

# Analysis, Modeling, and Representation of COVID-19 Spread: A Case Study on India

Rahul Mishra<sup>ID</sup>, Graduate Student Member, IEEE, Hari Prabhat Gupta<sup>ID</sup>, and Tanima Dutta<sup>ID</sup>

**Abstract**—Coronavirus outbreak is one of the challenging pandemics for the entire human population on Earth. Techniques, such as the isolation of infected people and maintaining social distancing, are the only preventive measures against the pandemic. The actual estimation of the number of infected peoples with limited data is an indeterminate problem faced by data scientists. There are several techniques in the existing literature, including reproduction number and case fatality rate, for predicting the duration of a pandemic and infectious population. This article presents a case study of different techniques for analyzing, modeling, and representing the data associated with a pandemic such as COVID-19. We further propose an algorithm for estimating infection transmission states in a particular area. This work also presents an algorithm for estimating end time of a pandemic from the susceptible infectious and recovered model. Finally, this article presents the empirical and data analysis to study the impact of transmission probability, rate of contact, infectious, and susceptible population on the pandemic spread.

**Index Terms**—Coronavirus, infectious, isolation, pandemic.

## I. INTRODUCTION

**A**DISEASE is a harmful deviation from the normal condition that adversely affects human health [1]. Every disease is a medical challenge, which may have symptoms and signs for their identification. These symptoms facilitate easier spotting of the disease in the early stage. However, some diseases do not have symptoms, which makes them untraceable. The rapid advancement in medical science helps in curing most of the existing diseases. However, some diseases are spread due to unidentified viruses or bacteria, which creates pandemics such as COVID-19. The newly identified novel coronavirus is responsible for the outbreak of COVID-19 that has the potential for maximal death [2]. The effect of the pandemic (COVID-19) and its preventive measures, such as quarantine and social distancing, can be analyzed using data analysis, modeling, and representation techniques of data mining [3], [4]. These techniques help in predicting disease's behavior, thus providing the convenience of taking necessary preventive measures.

Reproduction number is an essential parameter for predicting the behavior of a pandemic in terms of infection

spread. It estimates the spread of secondary infections from the primarily infected people. Next, the case fatality rate captures the number of deaths during a pandemic spread. It plays a vital role in estimating the actual damage to human lives. Furthermore, there are four states of infection transmission in a pandemic, i.e., no contact, local, untraceable, and source missing. The first state is the least harmful and often occurs at the initial stage of the pandemic, whereas the last state is critical, indicating that the spread of infection is severe. In the literature, different pandemic models are discussed, including susceptible, infectious, and recovered (SIR), susceptible infectious (SI), and susceptible, infectious, and susceptible (SIS) model.

We are currently living in the information-centric era that encourages prediction and analysis of different parameters associated with a pandemic disease. The predictions are performed by the models that take the historical data to learn different patterns. The performance of these models depends heavily upon data representation and storage. Therefore, in this article, we summarize different techniques presented in the existing literature for estimating various parameters of the pandemics. We further discuss the models that predict the end time and the total population infected and susceptible.

## A. Related Work

The impact of the heterogeneous recovery rate of a pandemic on a complex network is analyzed by de Arruda *et al.* [1]. They have used the modified version of the SIS model to determine the infectivity. Li *et al.* [2] developed a new model named SEIQR. It incorporates susceptible (*S*), exposed (*E*), infectious (*I*), confirmed (*Q*), and recovered (*R*). SEIQR helped in estimating the imported cases of COVID-19 in the discrete-time interval. Zhou *et al.* [5] proposed the mechanism of estimating the basic reproduction number using the SEIR model, where they have estimated the transmission probability of COVID-19 in China. Next, Van den Driessche and Watmough [6] proposed the mechanism for analyzing the disease-free equilibrium on the heterogeneous population. They proposed an epidemic model named SEIT, which incorporates susceptible (*S*), exposed (*E*), infected (*I*), and treated (*T*). The methodology for epidemiological modeling is discussed by Ranjan [7], where he has predicted the reproduction number of the COVID-19.

Furthermore, Zhang *et al.* [8] proposed an improved version of the SIR model for characterizing the dynamics of a

Manuscript received June 26, 2020; revised December 17, 2020, February 7, 2021, and April 18, 2021; accepted April 25, 2021. Date of publication May 17, 2021; date of current version August 2, 2021. (Corresponding author: Rahul Mishra.)

The authors are with the Department of Computer Science and Engineering, IIT (BHU) Varanasi, Varanasi 221005, India (e-mail: rahulmishra.rs.cse17@iitbhu.ac.in; hariprabhat.cse@iitbhu.ac.in; tanima.cse@iitbhu.ac.in).

Digital Object Identifier 10.1109/TCSS.2021.3077701

TABLE I  
SUMMARY OF SOME EXISTING WORK ON PANDEMIC MODELING

Paper	Objective	Pandemic parameters			Pandemic model
		$R_o$	$C_r$	State	
[1]	To study the recovery rate distribution	✗	✗	✗	SIS
[2]	To identify discrete time imported cases	✓	✗	✓	SEIQR
[5]	To estimate $R_o$ of reported cases	✓	✓	✗	SEIR
[6]	To analyze disease free equilibrium	✓	✓	✗	SEIT
[7]	Epidemiological modeling	✓	✗	✓	SIR
[8]	To characterize the dynamics of pandemic	✗	✗	✓	SIR
[9]	To study the influence of preventive measures	✗	✗	✓	SIS
[10]	To estimate pandemic threshold of spread	✗	✗	✗	SIS

pandemic on social contacts. The layered architecture for modeling the pandemic is discussed in [9]. Finally, Shi *et al.* [10] examined the pandemic threshold for identifying the spread of disease. Table I presents a comparative summary of some existing work on pandemic modeling. It compares the pandemic models and parameters, including reproduction number ( $R_o$ ), case fatality rate ( $C_r$ ), and state of infection transmission.

### B. Major Contributions

To the best of our knowledge, this is the first work as a case study that covers data analysis, modeling, and representation of the coronavirus spread in India. This article makes the following major contributions:

- 1) *Reproduction Number*: First, we describe the role of the reproduction number for estimating the infected population during pandemic such as COVID-19. We also determine the average value of the reproduction number for India up to the specified date by analyzing the existing dataset.
- 2) *Case Fatality Rate*: This work analyzes the role of case fatality rate while estimating the damage caused by the pandemic. We describe the necessity of case fatality rate along with the reproduction number while estimating the infected and recovered population.
- 3) *Infection Transmission States*: Next, we describe the different infection transmission states of a pandemic such as COVID-19. This work also proposes an algorithm for iteratively estimating the infection transmission states.
- 4) *Data Modeling*: Furthermore, we present different models of a pandemic, including SIR, SI, and SIS. We also propose an algorithm for estimating the end time of a pandemic using the SIR model.
- 5) *Empirical Analysis*: Finally, this article presents an empirical analysis to study the impact of transmission probability, rate of contact, infectious, and susceptible population.

The rest of this article is organized as follows. Section II illustrates the COVID-19 open-access dataset and the parameters used for measuring the impact of a pandemic. Section III

illustrates the different mathematical models of the pandemic. Section IV illustrates the empirical results and this article is concluded in Section V.

## II. DATASET AND PARAMETERS FOR MEASURING THE SPREAD OF A PANDEMIC DISEASE

In this section, we first illustrate an open-access dataset of COVID-19 that we use in the rest of this article. We next define the parameters used for measuring the impact of the pandemic.

### A. Dataset of COVID-19

This work uses the COVID-19 dataset available in [11] for predicting and analyzing the pandemic in India. The dataset illustrates the detailed description of the patients infected from coronavirus. Table II shows the first eight entries of the dataset having different columns, including patient number, state patient number, date announcement, age bracket, gender, detected city, detected state, state code, current status, nationality, type of transmission, and status change date. The column header *patient number* indicates the actual patient count in India, and *state patient number* is a combination of state code, detected city code, and patient count. The column *current status* indicates the current state of the patient. Next, *nationality* followed by *type of transmission* that depicts whether the transmission of infection is imported from other countries or local. Finally, *status change date* illustrates the status change of infected patient to recovered or deceased.

### B. Pandemic First Parameter: Reproduction Number

The pandemic disease modeling incorporates the reproduction number (denoted by  $R_o$ ) to understand the disease outbreak by wrapping up all the infection transmission in a particular area. The reproduction number specifies the number of people being infected by the infectious population [5]. We can define  $R_o$  as follows.

*Definition 1*: The reproduction number of a given region is the ratio of the total number of people infected to the total number of people who are infecting, i.e.,

$$R_o = \frac{\text{Total number of people infected}}{\text{Total number of people who are infecting}} \quad (1)$$

where “the total number of people infected” means the number of secondary cases from the primary cases. Next, “the total number of people who are infecting” means the number of patients who have infected other people.

$R_o$  is an essential metric that determines: *how contagious the disease is?* A disease is said to be highly contagious when an infected individual transmits the infection to a large number of people. The mechanism of providing a proper treatment or appropriate handling of any disease requires full information about how the infection develops in the body, the mode of transmission, and the social structure where it can spread. If all this information is available, then the outbreak can be detected and monitored easily. The reproduction number  $R_o$  captures all this necessary information by incorporating three components: population density, rate of infection, and rate of

TABLE II  
DATASET OF COVID-19 [11] (TOTAL ROWS ARE 20 000+ AS ON APRIL 23, 2020)

Patient Number	State Patient Number	Date Announced	Age Bracket	Detected City	Detected State	State code	Current Status	Nationality	Type of transmission	Status Change Date
1	KL-TS-P1	30/01/2020	20	Thrissur	Kerala	KL	Recovered	India	Imported	14/02/2020
2	KL-AL-P1	02/02/2020		Alappuzha	Kerala	KL	Recovered	India	Imported	14/02/2020
3	KL-KS-P1	03/02/2020		Kasaragod	Kerala	KL	Recovered	India	Imported	14/02/2020
4	DL-P1	02/03/2020	45	East Delhi	Delhi	DL	Recovered	India	Imported	15/03/2020
5	TS-P1	02/03/2020	24	Hyderabad	Telangana	TG	Recovered	India	Imported	02/03/2020
6		03/03/2020	69	Jaipur	Rajasthan	RJ	Recovered	Italy	Imported	03/03/2020
7		04/03/2020	55	Gurugram	Haryana	HR	Recovered	Italy	Imported	29/03/2020
8		04/03/2020	55	Gurugram	Haryana	HR	Recovered	Italy	Imported	29/03/2020

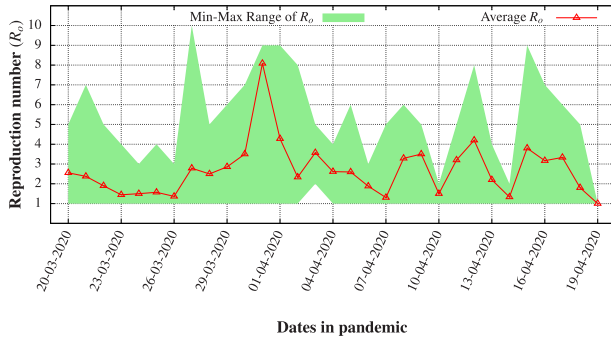


Fig. 1. Illustration of  $R_o$  for COVID-19 spread in India.

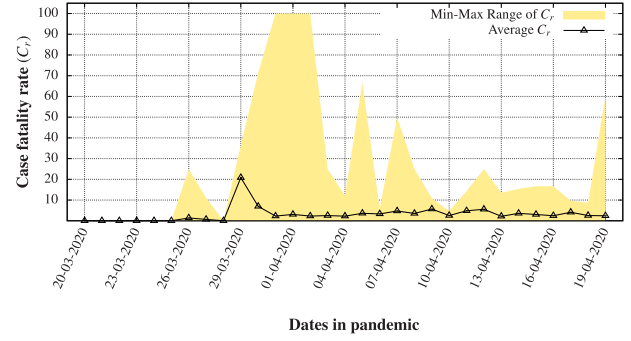


Fig. 2. Illustration of  $C_r$  in India on specified dates.

recovery [6]. The estimation accuracy of  $R_o$  does not depend on the dataset but relies on the mechanism of the calculated number of peoples being infected.

Fig. 1 shows the value of  $R_o$  for the given dates. It also shows the minimum and maximum values of  $R_o$  in the states of India. We conclude that the difference between the minimum and maximum values of  $R_o$  is very high, which indicates that the pandemic is very high in some states while others at lower threat. The value of  $R_o$  changes over time, as shown in Fig. 1. Here, though the value of  $R_o$  is time-varying but is always greater than 1. As the value of  $R_o \geq 1$  indicates the outbreak of the pandemic. Furthermore, we observed from Fig. 1 that the value of  $R_o$  achieves sharp peaks and sudden decrements. It is because the reported infection transmission on some specified date can sharply increase or decrease independent of previous days. Table III shows the value of  $R_o$  for the spread of COVID-19 in different states of India from the dataset available in [11] up till April 18, 2020. The average  $R_o$  value of some states is less than 1, which is a good indication of lower spread.

#### C. Pandemic Second Parameter: Case Fatality Rate

The case fatality rate (denoted by  $C_r$ ) provides an estimation of the number of people who dies among the total infected people from the disease during its spread in a particular area.

**Definition 2:** Case fatality rate of a given region is the ratio of the total number of people who died from the pandemic to the total number of the people who are infected, i.e.,

$$C_r = \frac{\text{Total death from the pandemic}}{\text{Total number of people who are infected}} \times 100. \quad (2)$$

The reproduction number only provides the rate of disease spread but cannot estimate how many people lose their lives in the entire cycle of the spread. Therefore, the case fatality rate is crucial for such estimation. The pandemic such as measles with  $R_o$  around 18 has a lower case fatality rate to that of Ebola with  $R_o$  around 1.5 [12]. Fig. 2 shows the value of  $C_r$  for the given dates. It also shows the minimum and maximum values of  $C_r$  in the states of India. We conclude that the difference between the minimum and maximum values of  $C_r$  is very high, which indicates that a pandemic is very high in some state, while others are free. The result also illustrates the minimum value of  $C_r$  increases with time. It indicates that the pandemic is spreading in India at a slower rate. Table IV shows the case fatality rate of different states of India.

#### D. Pandemic Third Parameter: State of Given Region

Measures, such as social/physical distancing, reduce the spread of the disease infections by breaking the transmission chains. infection\_transmission of each individual can be easily obtained by tracing the infected person's contact with other people. In this work, we are using a dataset available in [11], which also provides a dedicated column for tracing out the infection source. Therefore, we can easily determine infection\_transmission of each individual. The spread of the infection is categorized in the following four states [7], [13]:

- 1) *State 1 (No Contact Transmission):* In this transmission state, the infection from an infected person does not proceed further to any other person. The infected person remains confined with the disease-causing virus without further dissipation. For a pandemic disease such

TABLE III  
STATEWISE REPRODUCTIVE NUMBER ( $R_o$ ) USING THE COVID-19 DATASET [11]

Andaman and Nicobar																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Andhra Pradesh																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	3	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.8
Arunachal Pradesh																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Assam																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bihar																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	5.33
Delhi																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3.34
Gujarat																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.5
Haryana																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6
Himanchal Pradesh																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jammu and Kashmir																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	3	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2.83
Karnataka																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	35	14	10	7	2	2	0	2	0	0	1	1	0	0	0	1	0	0	0	1	2.73
Kerala																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	7	5	1	3	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.8
Madhya Pradesh																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	2	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3.6
Maharashtra																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	2	2	2	1	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3.7
Odisha																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Punjab																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	7	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	3.91
Rajasthan																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	4	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.75
Tamil Nadu																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	6	0	2	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	3.30
Uttar Pradesh																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	7	0	1	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2.58
Uttarakhand																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
West Bengal																					
Person contact	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Avg.
Count	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.5
India																			Avg. = 1.79		



TABLE IV  
ILLUSTRATION OF CASE FATALITY RATE IN STATES OF INDIA

State/UT	Fatality rate (in%)	State/UT	Fatality rate (in%)
Andaman and Nicobar	0	Bihar	0.71
Arunachal Pradesh	0	Assam	0
Andhra Pradesh	0	Chandigarh	0
Chattisgarh	0	Delhi	0.044
Dadra Nagar Haveli	0	Goa	0
Gujrat	0.17	Haryana	0
Himanchal Pradesh	2.56	Kerla	0.46
Jharkhand	0	Karnataka	2.48
Jammu & Kashmir	0.246	Ladakh	0
Madhya Pradesh	0.126	Maharashtra	0.195
Manipur	0	Mizoram	0
Odisha	1.21	Puducherry	0
Punjab	0.72	Rajasthan	0.052
Tamil Nadu	0.061	Telangana	0.11
Tripura	0	Uttar Pradesh	0.069
Uttarakhand	0	West Bengal	0.874
<b>India</b>			<b>0.34 (in%)</b>

as COVID-19, this state is the most crucial state to control coronavirus spread through quarantine and social isolation. For example, the topmost layer in Fig. 3 demonstrates two persons (i.e., A and B) are infected with coronavirus, and there is no further transmission.

- 2) *State 2 (Local Transmission)*: The isolation compromise and delay in the detection of infection lead to this state. The infected person transfers the infection to the person in its close contact. This type of transmission is usually observed when the infected person comes in close contact with a person such as a friend or a family member. At this stage, it is easy to trace disease spread. The second layer in Fig. 3 indicates that the transmission is from infected person A to C and B to D.
- 3) *State 3 (Untraceable Transmission)*: The unaltered social contacts result in a third transmission state. In this state, the concept of transmitting and being transmitted is followed. Here, an infected person transfers the infection to another person who transfers to the third person. Here, the source of the infection is untraceable. Fig. 3 shows transmission State 3. In this state, A transfers infection to C that further propagates to E. Similar infection transmission is observed among B, D, and F.
- 4) *State 4 (Source Missing Transmission)*: This is the final infection transmission state where the source of infection is unidentified. It can be understood by an example where a person is infected despite not coming in close contact with an infected person and had no travel history from an infected country.

Algorithm 1 illustrates the steps for identifying the infected population in the different transmission states. In the algorithm, we introduce a variable *infection\_transmission* that predicts the chance of a person to lie in a particular state depending on its values (i.e., 0, 1, and more). *infection\_transmission* = 0 indicates that there is no person-to-person infection transmission and *infection\_transmission* = 1 indicates that infection transmission is in State 2. Furthermore, if the infection is transmitted from the first person to the

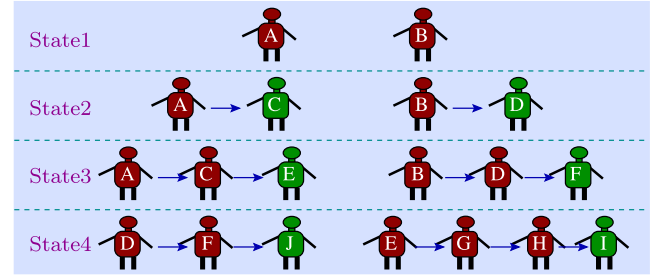


Fig. 3. Example scenario representing four transmission states.

---

**Algorithm 1:** Pandemic State Identification

---

**Input:** Population  $M$  of a particular area (state/country);  
**Output:** States in which all the infected population lie;  
 /\*Variable initialization\*/  
 1  $Set_1 = \{\}$ ,  $Set_2 = \{\}$ , and  $Set_3 = \{\}$ .  
 2  $infection\_transmission \leftarrow 0$ .  
 3 Identify the infectious population  $I$ .  
 4 **for**  $i \leftarrow 1$  to  $\|I\|$  **do**  
 5   **if**  $infection\_transmission_i == 0$  **then**  
 6      $Set_1 \leftarrow Set_1.append(person_i)$ .  
 7   **else if**  $infection\_transmission_i == 1$  **then**  
 8      $Set_2 \leftarrow Set_2.append(person_i)$ .  
 9   **else**  
 10     $Set_3 \leftarrow Set_3.append(person_i)$ .  
 11 The resultant sets  $Set_1$ ,  $Set_2$ , and  $Set_3$  contain the infected person in different transmission state, i.e., State 1, State 2, and State 3, respectively.

---

second person and then to the third or the source is untraceable, this indicates the transmission State 3. If the transmission state does not belong to the above three, then the infection transmission State 4 is reached.

Fig. 4 shows the number of peoples in the different states of infection transmission on the specified dates. The figure depicts the increment in the registered cases with the passing time. An interesting observation from this result is that the number of peoples in State 2 increases sharply than State 1. Table V shows the different states of infection transmission in India. Due to several cases in Maharashtra, a number of peoples in transmission State 2 are high. The table is created using the data available in [11] (April 18, 2020). An important conclusion from the table is that India is in Stage 2 (90.11%) and shifting toward Stage 3 (5.13%).

### III. PANDEMIC MODELS AND CONTACT TRACKING

Data modeling helps in quantifying the damage caused by the pandemic (such as COVID-19) and its probable end time. The models simulate the infection transmission from person to person and its geographical spread on the entire population. This section discusses the existing pandemic models. In addition, we estimate different parameters using the existing dataset.

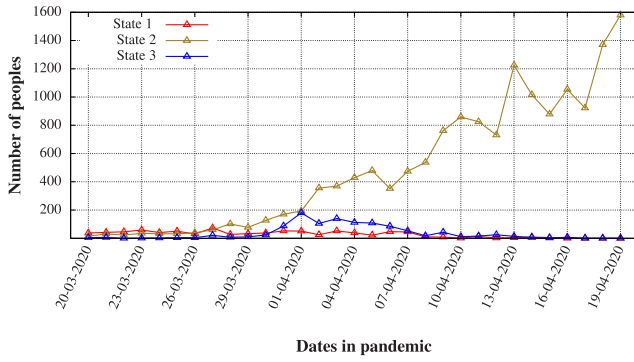


Fig. 4. Illustration of the number of peoples in the different states of infection transmission on the specified dates.

TABLE V  
ILLUSTRATION OF STATES OF TRANSMISSION IN INDIA

State/UT	State1	State2	State3
Andaman and Nicobar	5	1	2
Andhra Pradesh	18	731	64
Arunachal Pradesh	-	-	1
Assam	1	10	25
Bihar	24	111	8
Chandigarh	12	14	1
Chattisgarh	6	29	1
Dadra Nagar Haveli	-	-	-
Delhi	22	2192	34
Goa	7	-	-
Gujrat	121	2280	6
Haryana	29	219	16
Himanchal Pradesh	3	21	15
Jammu and Kashmir	21	380	6
Jharkhand	2	47	-
Karnataka	89	224	114
Kerla	259	156	22
Ladakh	7	-	-
Madhya Pradesh	17	1568	2
Maharashtra	82	5547	20
Manipur	1	-	1
Odisha	19	64	1
Puducherry	1	2	4
Punjab	30	223	25
Rajasthan	96	1811	28
Tamil Nadu	70	1046	513
Telangana	29	888	26
Tripura	-	2	-
Uttar Pradesh	25	1264	160
Uttarakhand	5	37	4
West Bengal	16	438	2
INDIA (Total)	1017	19305	1101
INDIA (in %)	4.74	90.11	5.13

#### A. SIR Model

SIR model [8] is one of the widely used models to figure out the damage caused by a pandemic. The SIR model incorporates the three types of people, i.e., susceptible, infectious, and recovered denoted as  $S$ ,  $I$ , and  $R$ , respectively. Here, the people of one type switch to another type with a specific course of action like the frequent contact with infectious converts  $S$  into  $I$ . If a pandemic covers the entire population  $M$  having  $S$ ,  $I$ , and  $R$  type of peoples, then  $M = S + I + R$ . The phases (type of people) of SIR models are shown in Fig. 5.

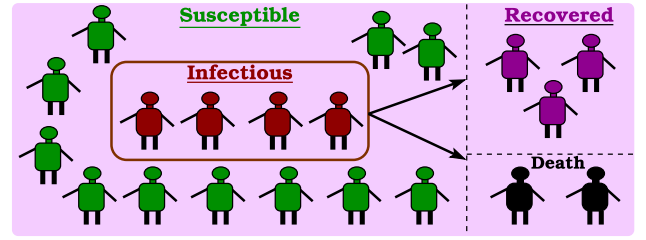


Fig. 5. Illustration of SIR population.

All the people in a particular area who are not infected with the disease and have no reported immunity against the infection are categorized as susceptible ( $S$ ). The absence of immunity in a person results from no prior exposure to the infection and no vaccination against such infection. During the disease outbreak such as COVID-19, some people came in close contact with the infected person and got infected afterward, forming a group called infectious ( $I$ ). These people are solely responsible for the spread of the disease in the community. The quarantine of such infected people preserves the community from being infectious. Every disease such as pandemic COVID-19 has a specific cycle for infection persistence. After this cycle, the infected population recovers from the infection with proper treatment. However, the severe spread of infection in the human body and inadequate cure can result in death. The summation of total recovered and total dead is termed recovered ( $R$ ) or sometimes removed.

The transmission of the disease changes the peoples count in a phase over time. The SIR model can be represented as [9]

$$R = \alpha_2 I - \gamma I \quad (3)$$

$$I = \alpha_1 S \quad (4)$$

$$M = S \cup I \cup R \quad (5)$$

where  $\alpha_1$  and  $\alpha_2$  are force of infection and recovery rate, respectively,  $\gamma$  denotes the death rate, and  $\cup$  is the union operator.

The force of infection  $\alpha_1$  is the rate at which the susceptible person becomes infectious on close contact with the infected person. Recovery rate  $\alpha_2$  is the rate at which an infected person recovers from the infection using the proper treatment. The force of infection is not a constant quantity and is proportional to the transmission rate  $\tau$ . The transmission rate  $\tau$  is the product of the rate of contact  $r_c$  and the probability of transmission  $p_t$  from susceptible to infectious

$$\tau = r_c \times p_t. \quad (6)$$

Using (6), we can define force of infection  $\alpha_1$  as

$$\alpha_1(I) = \tau \left\{ \frac{I}{M} \right\} \times 100 \quad (7)$$

where  $\{I/M\} \times 100$  represents the percentage of population infectious among the total population in a particular area.

To capture the population change in a phase, the SIR model adopts differential equations. A differential equation captures the progression of the disease. The rate of change

**Algorithm 2: SIR**


---

**Input:** Population  $M$  of a particular area facing pandemic;  
**Output:** Time  $T$  for which the pandemic persist;  
/\*Variable initialization\*/  
1  $\alpha_1 \leftarrow 0, \alpha_2 \leftarrow 0, \tau \leftarrow 0$ , and  $r_c \leftarrow 0$ .  
2 Time  $t \leftarrow 0$  and change in time  $\Delta t$ .  
3 Calculate reproduction number  $R_o$ .  
4 **while**  $R_o < 1$  **or**  $\frac{dS}{dt} = \frac{dI}{dt} = \frac{dR}{dt} = 0$  **do**  
5     Estimating transmission rate  $\tau$   
6     Calculate rate of contact ( $r_c$ ) among the persons.  
7     Calculate probability of transmission ( $p_t$ ).  
8      $\tau \leftarrow r_c \times p_t$ .  
9     Estimate *force of infection*  $\alpha_1$  using Eq. 7.  
10    Calculate  $R$  and  $I$  for given  $S$ .  
11     $I = \alpha_1 S$  and  $R = \alpha_2 I - \gamma I$  using Eqs. 4 and 3.  
12    Calculate the derivatives  $\frac{dS}{dt}$ ,  $\frac{dI}{dt}$ , and  $\frac{dR}{dt}$ .  
13     $t \leftarrow t + \Delta t$ .  
14 Either Disease Free or Endemic Equilibrium is reached.  
15 The pandemic will be ended after  $t$  time.

---

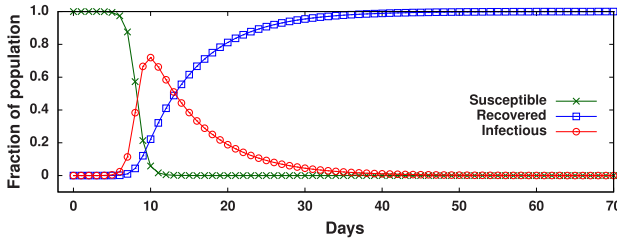


Fig. 6. Illustration of SIR model on the COVID-19 dataset of India up till April 18, 2020.

in  $S$ ,  $I$ , and  $R$  are defined as follows:

$$\frac{dS}{dt} = -\alpha_1(I)S \quad (8)$$

$$\frac{dI}{dt} = \alpha_1(I)S - \alpha_2 I \quad (9)$$

$$\frac{dR}{dt} = \alpha_2 I. \quad (10)$$

An essential step after estimating the rate of change in  $S$ ,  $I$ , and  $R$  using the differential equation is to determine the SIR model's equilibrium state. To describe the pandemic methodology, two equilibria exist side-by-side, i.e., disease-free equilibrium and endemic equilibrium [14]. Algorithm 2 illustrates the step of estimating the time  $t$  for which a pandemic persists using the SIR model.

Fig. 6 shows the relationship between SIR cases in India up till April 18, 2020. The figure demonstrates that infectious persons grow up to a limiting value, where recovery starts at a higher pace. It compels most people to come out of the susceptible, and the outbreak is about to die. During the construction of the result, we have taken the value of reproduction number  $R_o = 1.79$ .

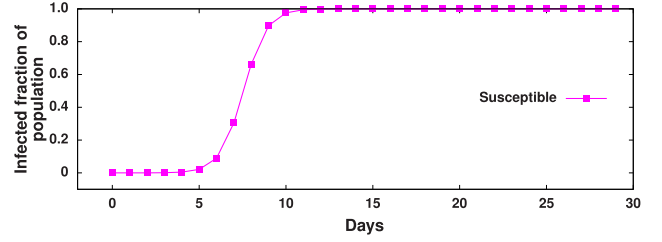


Fig. 7. Illustration of infectious population using the SI model on the COVID-19 dataset of India up till April 18, 2020.

### B. SI Model

Ranney *et al.* [15] proposed the SI model and assumed that the total population in a particular area during pandemic belongs either to susceptible or infectious. Apart from the SIR model, they do not consider the recovered cases in a particular location. Let, at time  $t = 0$ , the number of infectious and susceptible peoples be represented as  $I_o$  and  $S_o$ , respectively. Let  $\tau$  represent the transmission rate given in (6). Then, the rate of change in susceptible and infectious is given as

$$\frac{dS}{dt} = -\tau SI \quad (11)$$

$$\frac{dI}{dt} = \tau I(1 - I). \quad (12)$$

Upon further calculation as given in [15], the number of infectious during a pandemic breakdown can be obtained as

$$I = \frac{I_o e^{\tau t}}{1 - I_o + I_o e^{\tau t}}. \quad (13)$$

Fig. 7 shows the fraction of the infectious population for the past 30 days of the pandemic (COVID-19) breakout in India. The result shows that the curved we obtained is an S-shaped curve for infectious. The curve grows exponentially from the early stage of infection in the country, and the infection starts spreading. During the result estimation, we have taken the value of reproduction number  $R_o = 1.79$ .

### C. SIS Model

The SIS model is an extension of the SI model [10], where the authors assumed that people who recovered from the infection are still susceptible. The rate of change in susceptible is given as

$$\frac{dS}{dt} = -\tau SI + \alpha I \quad (14)$$

where  $\tau$  and  $\alpha$  represent the transmission rate and fraction of the infected population, respectively. The term  $\tau SI$  can be understood as an average infection transmitted by an individual to  $\tau M$  susceptibles. The rate of change in infectious is

$$\frac{dI}{dt} = \tau SI - \alpha I. \quad (15)$$

Using the assumption that the pandemic could infect the entire population such that  $M = S + I$ , we obtain

$$\begin{aligned} \frac{dI}{dt} &= \tau I(M - I) - \alpha I \\ &= (\tau M - \alpha)I - \tau I^2. \end{aligned} \quad (16)$$

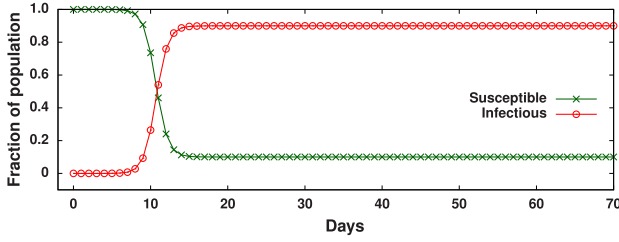


Fig. 8. Illustration of infectious and susceptible population curve of the SIS model on the COVID-19 dataset of India up till April 18, 2020.

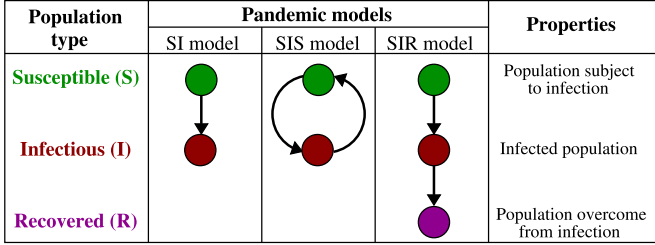


Fig. 9. Overview of pandemic models (SI, SIS, and SIR).

The result shown in Fig. 8 depicts the relationship between susceptible and the infectious fraction of population in India up till April 18, 2020. The result is estimated by taking the value of the reproduction number as  $R_o = 1.79$ , i.e., current status for India with a total 12 079 active cases. If the rate of infection in India is decreasing or not exceeding the current rate, then the spread of COVID-19 would be ended in the next 70 days. Fig. 8 shows that the rate of susceptible and the rate of infectious remain constant between 20 and 70 days. In the SIS model, people who are recovered from the infection are still susceptible. Moreover, if the rate of susceptible is constant, then no new person will be infected during the pandemic. Furthermore, at given recovery and case fatality rates, the total population being infected will be recovered in the next 70 days. Thus, no new population will be infected, which indicates that the pandemic is ended.

Fig. 9 shows the entities of different pandemic models, including SI, SIS, and SIR. Here, in the SIS model, susceptible and infectious are forming a cycle, where the susceptible population being infected are still susceptible.

#### IV. EMPIRICAL EVALUATION

This section empirically evaluates the performance of the SIR model on the population size of 1000. During the entire evaluation, we have taken the value of reproduction number ( $R_o$ ) as 1.79. The empirical results on the SIR model rely on the following expression [16]:

$$I_{total} = I_o + r_c \left( \frac{I}{M} \right) \times 100 \times S \times p_t \quad (17)$$

where  $I_o$  is the initial infected population,  $r_c$  is the rate of contact, and  $p_t$  is the transmission probability.

We have considered only the SIR model in the empirical analysis due to its broader applicability for modeling pandemic such as COVID-19. In addition, SI and SIS models are also

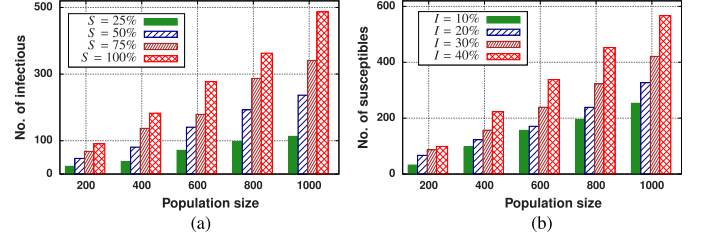


Fig. 10. Impact of population size on (a) infectious and (b) susceptible.

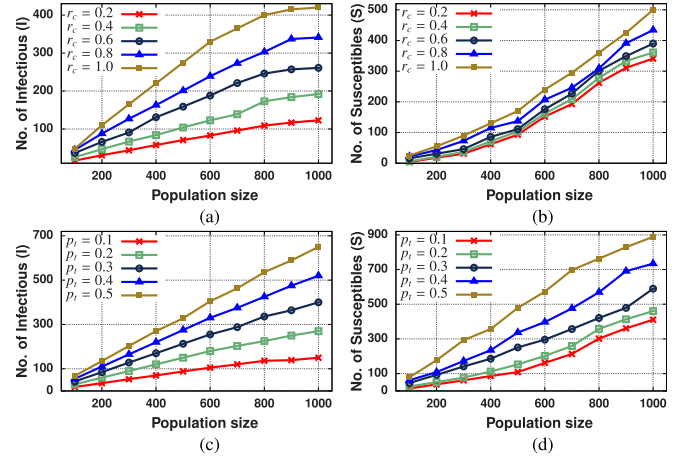


Fig. 11. Impact of rate of contact ( $r_c$ ) and probability of transmission ( $p_t$ ) on a number of infectious and susceptibles. (a) Infectious versus  $r_c$ . (b) Susceptible versus  $r_c$ . (c) Infectious versus  $p_t$ . (d) Susceptible versus  $p_t$ .

available, which have not considered the recovered population that plays an essential role in determining the extent of pandemic's spread. Furthermore, the SIS model assumes that the infected populations are still susceptible. This assumption is not suitable for COVID-19 as there persists an uncertainty of infectious being susceptible.

1) *Impact of Infectious and Susceptible*: At first, this work carried out an empirical analysis to estimate the infectious population when the susceptible increases from 25% to 100%. Here, we fix the rate of contact  $r_c = 1$  and the probability of transmission  $p_t = 0.3$ . Fig. 10(a) shows the increment in infectious with the increase in the % susceptible. If the susceptible reaches 100% on population size 1000, then the infection can be spread to 487 people. Similarly, Fig. 10(b) shows that the increment in the infectious accelerates the growth of susceptible population. Susceptible reaches 564 at the population size of 1000 with 40% infectious. It indicates that the effect of a pandemic (such as COVID-19) can be reduced by decreasing either the susceptible or infectious.

2) *Impact of Rate of Contact and Probability of Transmission*: Next, we study the impact of the rate of contact ( $r_c$ ) and the probability of transmission ( $p_t$ ) on susceptible and infectious with a population size of 1000. First, we consider the probability of transmission as a constant ( $p_t = 0.3$ ) and the rate of contact as a variable. Fig. 11(a) shows that the increment in the infectious is nearly linear with the rate of contact and population size. At higher population size ( $> 800$ ), the curve starts flattening as the infection has covered roughly



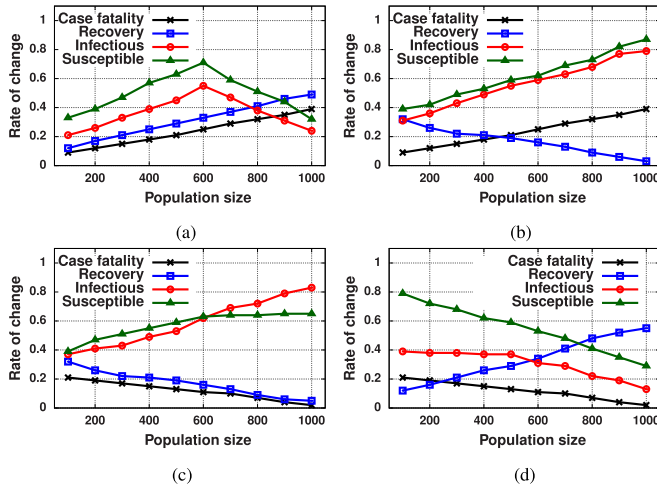


Fig. 12. Impact of case fatality rate and recovery rate on infectious and susceptible. (a) Both case fatality and recovery increase. (b) Increasing case fatality and decreasing recovery. (c) Increasing recovery and decreasing case fatality. (d) Both decrease.

half of the population. Similarly, susceptible increases linearly with increment in the rate of contact.

Furthermore, Fig. 11(c) and (d) shows that the transmission probability is highly sensitive toward the rate of infection spread and increment in the susceptible count. We have considered the rate of contact ( $r_c = 1$ ) for calculating the value of infectious and susceptible. Fig. 11(c) shows that the infectious count is directly proportional to the transmission probability and shows an exact linear increment with the increase in the population size. In a similar pattern, Fig. 11(d) shows that the number of susceptibles increases with an increment in the transmission probability.

3) *Impact of Case Fatality Rate and Recovery Rate*: Finally, we evaluate the impact of the case fatality and recovery rates on the rate of change in infectious and susceptible.

- 1) *Both Case Fatality and Recovery Rates Increase*: Fig. 12(a) shows the variation in infectious and susceptible when both case fatality and recovery rates increase. The result depicts the simultaneous increase in case fatality and recovery rates rectify the number of infectious and susceptible. This increment in infectious and susceptible continued until a thresholding limit, where the case fatality rate's influence starts degrading to that of recovery rate. We also observe from Fig. 12(a) that when the population size is 50, the change in case fatality and the recovery rates are nearly similar. As the population size increases, the recovery rate is more rapid than the case fatality rate, i.e., gap between recovery and case fatality rates is considerable. Furthermore, at the population size 600, infectious and susceptible cover the entire population of 1000 people; thereafter, infectious and susceptible will only decrease.
- 2) *Case Fatality Rate Increase and Recovery Rate Decrease*: It is the most adverse situation for a pandemic, where the case fatality is increasing, and recovery is decreasing. It indicates that the pandemic is in the

worst phase and can infect the entire population in a particular area. Fig. 12(b) shows the adverse situation where both infectious and susceptible are increasing.

- 3) *Both Case Fatality Rate and Recovery Rate Decrease*: It is similar to that of the previously discussed case, but one positive thing that mitigates severe damage is the decrement in fatality rate. Fig. 12(c) shows that the simultaneous decrement in both case fatality and recovery rates increase the infectious and susceptible.
- 4) *Case Fatality Rate Decreases and Recovery Rate Increases*: The last combination is the condition when the case fatality rate decreases and recovery rate increases. It is a favorable situation where there is an indication that the pandemic is ending. Fig. 12(d) shows that both infectious and susceptible decrease under such circumstances.

## V. CONCLUSION

The pandemic COVID-19 is spreading at a higher pace around the globe. The empirical and data analysis in this article tried to figure out the impact of various parameters, such as the probability of transmission, contact rate, infectious, and susceptible populations. The results indicated that the spread of COVID-19 could be analyzed, modeled, and represented in the tabular and graphical format. The data analysis has provided the complete in-site of the coronavirus spread and approximated the reproduction number of COVID-19 between 1.5 and 4 with a low fatality rate. This article has covered popular models for analyzing different aspects of the pandemic and identifies one model's advantage and disadvantage over others. The representations have provided a more natural way to visualize the actual effect of the pandemic spread.

## ACKNOWLEDGMENT

The authors would like to thank to Vinit Kumar Patra for doing the implementation and data preprocessing.

## REFERENCES

- [1] G. F. de Arruda, G. Petri, F. A. Rodrigues, and Y. Moreno, "Impact of the distribution of recovery rates on disease spreading in complex networks," *Phys. Rev. Res.*, vol. 2, no. 1, pp. 013046:1–013046:6, Jan. 2020.
- [2] M.-T. Li *et al.*, "Analysis of COVID-19 transmission in Shanxi province with discrete time imported cases," *Math. Biosci. Eng.*, vol. 17, no. 4, pp. 3710:1–3710:11, 2020.
- [3] P. C. Cooley *et al.*, "The model repository of the models of infectious disease agent study," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 4, pp. 513–522, Jul. 2008.
- [4] Z.-K. Zhang, C. Liu, X.-X. Zhan, X. Lu, C.-X. Zhang, and Y.-C. Zhang, "Dynamics of information diffusion and its applications on complex networks," *Phys. Rep.*, vol. 651, pp. 1–34, Sep. 2016.
- [5] T. Zhou *et al.*, "Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV," *J. Evidence-Based Med.*, vol. 13, no. 1, pp. 3–7, Feb. 2020.
- [6] P. van den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Math. Biosci.*, vol. 180, nos. 1–2, pp. 29–48, Nov. 2002.
- [7] R. Ranjan, "Predictions for COVID-19 outbreak in India using epidemiological models," *medRxiv*, early access, pp. 1–10, 2020, doi: 10.1101/2020.04.02.20051466.
- [8] Z. Zhang, H. Wang, C. Wang, and H. Fang, "Modeling epidemics spreading on social contact networks," *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 3, pp. 410–419, Sep. 2015.

- [9] Z. Wang, C. Xia, Z. Chen, and G. Chen, "Epidemic propagation with positive and negative preventive information in multiplex networks," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1454–1462, Mar. 2021, doi: [10.1109/TCYB.2019.2960605](https://doi.org/10.1109/TCYB.2019.2960605).
- [10] H. Shi, Z. Duan, and G. Chen, "An SIS model with infective medium on complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 387, nos. 8–9, pp. 2133–2144, Mar. 2008.
- [11] *COVID-19 India*. (2021). [Online]. Available: <https://www.covid19india.org/>
- [12] *Pandemic*. (2021). [Online]. Available: <https://theconversation.com/how-to-model-a-pandemic-134187>
- [13] *Four Stages Of Coronavirus Transmission*. (2021). [Online]. Available: <https://www.healthwire.co/four-stages-of-coronavirus-transmission-and-next-phase-of-indias-battle/>
- [14] E. Pardoux, "Deviations from the law of large numbers and extinction of an endemic disease," in *Mathematical Modeling of Random and Deterministic Phenomena*. Wiley, 2020, pp. 1–30, doi: [10.1002/9781119706922.ch1](https://doi.org/10.1002/9781119706922.ch1).
- [15] M. L. Ranney *et al.*, "Correlates of depressive symptoms among at-risk youth presenting to the emergency department," *Gen. Hospital Psychiatry*, vol. 35, no. 5, pp. 537–544, Sep. 2013.
- [16] Z. Ma, J. Liu, and J. Li, "Stability analysis for differential infectivity epidemic models," *Nonlinear Anal., Real World Appl.*, vol. 4, no. 5, pp. 841–856, Dec. 2003.



**Rahul Mishra** (Graduate Student Member, IEEE) received the M.Tech. degree in computer science and engineering from the Madan Mohan Malaviya University of Technology, Gorakhpur, India, in 2017. He is currently pursuing the Ph.D. degree in computer science and engineering with IIT (BHU) Varanasi, Varanasi, India.

His research interests include wireless sensor networks, deep learning, and fog computing.



**Hari Prabhat Gupta** received the Ph.D. degree in computer science and engineering from IIT Guwahati, Guwahati, India, in 2014.

He is currently working as an Assistant Professor at the Department of Computer Science and Engineering, IIT (BHU) Varanasi, Varanasi, India. His research interests include wireless sensor networks, machine learning, and distributed algorithms.



**Tanima Dutta** received the Ph.D. degree in computer science and engineering from IIT Guwahati, Guwahati, India, in 2014.

She is currently working as an Assistant Professor at the Department of Computer Science and Engineering, IIT (BHU) Varanasi, Varanasi, India. Her research interests include multimedia, digital forensics, wireless sensor networks, and algorithms.