; Donnelly, C.A.; Cauchemez, S.; Hanage, W.P.; Van Kerkhove, M.D.; Hollingsworth, T.D.; Griffin, J.; Baggaley, R.F.; Jenkins, H.E.; Lyons, E.J.; et al. Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* **2009**, *324*, 1557–1561. [CrossRef] [PubMed]
27. Heffernan, J.; Smith, R.; Wahl, L. Perspectives on the basic reproductive ratio. *J. R. Soc. Interface* **2005**, *2*, 281–293. [CrossRef]
28. Coronavirus COVID-19 Tracker (INDIA): Corona News Dashboard from India and World. Available online: <https://www.indiatoday.in/coronavirus-cases-tracker-dashboard/> (accessed on 14 December 2020).
29. Countries by Population Density 2020-StatisticsTimes.com. Available online: <http://statisticstimes.com/demographics/countries-by-population-density.php> (accessed on 2 December 2020).
30. Gururaja, K.; Sudhira, H. Population crunch in India: Is it urban or still rural? *Curr. Sci.* **2012**, *103*, 37–40.
31. India Goes Past China in Covid-19 Cases But With Lower Rate of Fatalities-India News-Hindustan Times. Available online: <https://www.hindustantimes.com/india-news/india-goes-past-china-in-covid-cases-but-with-lower-rate-of-fatalities/story-1rlSwXO3Z2oVLNGDjpknYM.html> (accessed on 2 December 2020).
32. Ma, Y.; Zhao, Y.; Liu, J.; He, X.; Wang, B.; Fu, S.; Yan, J.; Niu, J.; Luo, B. Effects of temperature variation and humidity on the mortality of COVID-19 in Wuhan. *medRxiv* **2020**. [CrossRef]
33. What Is Herd Immunity and How Can We Achieve It With COVID-19?—COVID-19—Johns Hopkins Bloomberg School of Public Health. Available online: <https://www.jhsph.edu/covid-19/articles/achieving-herd-immunity-with-covid19.html> (accessed on 2 December 2020).
34. World Health Statistics 2019: Monitoring Health for The SDGs, Sustainable Development Goals. Available online: <https://www.who.int/publications/i/item/world-health-statistics-2019-monitoring-health-for-the-sdgs-sustainable-development-goals> (accessed on 2 December 2020).