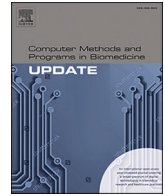




Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine Update

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine-update



A death, infection, and recovery (DIR) model to forecast the COVID-19 spread

Fazila Shams^a, Assad Abbas^{a,*}, Wasiq Khan^b, Umar Shahbaz Khan^c, Raheel Nawaz^d

^a Department of Computer Science, COMSATS University Islamabad, Islamabad, Pakistan

^b Department of Computing and Mathematics, Liverpool John Moores University, Liverpool L3 3AF, United Kingdom

^c National Centre of Robotics and Automation, H-12, Islamabad, Pakistan

^d Department of Operations, Technology, Events and Hospitality Management, Manchester Metropolitan University (MMU), Manchester M156BH, United Kingdom

ARTICLE INFO

Keywords:

COVID-19

Forecasting model

Time-series model, Death rate, DIR model

ABSTRACT

Background: The SARS-Cov-2 virus (commonly known as COVID-19) has resulted in substantial casualties in many countries. The first case of COVID-19 was reported in China towards the end of 2019. Cases started to appear in several other countries (including Pakistan) by February 2020. To analyze the spreading pattern of the disease, several researchers used the Susceptible-Infectious-Recovered (SIR) model. However, the classical SIR model cannot predict the death rate.

Objective: In this article, we present a Death-Infection-Recovery (DIR) model to forecast the virus spread over a window of one (minimum) to fourteen (maximum) days. Our model captures the dynamic behavior of the virus and can assist authorities in making decisions on non-pharmaceutical interventions (NPI), like travel restrictions, lockdowns, etc.

Method: The size of training dataset used was 134 days. The Auto Regressive Integrated Moving Average (ARIMA) model was implemented using XLSTAT (add-in for Microsoft Excel), whereas the SIR and the proposed DIR model was implemented using python programming language. We compared the performance of DIR model with the SIR model and the ARIMA model by computing the Percentage Error and Mean Absolute Percentage Error (MAPE).

Results: Experimental results demonstrate that the maximum% error in predicting the number of deaths, infections, and recoveries for a period of fourteen days using the DIR model is only 2.33%, using ARIMA model is 10.03% and using SIR model is 53.07%.

Conclusion: This percentage of error obtained in forecasting using DIR model is significantly less than the% error of the compared models. Moreover, the MAPE of the DIR model is sufficiently below the two compared models that indicates its effectiveness.

1. Introduction

The Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2) or COVID-19 (coronavirus disease 2019) originated in China around the end of 2019 and rapidly spread worldwide. COVID-19 was declared a Public Health Emergency of International Concern (PHEIC) on 30-th January 2020 by the World Health Organization (WHO) [1]. The first casualty was reported in China on 9-th January 2020 [2]. Considering the rapid spread and increasing casualties in several countries, WHO declared COVID-19 a global pandemic on 11-th March 2020 [3] and announced a worldwide emergency. Many preventive measures were taken by health authorities across the world. Several data mining

approaches were also used to extract the information from social media [4] and news [5] to ensure the timely availability of COVID-19 related information.

Several approaches for studying the virus spread have been reported, including models like Susceptible-Infectious-Recovered (SIR) [6], Auto Regressive Integrated Moving Average (ARIMA) [7] among others. The SIR model was used to predict the peak of the pandemic in Pakistan and Romania [3]. They forecasted the time-to-reach peak in both countries and also forecasted the worse situation in countries if no preventive measures were taken or no strict rules were imposed by the government of the countries. Using the SIR model it was predicted that by 5-th July 2020 around 3000,000 individuals will be infected [8] in Baluchistan,

* Corresponding author at: Department of Computer Science COMSATS University Islamabad, Park Road Islamabad.

E-mail address: assadabbas@comsats.edu.pk (A. Abbas).

<https://doi.org/10.1016/j.cmpbup.2021.100047>

Received 7 June 2021; Received in revised form 14 November 2021; Accepted 27 December 2021

Available online 28 December 2021

2666-9900/Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Pakistan, if lockdown and social distancing measures are either not followed sternly by the public or if the government fails to implement rules that made it mandatory for everyone to abide the SOPs.

In [6], the situation of India under lockdown was studied using the Susceptible-Infected-Removed (SIR) model (considering deaths and recoveries as the single term “removed”). The potential parameters of the SIR model, including the infection rate, the recovery rate, and the reproduction number (R_0) were estimated separately for each phase of lockdown. The value of R_0 calculated for the used dataset is 1.36515. However, the value differs in different phases of lockdown.

Autoregressive Integrated Moving Average (ARIMA) was used in Ref. [7] to predict the daily number of cases in Saudi Arabia. The dataset was first tested using four different models and ARIMA outperformed in that situation. The results from this model suggested that the Umrah and Hajj activities (religious Muslim activities) should be suspended in the country as these events are the gathering of millions of people and can spread the infectious disease more rigorously.

The SIR model was also implemented on the situation of other countries like China, Italy, Russia, Australia, South Korea, India, and the state of Texas in the USA by [9, 10, 11, 12] to find out the rate of infection and rate of recovery in these countries and to validate the classical parameters of the SIR model in this novel disease. A Stochastic SIR model was implemented in Ref. [13] by considering the data of the European region to find out the parameters of the SIR model. They used the calculated parameters to predict the future trend of infection in India during the early days of COVID-19. The first hundred days of COVID-19 were modeled in [14] using the SIR model for countries including China, Greece, Australia, Denmark, France, Germany, Switzerland, Italy, the United States, and Spain. The effective contract rate or rate of infection was considered as an impact of an externally imposed condition and hence, the SIR is modified to Forced SIR (FSIR).

The time-dependent SIR model was used to find out the rate of infection and rate of recovery at a particular time (t) [15]. The results show that less than 3% error was detected in the prediction of one day. Moreover, the infected people have been distributed into two categories, the detectable and the undetectable. As the number of detectable infected people does not describe the statistics of the pandemic, therefore, it was suggested that social distance should be maintained to keep ourselves safe from undetectable infected people.

SIR is an epidemiological model and is used in a closed population setting, assuming that the total population will remain the same till the end of the pandemic. It can compute the theoretical number of infected people, but it cannot compute or predict the number of deaths caused by the viral disease. ARIMA is a time-series model used to predict the future points in the series. It can be used to forecast the total number of cases, recoveries, and deaths caused by the virus. However, it uses the “backward looking” method and is poor at predicting long-term series, as the forecast eventually becomes a straight line. The model lacks an exponential smoothing method, which can severely impact its performance.

In this paper, we present a Death-Infection-Recovery (DIR) time-series model to forecast the number of deaths, infections, and recoveries from this disease. The model uses a triple exponential smoothing method to minimize long-term prediction error. It is important to highlight that our model assumes no cases of re-infection in the population and also, the total population will not remain the same till the end of the pandemic. The model can be used to forecast the virus spread for a minimum of one day and a maximum of fourteen days. The forecasting can help authorities to take measures accordingly by considering the dynamic behavior of the virus. It is important to mention that the research in this paper is restricted to forecast the virus spread in Islamabad; however, the model can be implemented in other cities/regions/countries.

The rest of the paper is organized as follows. Section 2 discusses the existing and proposed method to forecast the COVID-19 spread whereas Section 3 presents the results and discussion while Section 4 concludes

the paper and highlights the directions for future work.

2. Method

This section discusses the method used to forecast the behavior of the virus. The methodology adopted for this study has three main steps as shown in Fig. 1 below:

2.1. Extraction of data

2.1.1. Source of data

Data has been extracted from the official website of Government of Pakistan, administered by National Institute of Health (NIH) Islamabad, Islamabad, Pakistan [16]. The extracted data include daily number of total cases, total deaths and total recoveries for Islamabad region of Pakistan. The website was accessed on 20-th April 2021 for extracting the data.

2.1.2. Predictors

The predictors of proposed DIR model include rate of infection, rate of recovery, and rate of death for one day forecast and for fourteen days forecast, seasonality index and trend factor are also used as predictors of the model.

2.1.3. Outcome

Outcome of model include number of total of deaths, total cases, and total recoveries either for next one day or for next fourteen days.

2.1.4. Sample size

The models are implemented on the dataset of 134 days which includes data of COVID-19 spread in Islamabad from 23-rd November 2020 to 6-th April 2021.

2.1.5. Computer application/program

The already defined equations of the SIR model and the devised equations of the DIR model are implemented in python programming language to get results of forecast, whereas, ARIMA model is implemented using XLSTAT (add-in for Microsoft Excel).

2.2. Implementation of models

2.2.1. SIR model

The SIR model was modified in Ref. [6]. The SIR acronym for Susceptible-Infectious-Removed. The number of recovered and number of deceased are collectively termed as “Removed” from the population. The model is implemented in python programming language on the basis of equations defined to get desirable results.

2.2.2. ARIMA model

ARIMA Model was implemented in Ref. [7]. The forecast using ARIMA model depends on the number of lag observations in the dataset, number of times that the raw observations are differenced, and the size of the moving averages window. The model was implemented using XLSTAT (add-in for Microsoft Excel).

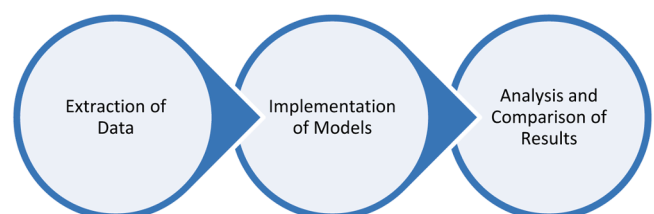


Fig. 1. Methodology adopted.

2.2.3. Proposed death, infection, and recovery (DIR) model

The basic structure of the proposed DIR model is shown in Fig. 2:

According to the DIR Model, a person can either be in one of the three stages namely, (i) infectious, (ii) recovered, or (iii) deceased. The unidirectional arrows in Fig. 1 indicate that the stages of the DIR model are irreversible, which means that once recovered, the person cannot be infected again, and hence there will be no case of re-infection. Moreover, it is self-evident that the deceased person is no longer part of the population; therefore, the total population will not remain the same by the end of the pandemic. If all the cases of coronavirus in the population will be reported accurately then the sum of recovered and deceased must be equal to the total infected people in the population as described in Eq. (1) below:

$$I = R + D \tag{1}$$

where $I, R,$ and D are the total numbers of infected, recovered, and deceased people, respectively. Moreover, $\beta, \gamma,$ and α represent the rate of infection, recovery, and death, respectively and can be calculated by Eq. (2), Eq. (3), and Eq. (4) as given below.

$$\beta = \frac{I^*}{N - D(t - 1) - R(t - 1)} \tag{2}$$

$$\gamma = \frac{R^*}{I^* - D^*} \tag{3}$$

$$\alpha = \frac{D(t)}{I(t)} \tag{4}$$

where N represent the total population and $I(t), R(t),$ and $D(t)$ are the number of infectious, recovered, and deceased, respectively at the time (t) . The values of $I^*, R^*,$ and D^* can be calculated by Eq. (5), Eq. (6), and Eq. (7) as given below.

$$I^* = I(t) - I(t - 1) \tag{5}$$

$$R^* = R(t) - R(t - 1) \tag{6}$$

$$D^* = D(t) - D(t - 1) \tag{7}$$

To predict the number of deceased, infectious, and recovered people at a time (t) using the DIR model, the following equations (Eq. (8), Eq. (9), and Eq. (10)) are used.

$$D(t) = D + D' \tag{8}$$

$$I(t) = I + I' \tag{9}$$

$$R(t) = R + R' \tag{10}$$

where $D, I,$ and R are the total number of deaths, infectious cases, and

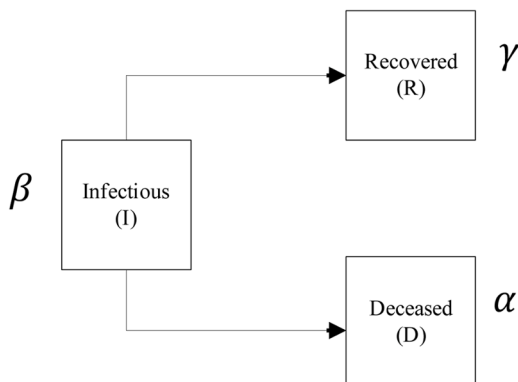


Fig. 2. Proposed DIR Model.

recoveries on the last entry of our dataset and $D', I',$ and R' can be calculated by Eq. (11), Eq. (12), and Eq. (13), as given below.

$$D' = avg(\alpha) * R^o \tag{11}$$

$$I' = I^* + avg(\beta) * I^* \tag{12}$$

$$R' = R^o + avg(\gamma) * R^o \tag{13}$$

Considering the total number of infectious cases, total deaths, and total recoveries obtained from the official website of the Government of Pakistan for the daily update of coronavirus statistics [16] from 23-rd November 2020 to 15-th January 2021, to forecast for 16-th January 2021 for Islamabad is explained below.

To predict the number of deaths, infectious cases, and recoveries at the time (t) , where $t = 16$ th day of January 2021, Eq. (8), Eq. (9), and Eq. (10) are used. The variables used in these equations are calculated first. The total number of deaths, infectious cases and recoveries on 15-th January 2021 in Islamabad as reported by NIH [16] are $D = 450, I = 3988,$ and $R = 37621$ respectively.

Eq. (5) gives the following output.

$$I^* = I(t) - I(t - 1)$$

$$I^* = I(15) - I(14)$$

$$I^* = 39888 - 39749$$

$$I^* = 139$$

Eq. (6) gives the following output.

$$R^* = R(t) - R(t - 1)$$

$$R^* = R(15) - R(14)$$

$$R^* = 37621 - 37413$$

$$R^* = 208$$

Eq. (7) gives the following output.

$$D^* = D(t) - D(t - 1)$$

$$D^* = D(15) - D(14)$$

$$D^* = 450 - 449$$

$$D^* = 1$$

To calculate the value of $\beta, \gamma,$ and $\alpha,$ Eq. (2), Eq. (3), and Eq. (4) are used as described below.

Eq. (2) gives the following output by putting the value of $I^* = 139$ and $N = 1,015,000$ (total population of Islamabad according to the latest census).

$$\beta = \frac{I^*}{N - D(t - 1) - R(t - 1)}$$

$$\beta = \frac{I^*}{N - D(14) - R(14)}$$

$$\beta = \frac{139}{1015000 - 449 - 37413}$$

$$\beta = 0.00142252$$

Eq. (3) gives the following output by putting values of $I^* = 139, R^* = 208,$ and $D^* = 1$

$$\gamma = \frac{R^*}{I^* - D^*}$$

$$\gamma = \frac{208}{139 - 1}$$

$$\gamma = 1.507246377$$

Eq. (4) gives the following output.

$$\alpha = \frac{D(t)}{I(t)}$$

$$\alpha = \frac{D(15)}{I(15)}$$

$$\alpha = \frac{450}{39888}$$

$$\alpha = 0.011281588$$

In the same way, the values of α , β , and γ are calculated for each day in the dataset and their average values are computed. In this example, the average values of the three aforementioned variables as calculated from the dataset are given below:

$$avg(\beta) = 0.00241506$$

$$avg(\gamma) = 1.546798259$$

$$avg(\alpha) = 0.010801059$$

Therefore, Eq. (11) gives,

$$D' = avg(\alpha) * D$$

$$D' = 0.010801059 * 450$$

$$D' = 5$$

Eq. (12) gives,

$$I' = I + avg(\beta) * I$$

$$I' = 139 + 0.00241506 * 139$$

$$I' = 139$$

Eq. (13) gives,

$$R' = R + avg(\gamma) * R$$

$$R' = 208 + 1.546798259 * 208$$

$$R' = 530$$

The forecasted values for 16-th January 2021 are given below.

Eq. (8) gives,

$$D(t) = D + D'$$

$$D(16) = 450 + 5$$

$$D(16) = 455$$

Eq. (9) gives,

$$I(t) = I + I'$$

$$I(16) = 39888 + 139$$

$$I(t) = 40,027$$

Eq. (10) gives,

$$R(t) = R + R'$$

$$R(16) = 37,621 + 530$$

$$R(16) = 38,151$$

The DIR model can be used to calculate the basic reproduction number of the virus spread. The reproduction number (R_o) determines the number of people that can be affected by one infected person. R_o can be calculated by obtaining the ratio of the rate of infection by the rate of recovery and rate of death. R_o can be computed by Eq. (14) as given below:

$$R_o = \frac{avg(\beta)}{avg(\gamma) + avg(\alpha)} \tag{14}$$

Other than reproduction rate, death rate, infection rate, and recovery rate, the trend factor and seasonality index need to be determined to forecast the number of deceased, infectious cases, and recovered people for the next fourteen days at a time (t). The trend factor (T) indicates the increase or decrease in the values of the particular series, whereas the seasonality index \hat{S} indicates the repeating short-term cycle in the series of data. These factors can be calculated by Eq. (15) and Eq. (16), as given below:

$$T = [\delta(T_t - T_{(t-1)})] + [(1 - \delta)T_{(t-1)}] \tag{15}$$

$$\hat{S} = \phi \frac{y_t}{T_t} + (1 - \phi) \hat{S}_{t-1} \tag{16}$$

The forecasting value (F) can be calculated by Eq. (17), as given below.

$$F = (T_t + T_t) \hat{S}_t \tag{17}$$

where T is the smoothed observed value, y is the observation, and δ and ϕ are the coefficients of trends and seasonality, respectively.

To have a rough estimate of virus spread, the DIR model can be used to forecast the spread of the virus for any number of days. However, depending on the sensitivity of decision making, the amount of error obtained from the model in predicting the spread should be the least.

2.3. Analysis and comparison of results

To analyze and compare the results obtained by the models, the Percentage Error and Mean Absolute Percentage Error (MAPE) are used

The percentage error is computed by dividing the absolute value of the difference between actual and predicted values by the actual value and multiplying it by 100 to get the percentage. The formula to compute the percentage error is given below in Eq. (18)

$$Percentage\ Error = \frac{Error}{Actual\ Value} * 100 \tag{18}$$

where,

$$Error = abs(Actual\ value - Predicted\ Value)$$

And “abs” means the absolute value that is ignoring the negative sign.

The formula to calculate the MAPE is given below in Eq. (19)

$$MAPE = \frac{\sum percentage\ error}{n} \tag{19}$$

Where Σ means the Sum of all values and n is the number of observations.

3. Results and discussion

In this section, we present the results and discuss the performance of the DIR model in contrast with the other techniques. The model is used to forecast the total number of deaths, infectious cases, and recoveries

for the next fourteen days using Eq. (17). It is important to mention that the percentage error of the predicted values starts increasing from 2.33% and MAPE starts increasing from 1.74%, if we predict for more than fourteen days, therefore, we restrict to forecast for a maximum of fourteen days. The forecast for the long-term can be beneficial for decision-making authorities; however, forecasting for the minimum number of days can help to understand the dynamic behavior of the virus. The value of R_0 calculated using Eq. (14) is 0.000274717 and the results of forecasted values using the DIR model for the next fourteen days that is 21-st April 2021 – 4-th May 2021 are shown below in Table 1:

However, using Eq. (8), Eq. (9), and Eq. (10), deaths, infectious, and recoveries can be calculated at any time (t) as explained in Section 2 above. This provides a prediction for a short time and is much more recommended for individuals seeking daily forecast by considering the dynamic behavior of the virus. In the long term, certain natural/uncontrolled parameters may change and can cause an error in our predictions and also cannot explain the dynamic behavior of the virus.

The results of the DIR model were compared with the results of the SIR and ARIMA models. Table 2 shows the actual and predicted cases whereas Table 3 shows the actual and predicted deaths. Similarly, Table 4 shows the actual and predicted recoveries along with their% error and MAPE computed for the predicted results of the DIR model for the Islamabad region of Pakistan.

The classical SIR model assumes that the total population will remain the same till the end of the pandemic. However, it has been observed that the virus is causing a great number of deaths; therefore, the total population will not remain the same. On the other hand, the DIR model considers the daily updated total population for calculating each next day's rate of virus spread.

Using the data of total cases, total recoveries, and total deaths from 23-rd November 2020 to 6-th April 2021 considering, "deaths" and "recoveries" as "removed" in the SIR model [6], the rate of infection and rate of recoveries are calculated and the predicted values for next fourteen days are shown in Table 5 and Table 6 below. Table 5 shows the actual and predicted cases, Table 6 shows actual and predicted recoveries along with their% error and MAPE computed for the predicted results of the SIR model for the Islamabad region.

The graphs in Figs. 3 and 4 below show the comparison of the DIR and SIR model for total cases and total recoveries. The X-axis represents the number of days and the y-axis represents total cases on a scale of thousand. Fig. 3 shows the comparison of predicted values for total cases computed using SIR and DIR models along with the actual values, Fig. 4 shows the comparison of predicted recoveries using SIR and DIR models along with the actual values for the Islamabad region.

The ARIMA model is a time series model used to predict the future trend. The model can be used to predict the future trend of coronavirus spread in the region but it seems that the% error generated by ARIMA is higher than that of the DIR model. Table 7 shows the actual and

Table 1
Forecasting for next fourteen days using DIR model in Islamabad.

Days	Date	Total cases	Total deaths	Total Recoveries
1	2021-04-21	71,399	653	57,986
2	2021-04-22	71,813	657	58,325
3	2021-04-23	72,227	661	58,664
4	2021-04-24	72,641	665	59,004
5	2021-04-25	73,055	670	59,343
6	2021-04-26	73,469	674	59,682
7	2021-04-27	73,884	678	60,021
8	2021-04-28	74,298	682	60,360
9	2021-04-29	74,712	686	60,699
10	2021-04-30	75,126	690	61,982
11	2021-05-01	75,540	694	62,659
12	2021-05-02	75,955	698	62,986
13	2021-05-03	76,369	699	63,051
14	2021-05-04	76,783	700	63,420

Table 2
Performance of DIR model during the prediction of total cases.

Days	Date	Predicted cases	Actual cases	Error	% Error
1	07/04/2021	63,357	63,499	142	0.2239%
2	08/04/2021	63,930	64,173	243	0.3782%
3	09/04/2021	64,504	64,902	398	0.6136%
4	10/04/2021	65,077	65,700	623	0.9479%
5	11/04/2021	65,651	66,380	729	1.0987%
6	12/04/2021	66,224	66,983	759	1.1329%
7	13/04/2021	66,798	67,491	693	1.0273%
8	14/04/2021	67,371	68,066	695	1.0209%
9	15/04/2021	67,945	68,665	720	1.0492%
10	16/04/2021	68,518	68,906	388	0.5630%
11	17/04/2021	69,091	69,556	465	0.6678%
12	18/04/2021	69,665	70,079	414	0.5908%
13	19/04/2021	70,238	70,609	371	0.5248%
14	20/04/2021	70,812	70,984	172	0.2425%
Sum of% Error					10.0814%
MAPE					0.7201%

Table 3
Performance of DIR model during the prediction of total deaths.

Days	Date	Predicted deaths	Actual deaths	Error	% Error
1	07/04/2021	591	591	0	0.0301%
2	08/04/2021	594	597	3	0.4424%
3	09/04/2021	598	601	3	0.5757%
4	10/04/2021	601	607	6	1.0345%
5	11/04/2021	604	611	7	1.1617%
6	12/04/2021	607	617	10	1.6073%
7	13/04/2021	611	619	8	1.2924%
8	14/04/2021	618	625	7	1.1200%
9	15/04/2021	621	626	5	0.7987%
10	16/04/2021	625	631	6	0.9509%
11	17/04/2021	630	636	6	0.9434%
12	18/04/2021	635	642	7	1.0903%
13	19/04/2021	641	645	4	0.6202%
14	20/04/2021	647	649	2	0.3082%
Sum of%Error					11.9758%
MAPE					0.8554%

predicted cases whereas Table 8 shows the actual and predicted deaths. Similarly, Table 9 shows the actual and predicted recoveries along with their% error and MAPE computed for the predicted results of the ARIMA model for the Islamabad region of Pakistan.

The graphs in Figs. 5, 6, and 7 show the comparison of results from the ARIMA and DIR models. The X-axis represents the number of days and the Y-axis represents total cases on a scale of thousand. Fig. 5 shows the comparison of predicted cases using ARIMA and DIR models along with the actual values, Fig. 6 shows the comparison of predicted deaths using ARIMA and DIR models along with actual values, and Fig. 7 shows the comparison of predicted recoveries using ARIMA and DIR models along with the actual values for the Islamabad region.

Several researchers modified the SIR model to minimize its limitations, but the results were not up to the mark. In some researches, researchers consider the death and recoveries as a single unit and termed it as removed from the disease [6], but it was not much practical approach as it certainly will not help to predict the deaths caused by the pandemic. The ARIMA model can be possibly used to forecast the total cases, total recoveries, as well as total deaths for coming days as it is a time-series model that uses the combination of automatic regression and moving averages, but the results obtained from the model were having a high percentage of error as the model uses the backward-looking method to forecast a long series and eventually, the graph becomes a straight line. The straight-line prediction is not highly useful when the virus has a dynamic behavior. The DIR model is designed to forecast the deaths, infections, and recoveries with the minimum percentage of error and hence minimum MAPE. The rate of infection, rate of recovery, and rate of death are updated daily to cover the dynamic performance of the virus

Table 4
Performance of DIR model during the prediction of total recoveries.

Days	Date	Predicted recoveries	Actual recoveries	Error	% Error
1	07/04/2021	50,313	50,530	217	0.4291%
2	08/04/2021	50,562	50,722	160	0.3157%
3	09/04/2021	50,811	51,238	427	0.8342%
4	10/04/2021	51,078	52,298	1220	2.3322%
5	11/04/2021	52,108	52,904	796	1.5046%
6	12/04/2021	52,357	53,441	1084	2.0290%
7	13/04/2021	52,805	53,961	1156	2.1415%
8	14/04/2021	53,454	54,602	1148	2.1023%
9	15/04/2021	54,103	55,222	1119	2.0267%
10	16/04/2021	54,552	55,828	1276	2.2865%
11	17/04/2021	55,300	56,399	1099	1.9482%
12	18/04/2021	55,549	56,805	1256	2.2112%
13	19/04/2021	55,998	57,316	1318	2.3002%
14	20/04/2021	56,546	57,631	1085	1.8821%
Sum of% Error					24.3434%
MAPE					1.7388%

Table 5
Performance of SIR model during the prediction of total cases.

Days	Date	Predicted cases	Actual cases	Error	% Error
1	07/04/2021	63,778	63,499	-279	0.4392%
2	08/04/2021	64,763	64,173	-590	0.9196%
3	09/04/2021	65,731	64,902	-829	1.2774%
4	10/04/2021	66,682	65,700	-982	1.4948%
5	11/04/2021	67,617	66,380	-1237	1.8628%
6	12/04/2021	68,535	66,983	-1552	2.3165%
7	13/04/2021	69,437	67,491	-1946	2.8832%
8	14/04/2021	70,324	68,066	-2258	3.3167%
9	15/04/2021	71,195	68,665	-2530	3.6844%
10	16/04/2021	72,051	68,906	-3145	4.5646%
11	17/04/2021	72,893	69,556	-3337	4.7975%
12	18/04/2021	73,720	70,079	-3641	5.1959%
13	19/04/2021	74,533	70,609	-3924	5.5579%
14	20/04/2021	75,333	70,984	-4349	6.1262%
Sum of% Error					44.4367%
MAPE					3.1741%

and to have a more accurate prediction. Moreover, to forecast for the long-term, the trend in the data and seasonality were also calculated to exponentially smooth the error in the dataset. The maximum% error obtained from the DIR model in predicting the total cases, total deaths, and total recoveries of Islamabad for fourteen days is 1.13%, 1.60%, 2.33%, respectively. The MAPE obtained in predicting total cases, total deaths, and total recoveries using the DIR model is 0.72%, 0.86%, and 1.74%, respectively. Using the ARIMA model, the maximum% error obtained in predicting total cases is 8.15%, predicting total deaths is 6.09%, and predicting total recoveries is 10.03%. The MAPE obtained using the ARIMA model in predicting total cases, total deaths, and total recoveries is 4.50%, 2.93%, and 5.20%, respectively. The maximum% error in predicting the total number of cases and recoveries using the SIR model is 6.13%, and 53.07% with 3.17%, and 26.67% MAPE, respectively. Based on the above discussion, we can say that the DIR model outperforms the SIR and ARIMA models.

Table 6
Performance of SIR model during the prediction of total recoveries.

Days	Date	Predicted recoveries	Actual recoveries	Error	% Error
1	07/04/2021	52,610	50,530	-2080	4.1173%
2	08/04/2021	54,745	50,722	-4023	7.9308%
3	09/04/2021	56,965	51,238	-5727	11.1780%
4	10/04/2021	59,276	52,298	-6978	13.3432%
5	11/04/2021	61,681	52,904	-8777	16.5900%
6	12/04/2021	64,183	53,441	-10,742	20.1005%
7	13/04/2021	66,786	53,961	-12,825	23.7681%
8	14/04/2021	69,496	54,602	-14,894	27.2769%
9	15/04/2021	72,315	55,222	-17,093	30.9530%
10	16/04/2021	75,248	55,828	-19,420	34.7860%
11	17/04/2021	78,301	56,399	-21,902	38.8337%
12	18/04/2021	81,477	56,805	-24,672	43.4330%
13	19/04/2021	84,782	57,316	-27,466	47.9207%
14	20/04/2021	88,221	57,631	-30,590	53.0799%
Sum of% Error					373.3110%
MAPE					26.6651%

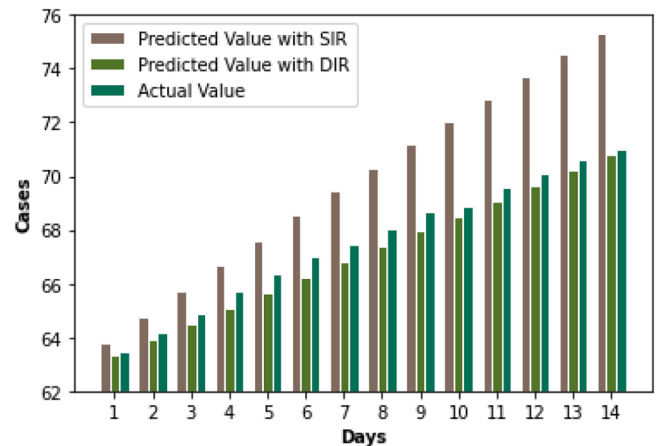


Fig. 3. Prediction for total cases using SIR and DIR model in Islamabad.

The DIR model is limited to consider that there is no case of reinfection in the population, however, total population will not remain the same till the end of pandemic. The model is used to forecast the number of deaths, infectious and recoveries not more than fourteen days. Moreover, the DIR model is also applied to all other regions of Pakistan including Punjab, Sindh, Baluchistan, Khyber Pakhtun Khwa, Azad Jammu and Kashmir, and Gilgit Baltistan to forecast the behavior of virus spread but due to restricted space, only the results of forecast for Islamabad region is discussed and compared with the other models in this study. The daily updates of forecasting in Pakistan for the next fourteen days along with the error in the previous predictions are available on the website [17].

4. Conclusions and future work

In this paper, we presented the DIR model, as a time-series model to

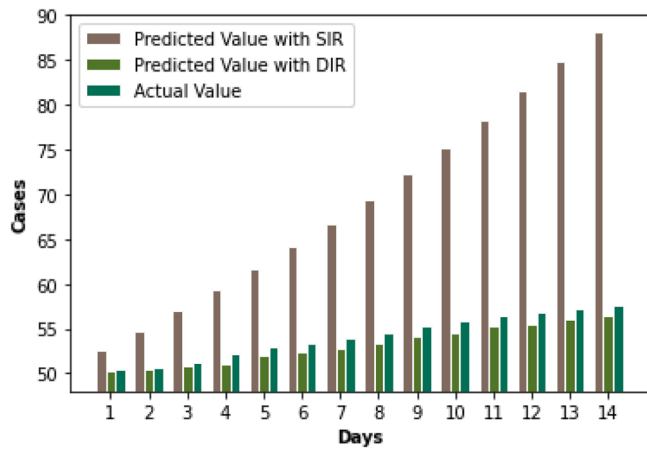


Fig. 4. Prediction for total recoveries using SIR and DIR model in Islamabad.

Table 7 Performance of ARIMA model during the prediction of total cases.

Days	Date	Predicted cases	Actual cases	Error	% Error
1	07/04/2021	63,427	63,499	72	0.1135%
2	08/04/2021	63,696	64,173	477	0.7432%
3	09/04/2021	63,902	64,902	1000	1.5404%
4	10/04/2021	64,076	65,700	1624	2.4721%
5	11/04/2021	64,229	66,380	2151	3.2411%
6	12/04/2021	64,366	66,983	2617	3.9062%
7	13/04/2021	64,493	67,491	2998	4.4418%
8	14/04/2021	64,611	68,066	3455	5.0759%
9	15/04/2021	64,722	68,665	3943	5.7429%
10	16/04/2021	64,826	68,906	4080	5.9209%
11	17/04/2021	64,925	69,556	4631	6.6573%
12	18/04/2021	65,020	70,079	5059	7.2185%
13	19/04/2021	65,111	70,609	5498	7.7862%
14	20/04/2021	65,199	70,984	5785	8.1503%
Sum of% Error MAPE					63.0104%
					4.5007%

Table 8 Performance of ARIMA model during the prediction of total deaths.

Days	Date	Predicted deaths	Actual deaths	Error	% Error
1	07/04/2021	594	591	-3	0.4707%
2	08/04/2021	596	597	1	0.1391%
3	09/04/2021	598	601	3	0.4993%
4	10/04/2021	600	607	7	1.2291%
5	11/04/2021	601	611	10	1.6539%
6	12/04/2021	602	617	15	2.4118%
7	13/04/2021	603	619	16	2.5454%
8	14/04/2021	604	625	21	3.3136%
9	15/04/2021	605	626	21	3.3112%
10	16/04/2021	606	631	25	3.9303%
11	17/04/2021	607	636	29	4.5469%
12	18/04/2021	608	642	34	5.3078%
13	19/04/2021	609	645	36	5.6231%
14	20/04/2021	610	649	39	6.0851%
Sum of% Error MAPE					41.0671
					2.9334%

forecast the virus spread for the next fourteen days. The results obtained from the DIR model were compared with the SIR model and the ARIMA model. Experimental results demonstrate that the DIR model outperforms both the SIR and the ARIMA models in all the three parameters of virus spread namely, the rate of infection, rate of recovery, and rate of death. The maximum% error obtained by the DIR model in all the three parameters is 2.33% with 1.74% MAPE, whereas the maximum error in predicting the virus spread using the SIR and ARIMA model is 53.07% and 10.03% with 26.67% and 5.20% MAPE, respectively. The model can

Table 9 Performance of ARIMA model during the prediction of total recoveries.

Days	Date	Predicted recoveries	Actual recoveries	Error	% Error
1	07/04/2021	50,543	50,530	-13	0.0256%
2	08/04/2021	50,741	50,722	-19	0.0369%
3	09/04/2021	50,892	51,238	346	0.6748%
4	10/04/2021	51,020	52,298	1278	2.4440%
5	11/04/2021	51,132	52,904	1772	3.3493%
6	12/04/2021	51,234	53,441	2207	4.1307%
7	13/04/2021	51,327	53,961	2634	4.8819%
8	14/04/2021	51,413	54,602	3189	5.8398%
9	15/04/2021	51,495	55,222	3727	6.7498%
10	16/04/2021	51,572	55,828	4256	7.6243%
11	17/04/2021	51,645	56,399	4754	8.4300%
12	18/04/2021	51,714	56,805	5091	8.9617%
13	19/04/2021	51,781	57,316	5535	9.6567%
14	20/04/2021	51,845	57,631	5786	10.0389%
Sum of% Error MAPE					72.8445%
					5.2032%

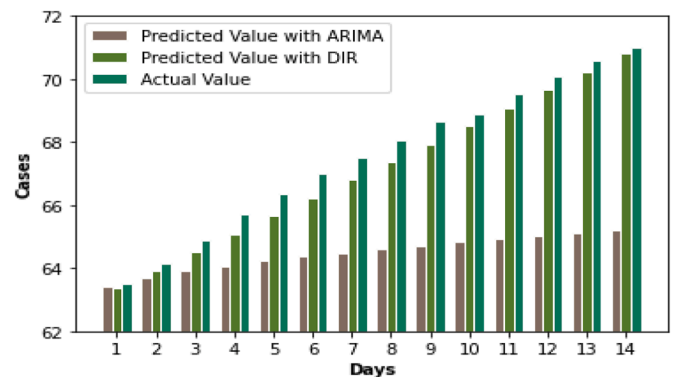


Fig. 5. Prediction for total cases using ARIMA and DIR model in Islamabad.

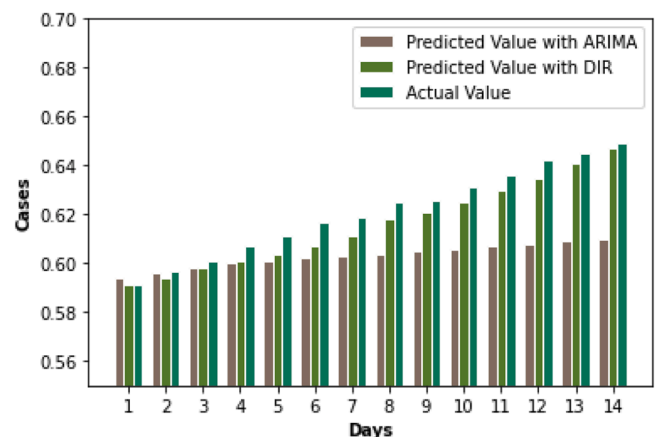


Fig. 6. Prediction for total deaths using ARIMA and DIR model in Islamabad.

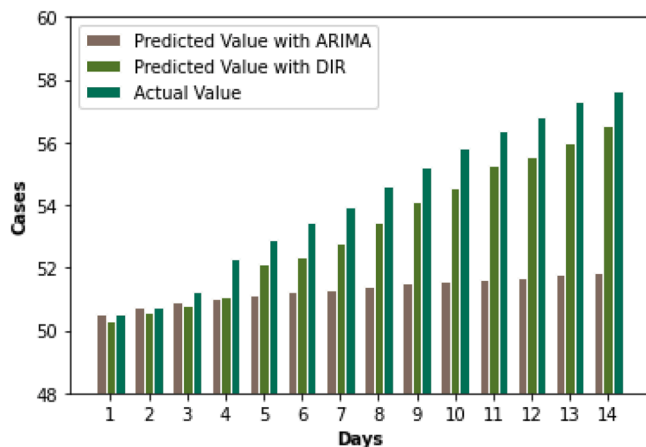


Fig. 7. Prediction for total recoveries using ARIMA and DIR model in Islamabad.

be used in other cities/regions/countries to study the dynamic behavior of the virus by forecasting the spread for a minimum of one day or can be used to help governments in planning non-pharmaceutical interventions like social and travel restrictions and lockdowns etc. The model can be improved by incorporating other parameters, such as weather conditions and herd immunity indicators.

Declaration of Competing Interest

The authors declare that there is no conflict of interest in any aspect.

Acknowledgement

We are thankful to the National Institute of Health Islamabad, Pakistan for providing the access to the daily data about COVID-19 cases to carry out this research.

References

- [1] F. Rahal, S. Rezak, F. Zohra, and B. Hamed, "Impact of meteorological parameters on the Covid-19 incidence. The case of the city of Oran, Algeria," vol. 19, Jan. 2020.
- [2] R. Salgotra, M. Gandomi, A. G., "Evolutionary modelling of the COVID-19 pandemic in fifteen most affected countries," *Chaos, Solit. Fractals*, vol. 140, p. 110118, Nov, 2020.
- [3] M. Ozair, T. Hussain, M. Hussain, A.U. Awan, D. Baleanu, K.A. Abro, "A mathematical and statistical estimation of potential transmission and severity of COVID-19: a combined study of Romania and Pakistan," *Biomed. Res. Int.* 2020 (2020). Dec.
- [4] A. Said, T.D. Bowman, R.A. Abbasi, N.R. Aljohani, S.U. Hassan, R. Nawaz, "Mining network-level properties of Twitter altmetrics data," *Scientometrics* 120 (1) (2019) 217–235. Jul.
- [5] P. Thompson, R. Nawaz, I. Korkontzelos, W. Black, J. McNaught, and S. Ananiadou, "News search using discourse analytics," 2013 Digital Heritage International Congress (DigitalHeritage), vol. 1, pp. 597–604, 2013.
- [6] D.K. Bagal, A. Rath, A. Barua, D. Patnaik, "Estimating the parameters of susceptible-infected-recovered model of COVID-19 cases in India during lockdown periods," *Chaos, Solit. Fractals* 140 (2020), 110154.
- [7] S. Alzahrani, I. Aljamaan, and E. A. Fakhri, "Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions," *J. Infect. Public Health*, vol. 13, no. 7, pp. 914–919, 2020.
- [8] M. Arif, "Estimation of the final size of the COVID-19 epidemic in Balochistan Province, Pakistan," *Int. J. Front. Sci.*, vol. 4, no. 2, pp. 78–80, June. 2020.
- [9] N.A. Kudryashov, M.A. Chmykhov, M. Vigdorowitsch, "Analytical features of the SIR model and their applications to COVID-19," *Appl. Math. Model.* 90 (2021) 466–473.
- [10] I. Cooper, A. Mondal, C.G. Antonopoulos, "A SIR model assumption for the spread of COVID-19 in different communities," *Chaos, Solit. Fractals* 139 (2020), 110057.
- [11] G.C. Calafiore, C. Novara, C. Possieri, "A time-varying SIRD model for the COVID-19 contagion in Italy," *Annu. Rev. Control* 50 (2020) 361.
- [12] J. Huang, L. Zhang, X. Liu, Y. Wei, C. Liu, X. Lian, Z. Huang, J. Chou, X. Liu, X. Li, and K. Yang, "Global prediction system for COVID-19 pandemic," *Sci. bull.*, Aug. 2020.

- [13] A. Simha, R.V. Prasad, and S. Narayana, "A simple stochastic SIR model for COVID 19 infection dynamics for Karnataka: learning from Europe," Mar. 2020.
- [14] E. Kaxiras, G. Neofotistos, E. Angelaki, "The first 100 days: modeling the evolution of the COVID-19 pandemic," *Chaos, Solit. Fractals* 138 (2020), 110114.
- [15] Y. Chen, P. Lu, C. Chang, T. Liu, "A time-dependent SIR model for COVID-19 with undetectable infected persons," *IEEE Trans. Netw. Sci. Eng.* 7 (4) (2020) 3279–3294. Sep.
- [16] "COVID-19 health advisory platform by Ministry of National Health Services Regulations and Coordination." <https://covid.gov.pk/> (accessed Apr. 20, 2021).
- [17] "COVID-19 forecasting in Pakistan." <https://covid-19forecasting.000webhostapp.com> (accessed Apr. 20, 2021).



Fazila Shams is a graduate student of the Computer Science department at the COMSATS University Islamabad, Islamabad, Pakistan and is perusing her Masters in Software Engineering. She has done her Bachelors in Software Engineering from Bahria University Islamabad, Islamabad, Pakistan. She can be reached at the e-mail: fazila.malikshams@gmail.com



Assad Abbas received Ph.D. in Electrical and Computer Engineering from North Dakota State University, USA. Currently, he is working as an Assistant professor of computer science at COMSATS University Islamabad, Islamabad Pakistan. His research interests are mainly but not limited to Smart Health, Big Data Analytics, Recommendation Systems, Patent Analysis, Software Engineering, and Social Network Analysis. Moreover, his research has appeared in several reputable international venues. He is also serving as the referee for numerous prestigious journals and as the technical program committee member for several conferences. Moreover, he is a member of IEEE and IEEE-HKN and a Professional Member of the ACM. He can be reached at the e-mail: assadabbas@comsats.edu.pk



Wasiq Khan is a Senior Lecturer in Artificial Intelligence & Data Sciences within the Department of Computer Science at Liverpool John Moores University, UK. Wasiq received his B.Sc. in Mathematics, Physics, and Geography following an M.Sc. in Computer Science from Pakistan. At Bradford University UK, He received an M.Sc. in Artificial Intelligence for Board Games following a Ph.D. in Speech Processing & Intelligent Reasoning and a Post Graduate Certificate in Teaching & Learning in Higher Education (PGCHEP).



Umar Shahbaz Khan is currently working in the field of robotics and mainly towards the development of myo-electric controlled prosthesis. In 2016 he secured funding for this project from Ignite for an amount of PKR 14.12 Mn, the project is near the completion phase. His areas of interests include robotics, Mechatronics, embedded systems and image processing. Currently, he is the head of the Mechatronics Engineering Department at NUST College of Electrical and Mechanical Engineering.



Raheel Nawaz is currently the Director of Digital Technology Solutions and a Reader in analytics and digital education with Manchester Metropolitan University (MMU). He has founded and/or headed several research units specializing in artificial intelligence, data science, digital transformations, digital education, and apprenticeships in higher education. He has led numerous funded research projects in U.K., EU, South Asia, and the Middle East. He has held adjunct or honorary positions with many research, higher education, and policy organizations, both in the U.K., and overseas. He regularly makes media appearances and speaks on a range of topics, especially artificial intelligence and higher education. Before becoming a full-time academic, he served in various senior leadership positions in the private higher and further education sector; and was an Army Officer before that.