



# Prediction of COVID-19 Data Using Hybrid Modeling Approaches

Weiping Zhao<sup>1</sup>, Yunpeng Sun<sup>2\*</sup>, Ying Li<sup>2\*</sup> and Weimin Guan<sup>2\*</sup>

<sup>1</sup> School of Asian Languages, Zhejiang Yuexiu University of Foreign Language, Shaoxing, China, <sup>2</sup> School of Economics, Tianjin University of Commerce, Tianjin, China

## OPEN ACCESS

### Edited by:

Mihajlo Jakovljevic,  
Hosei University, Japan

### Reviewed by:

Vena Pearl Bongolan,  
University of the Philippines  
Diliman, Philippines  
Dairi Abdelkader,  
Oran University of Science and  
Technology - Mohamed  
Boudiaf, Algeria  
Seyed Amir Abbas Oloomi,  
Islamic Azad University, Iran

### \*Correspondence:

Yunpeng Sun  
tjwade3@126.com  
Ying Li  
yingzij2006@126.com  
Weimin Guan  
weibe89@163.com

### Specialty section:

This article was submitted to  
Health Economics,  
a section of the journal  
Frontiers in Public Health

Received: 20 April 2022

Accepted: 23 May 2022

Published: 22 July 2022

### Citation:

Zhao W, Sun Y, Li Y and Guan W  
(2022) Prediction of COVID-19 Data  
Using Hybrid Modeling Approaches.  
*Front. Public Health* 10:923978.  
doi: 10.3389/fpubh.2022.923978

A major emphasis is the dissemination of COVID-19 across the country's many regions and provinces. Using the present COVID-19 pandemic as a guide, the researchers suggest a hybrid model architecture for analyzing and optimizing COVID-19 data during the complete country. The analysis of COVID-19's exploration and death rate uses an ARIMA model with susceptible-infectious-removed and susceptible-exposed-infectious-removed (SEIR) models. The logistic model's failure to forecast the number of confirmed diagnoses and the snags of the SEIR model's too many tuning parameters are both addressed by a hybrid model method. Logistic regression (LR), Autoregressive Integrated Moving Average Model (ARIMA), support vector regression (SVR), multilayer perceptron (MLP), Recurrent Neural Networks (RNN), Gate Recurrent Unit (GRU), and long short-term memory (LSTM) are utilized for the same purpose. Root mean square error, mean absolute error, and mean absolute percentage error are used to show these models. New COVID-19 cases, the number of quarantines, mortality rates, and the deployment of public self-protection measures to reduce the epidemic are all outlined in the study's findings. Government officials can use the findings to guide future illness prevention and control choices.

**Keywords:** hybrid modeling approach, ARIMA models, SIR and SIER models, COVID-19, Pakistan

## INTRODUCTION

The economy of Pakistan, home to over 197 million people, is rated as being of moderate to medium complexity. Since the first case in Pakistan in February 2020, the country has been on high alert. Government officials and healthcare professionals recommend preventive steps to limit the disease's spread. There are four provinces and three regions in Pakistan: Punjab, Sindh, Khyber Pakhtunkhwa (KPK), Baluchistan, Gilgit-Baltistan, and Azad Jammu & Kashmir (AJK) (1). The first two instances of COVID-19 in Pakistan sounded the warning bell on February 26, 2020 (2). Determine the effects of social isolation remains a pressing issue, and the worldwide health catastrophe of COVID-19 has highlighted the importance of research and scientific progress in the face of this global health epidemic. In addition to the evident deaths and illnesses, the epidemic has caused emotional anguish and stress in the residents, who have been under lockdown and quarantine for the past few months (3). LR, ARIMA, SVR, MLP, RNN, GRU, and long short-term memory (LSTM) are utilized for the same purpose. The root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are used to show these models, and there are a variety of mathematical approaches to the susceptible-exposed-infectious-removed (SEIR) model (4). Health facilities, transportation, and strong economic activity or densely occupied migratory workers might all affect the relevance of infectiousness in a district's population mobility (5). The logistic growth model (6), stochastic susceptible-infectious-removed (SIR) model (7), and SEIR model are used to investigate and predict the future trends of COVID-19 (8). According to Pakistan's



COVID-19 statistics, men accounted for 74% of cases in June 2020. However, this might be due to a bias in testing (9). The study consists of five sections. The first section describes the introduction, and section literature review consists of the literature review to describe the supporting references. Section research methodology determined the methodology and research design. Section results and discussions explains the results in the figures and tables, and section conclusions concludes with the results and analysis of the study.

## LITERATURE REVIEW

Infections of the respiratory tract are the most frequent globally, resulting in a high mortality rate and a substantial financial load on healthcare systems (10). Mobile workforce prioritization reduces the spread of diseases by 24.14% compared to other situations. Quezon City's population is best protected by preferring the elderly (439%) over other alternatives (11). The COVID-19 virus has been implicated in an epidemic of atypical pneumonia (12), and the virus's human-to-human transmissibility has been documented domestically and globally (13). Pakistan has seen a comparatively low number of infections and fatalities from the COVID-19 epidemic. There are 585,435 cases and 13,076 fatalities in the country as of March 4, 2021 (14). The mathematical modeling and forecasting of epidemics have been resurrected after COVID-19's outbreak are developing new prediction algorithms to track the spread of the disease and foretell its future course (15, 16). However, quarantine is more effective at restricting the transmission of the virus, a control plan that maximizes and is likely to be successful. For release from quarantine, continuous, and stringent compliance with prescribed social and health-enhancing activities is required (17). Since no vaccine has been created for infectious diseases, such as COVID-19, vaccinations cannot be used to treat them. A limited supply of vaccines controlling the spread of COVID-19 requires a focus on early detection and prevention (18). COVID-19 many prediction models are currently available, including the logistic regression analysis LSTM, ARIMA, and TBATS models (19). Alzahrani et al. (20) described that the ARIMA prediction model is used to propagate pandemic outbreaks in Saudi Arabia. COVID-19 lockdown effectiveness

is examined using the age-structured variant of the SEIR-type for pandemics (21, 22). The COVID-19 epidemic dynamics are predicted using a modified SEIR compartmental model (23). The dynamic process of epidemic spread is described using a simple SEIR model developed (24). The average age of the people of Quezon City is significantly lower than that of the most vulnerable demographic (22). COVID-19's spread may be predicted using SEIR's mathematical modeling. The SEIR model is utilized to build the model (25). The model study employs the generation matrix approach (26) to get the fundamental reproduction number and the global stability for COVID-19 spreading by incorporating vaccination and isolation variables as model parameters. The number of COVID-19 cases in Pakistan is simulated using secondary data (27). Compared to the Q-SEIR model, the infection rates predicted by the age-stratified probability are significantly lower. Quarantine delays peaks for exposed and infected groups and "flattens" the curve or lowers each compartment's projected value (28). The effectiveness of various actions following the outbreak may be evaluated (29), which appears to be a daunting problem for general statistical approaches. As a general rule, the SIR paradigm is commonly used to describe the progression of an illness from one person to another through three distinct but incompatible stages (15). The long-term dependencies in huge sequences of hundreds or thousands of steps are the fundamental restriction of RNN learning. LSTM networks can overcome these restrictions (30). Precautions include immunizing pupils first, mandating mask use, and maintaining a physical distance of at least 25 feet between students and teachers (31). Chinese forecasters employed LSTM for COVID-19 forecasting and found it more accurate than the dynamic SEIR model (32). The system uses a SIR model and an LSTM to depict the current state of the pandemic and estimate its course into the future with reasonable accuracy (33). The RMSE calculates the difference between the actual and fitted data, while the R shows the relationship between the model's output and the actual data (34). LSTM networks, polynomial neural networks, and neural networks may all be used to anticipate COVID-19 situations (35).

## RESEARCH METHODOLOGY

### Data Collection

The research data is collected from March 1, 2020 to September 30, 2020, and the number of confirmed COVID-19 cases is utilized to conduct the investigation using SIR and SEIR models.

**Abbreviations:** MSE, Mean squared error; NMSE, Normalized mean squared error; MAE, Mean absolute error; MAPE, Mean absolute percentage error.

**TABLE 1** | Validation matrix for COVID-19 confirm case (actual vs. predicted) without temperature.

Localities	Methods	RMSE	MAE	MAPE
Punjab	ARIMA	20.176	407.102	0.025101
	LR	19.49	380.011	0.023831
	SVR	18.17	330.19	0.026613
	MLP	17.35	301.181	0.020162
	RNN	14.5	210.451	0.019541
	GRU	13.44	180.178	0.01861
	LSTM	12.57	141.91	0.018111
Sindh	ARIMA	20.176	405.102	0.026101
	LR	20.49	380.011	0.028831
	SVR	19.17	330.19	0.027613
	MLP	18.35	307.181	0.020162
	RNN	15.53	210.451	0.019541
	GRU	14.41	180.178	0.01871
	LSTM	12.15	155.91	0.018511
KPK	ARIMA	20.76	409.102	0.027101
	LR	19.49	380.011	0.023831
	SVR	18.72	330.19	0.026613
	MLP	17.53	301.181	0.020162
	RNN	15.5	217.451	0.019541
	GRU	14.44	176.781	0.01891
	LSTM	13.51	146.91	0.017911
Baluchistan	ARIMA	20.176	407.102	0.027101
	LR	19.49	395.011	0.026131
	SVR	18.17	331.19	0.025613
	MLP	17.35	301.181	0.020162
	RNN	14.5	210.451	0.019541
	GRU	13.44	180.178	0.01891
	LSTM	11.65	135.91	0.017911
Azad Jammu & Kashmir	ARIMA	19.1	407.102	0.025101
	LR	18.49	380.011	0.024831
	SVR	17.11	330.19	0.026713
	MLP	16.35	305.181	0.0199162
	RNN	13.51	208.451	0.018941
	GRU	12.44	181.178	0.01881
	LSTM	11.57	138.91	0.018011

The COVID-19 cases data of Pakistan is collected from the WHO <https://www.who.int/.com>. The data relating to several total tests confirmed cases, critical cases, death cases, quarantine, and fully vaccinated persons in all provinces in Punjab, Sindh, Baluchistan, KPK, and Jammu & Kashmir is collected from the National Conference on Citizenship (NCO) Govt. of Pakistan website.”

### Research Design

Rao et al. (36) propose a non-linear incidence rate and the recovery rate in the dimensional model discussed. Hospital bed-population ratio “ $a_1 > 0$ ” and contaminated “ $I$ ” are both factors that affect the patient’s ability to recover; according to the healing

rate (37).

$$b = b_0 + \frac{(b_1 - b_0) a_1}{1 - a_1} \tag{1}$$

The maximum and minimum per capita recovery rates are represented by the parameters  $a_1$  and  $a_0$  can use the function to generalize a nonlinear incidence rate.

$$f(S, I) = \frac{\alpha_1 SI}{b_1 + b_2 S + b_3 I} \tag{2}$$

Thus, the system of differential equations is given by

$$\frac{dS}{dt} = (1 - q) a - \sigma_1 S - f(S, I) + \gamma R, \tag{3}$$

$$\frac{dI}{dt} = f(S, I) + (\sigma_2 + b) I, \tag{4}$$

$$\frac{dR}{dt} = qa - (\sigma_2 + \gamma) R + bI, \tag{5}$$

$S(t)$  is the susceptible population,  $I(t)$  is the population that is infected, and  $R(t)$  is the population that has recovered, so that  $N = S + I + R$ .

### Logistic and SEIR Hybrid Model

As a result, this model can only forecast cumulative infections but not the present number of confirmed cases. The SEIR model can simulate population shifts resulting from infectious disease outbreaks. According to Nawaz et al. (38), if the number of confirmed COVID-19 diagnoses each day ( $I$ ) is projected, the number of healthy persons ( $S$ ), exposed people ( $E$ ), and infected people ( $I$ ) need to be computed. It is also necessary to examine the number of recovered persons ( $R$ ), the infection rate of healthy people becoming latent, the infection rate of exposed people getting infected, and the recovery rate of diseased people. The rationale behind the SEIR and the logistic model is that the logistic model’s parameters are the initial number of infections  $P_0$ , the cumulative number of infections  $K$ , and the infection rate ( $r$ ), which may be automatically produced after dataset training. We may utilize ( $r$ ) in the logistic model to initialize in the SEIR model. The SEIR model’s starting numbers of infected people ( $I$ ) and exposed people ( $E$ ) are set to zero when  $P_0$  is used in the logistic model. When calculating the number of rehabilitated persons ( $R$ ), the first day’s total and the number of dead people in the data are used as a starting point. It can start SEIR model ( $S$ ) by multiplying the Logistic model ( $K$ ) by a coefficient larger than (1) if the starting number of healthy persons within a specified range is similar to or more than the ultimate cumulative number of illnesses. It may make reasonably simple adjustments to the infection and recovery rates of those exposed and those who have been affected. The accuracy of the predictions will rise. The algorithm and implementation of the Logistic SEIR model.

Logistic function as defined in Equation 1, logistic function ( $t, K, P_0, r$ ):

$$Return = kQ_0 e^{rt} / [k + Q_0(e^{rt} - 1)] \tag{1}$$

**TABLE 2 |** The results of SEIRD parameters.

Parameters	Initial Values used in this study <i>Punjab; Sindh; KPK; Baluchistan; AJK</i>	Source
Susceptible (S)	S = 33,000,000; 29,000,000; 27,000,000; 23,000,000; 21,000,000 S rate = 33,000,000/64000,000 = 0.5156; 0.4531; 0.4218; 0.359375; 0.328125	
Exposed (E)	E = 99,219 E ratio = 99,219/64,000,000 = 0.001550; 0.001571; 0.001610; 0.00154; 0.001510	https://www.undp.org.com
Infected (I)	I = 2114 I ratio = 0.00003303	https://www.undp.org.com
Recovered (R)	R = 95,421; 13,600; 130137; 35,971; 15,916 R ratio = 95,421/64,000,000 R = 0.00149; 0.0002125; 0.002033; 0.0005620; 0.0002486	https://www.undp.org.com
Deaths (D)	D = number of deaths due to COVID-19 = 2231; 2517; 1263; 148; 182	

Define the SEIR model function according to Equations 2–5:

$$diff\_SEIR (Y = [S, E, I, R]):$$

$$X(0) = -\beta \times Y(0) \times Y(2) \tag{2}$$

$$X(i) = \beta \times Y(0) \times Y(2) - Y(1) \times b \tag{3}$$

$$X(ii) = b \times Y(1) \times Y(2) \tag{4}$$

$$X(iii) = \gamma \times Y(2) \tag{5}$$

Reappearance Y

- (1) The initial number of infected persons ( $Q_0$ ) and the ultimate cumulative number of infected people ( $k$ ) may be determined using the nonlinear least-squares approach to fit the logistic function on the training set ( $r$ ).
- (2) In this step, use the previously collected parameters to initialize the SEIR model parameters partially and then initialize the remaining parameters.
- (3) The SEIR model differential equation is solved in Matlab, and the projected number of confirmed patients is obtained.

From 7 to 4, the number and complexity of setting a single SEIR model parameter are considerably reduced by the hybrid model suggested in this article. You can get better prediction results by fine-tuning the parameters within a narrower range.

**RMSE AND MAE**

The SEIR individual's prediction is not good enough. Seven parameters need to be debugged repeatedly to get a decent prediction effect from the SEIR model, which is a time-consuming and hard operation. A quantitative study of the model's prediction effect uses the RMSE and MAE.

$$RMSE = \sqrt{\frac{1}{M} \sum_{T=1}^M (z_t - q_t)^2} \tag{6}$$

$$MAE = \frac{\sum_{j=1}^m |x_j - y_j|}{m} \tag{7}$$

**TABLE 3 |** The results of SEIRV coefficients and portrayals.

Coefficients	Initial values used in this study	Source
$\beta =$ Coefficients of infection	$\beta = \beta_0 * k$ $\beta =$ coefficient of infection rate K = frequency of exposure	(39)
$\beta_0 =$ probability of infection per espoure	$\beta_1 = 6.31$ $\beta_2 = 1.11$ $\gamma = 1/T_i$ $T_i =$ average recovery time $T_i = 14$ days $\gamma = 1/14 = 0.071$	(40)
$\gamma =$ Coefficient of Migration rate	$\delta_e =$ average value of Latency =5.1 days =1/5.1 =0.196	(13)
$\delta =$ Coefficient of Latency		

Where  $q_t$  and  $z_t$  represent the  $t^{th}$  predicted and true values,  $M$  is the total number of samples for which a predicted, and true values exist. No matter how one looks at the RMSE model provided in this study, it appears to have certain advantages over other models. **Figure 1** shows the model's coefficients.

**Algorithmic Demonstration**

**Algorithmic Representation of the Method**

**Step-1: Start**

**Input:** N data sets

**Step-2: Processing:** Data Cleaning

**Tuning of Parameter:** Based on Logistic growth

**Step-3: Construction of Model:**

Establish a prediction model based on the SIER model.

**Step-4: Validation:**

**Step-5: The parameters are appropriate**

**TABLE 4 |** The MAE and RMAE of the hybrid models.

Localities	Methods	
	Logistic hybrid model	SEIRV model
<b>Punjab</b>		
MAE	0.89	2.51
RMSE	2.72	30.71
<b>Sindh</b>		
MAE	0.91	3.53
RMSE	3.86	34.66
<b>KPK</b>		
MAE	0.84	2.23
RMSE	3.11	28.47
<b>Baluchistan</b>		
MAE	0.78	2.27
RMSE	2.31	30.17
<b>KPK</b>		
MAE	0.719	2.11
RMSE	2.29	24.56

**If Yes - - - Obtain Forecast Result**

**If No - - - Repeat step 2-5**

According to the NCDC, the first incidence of COVID-19 will be reported in January 2020.

Three incidents occurred in February and are consistent throughout the month.

## RESULTS AND DISCUSSIONS

### Hybrid Model-Analysis Results

The hybrid model's SIR and SEIR have been used to study and predict the spread of COVID-19 disease in this review. We used LR, ARIMA, SVR, MLP, RNN, GRU, and LSTM for the same reason. These models are evaluated using RMSE, MAE, and MAPE for their display. From March 2020 to September 2020, the number of confirmed COVID-19 cases is utilized to conduct the investigation using SIR and SEIR models. **Table 1** presents the results for each of the study areas.

The limit LSTM model outperformed all of our other recognized models in performance. In addition, the model's productivity in the AJK district is higher than in other regions. According to information provided by Pakistan's Public Order and Activity Center (NCOC) and used in the evaluation of execution, RMSE for the SEIR model results is expected to have a high value in all regions of Pakistan up to and including the year 2020. The RMSE of the two models is approximately multiple times more than MAPE and MAE, which further confirms our crossover model's predominance. Compared to the SEIR model, three borders should be effectively defined, and more information is anticipated to dissect and introduce the boundaries in the half and half model. Longer-term expectations need to grow the recovery rate more because of quantitative examinations by SIR and SEIR models of RMSE and the recommended model the plague develops. The SEIR model then registers Punjab's  $R_0$

**TABLE 5 |** The results of all provinces of population size.

Regions	Rate	MAE	RMSE
Punjab	K	1.58	12.531
	1.15*K	0.96	8.137
	1.30*K	1.15	3.72
	1.45*K	1.51	5.435
	1.60*K	1.57	6.451
Sindh	K	1.58	11.153
	1.15*K	1.17	8.731
	1.30*K	2.15	4.72
	1.45*K	3.53	4.735
	1.60*K	3.87	5.415
Baluchistan	K	1.66	10.531
	1.15*K	1.11	7.137
	1.30*K	2.15	5.72
	1.45*K	3.53	5.253
	1.60*K	3.71	6.135
KPK	K	1.37	9.531
	1.15*K	1.01	7.107
	1.30*K	1.15	3.52
	1.45*K	1.51	5.351
	1.60*K	1.57	6.401
AJK	K	1.28	8.131
	1.15*K	0.91	3.137
	1.30*K	1.01	2.112
	1.45*K	1.23	2.435
	1.60*K	1.37	3.451

value, greater than that of other regions. Taking a look at the trend is almost guaranteed that the number of instances will increase. Medical professionals, healthcare workers, and others who provide basic support must be protected by established clinical standards. **Table 2** contains the starting values utilized in the hybrid SEIR model for calculation purposes and the source data for each region of interest in this study.

As a result of running the SEIR model, researchers can understand how quickly the virus might spread throughout Pakistan. The results of this simulation are extremely useful in strengthening the strategic capacity to protect Pakistan's provinces against COVID-19 conditions. Changes in parameters such as susceptible ( $S$ ), exposed ( $E$ ), infected ( $I$ ), recovered ( $R$ ), and mortality rates affect the hybrid model ( $D$ ). Adding to the complexity is the requirement to use extra data to establish the starting values of the various SEIR model parameters. According to the current statistics, the estimated peak number of infected persons is higher in Punjab than in any other province. At the same time, the recovery ( $R$ ) and death ( $D$ ) rates are likewise higher in Punjab. According to statistics from the number of COVID-19 cases analyzed, the COVID-19 spreading model is a SEIRV model in the provinces of Pakistan (Punjab, Sindh, Baluchistan, KPK, and AJK). The study's findings may model the transmission of the COVID-19 virus throughout the country

**TABLE 6** | Determination of the impact of seven parameters.

Region	$\alpha$				$\gamma$			
	1	1/5	1/10	1/15	1/10	1/14	1/18	1/20
<b>Punjab</b>								
MAE	1.45	0.94	1.11	1.38	1.53	1.04	0.91	1.05
RMSE	11.07	3.91	7.65	7.871	9.51	6.04	3.41	4.11
<b>Sindh</b>								
MAE	1.31	1.27	1.21	1.35	1.61	1.14	0.97	1.11
RMSE	13.11	3.73	7.28	7.371	8.53	5.84	3.57	5.23
<b>KPK</b>								
MAE	1.43	0.94	1.11	1.45	1.57	1.21	0.87	0.99
RMSE	9.37	4.10	6.18	7.717	7.35	5.04	4.91	5.91
<b>Baluchistan</b>								
MAE	1.37	1.04	1.17	1.39	1.46	1.24	0.93	1.01
RMSE	9.07	4.21	5.81	6.671	7.55	6.11	3.87	6.01
<b>AJK</b>								
MAE	1.13	0.74	0.99	1.21	1.33	1.14	0.74	0.81
RMSE	8.71	2.52	6.56	5.713	6.31	4.41	3.91	4.61

by taking into account vaccination and isolation periods in the COVID-19 population.

### SEIR Intervention of Vaccination

The model is a development of SEIR that incorporates vaccination as an additional intervention. The suggested model uses the predictors listed in the methodological section’s parameter table. The structural assumptions of a compartmental epidemiological model have the biggest impact on the results. The SEIRV model aims to evaluate the influence of vaccination on an ongoing pandemic, such as the present COVID-19 one, and examine “international and optimum” vaccine distribution options SEIRV models. The SEIRV model’s mean delay coefficient peak values are shown in **Table 3**.

The study’s strategic mixture model fits well with the information gathered from NCOC and various data sets. Furthermore, this investigation examined the impact of openness and its level on each location. **Table 4** records the MAE and RMSE for the mixed-race SIERV model affirmed cases. Even though the available information on COVID-19 is extremely limited, the SIERV model performed well. The RMAE’s benefits for recovered and confirmed cases are always <5%, well below the margin of error for confusion. In July and August of 2020, when the entire country is put on lockdown due to passing cases. However, the model achieves an RMAE of just 2.2% on average.

As far as the number of cases and deaths are concerned, the COVID-19 in Pakistan has been rather limited. Five hundred eighty-five thousand four hundred thirty-five cases and 13,076 deaths were recorded throughout the country on March 4, 2021. It has been determined that Sindh, southern Punjab, and KPK are the areas with the highest risk of exposure to COVID-19, based on population density and proximity to family members and access to water, sanitation, and cleanliness. Comparatively,

Pakistan has an extraordinarily low number of people afflicted by the COVID-19 and a low number of deaths, especially compared to developed nations like the United States and Germany.

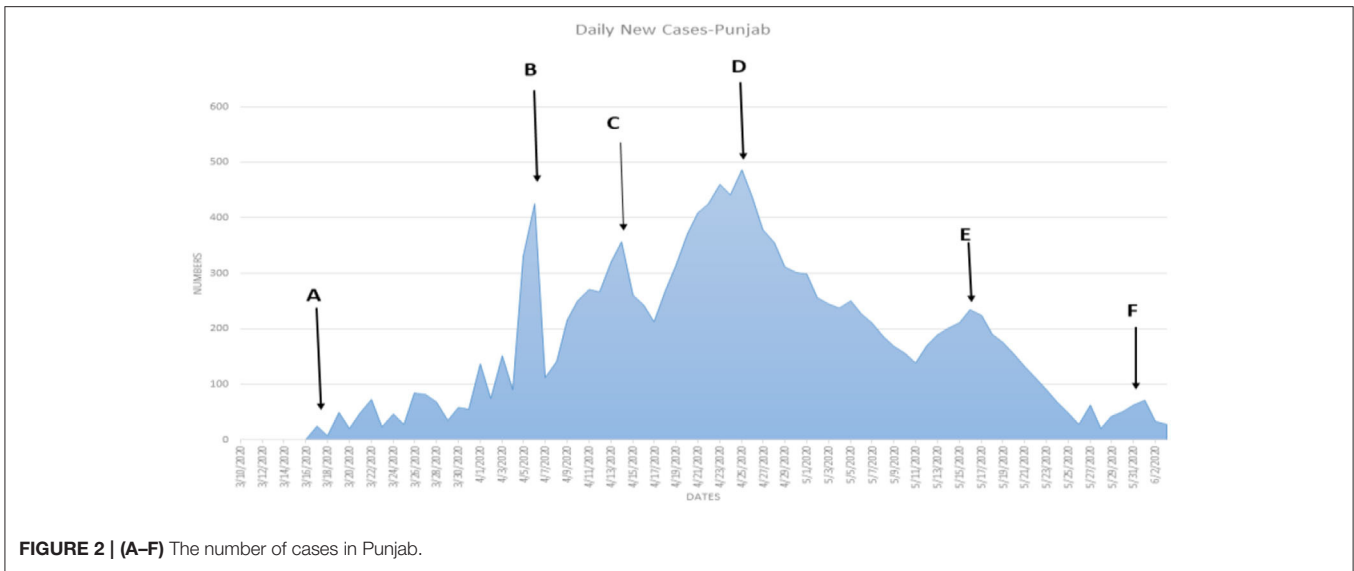
It is not until the crossover arrangement has been selected that the final mixture model is attempted with the last informative index separated from the beginning. **Table 5** provides a variety of error estimates for this configuration: Use the MAE for quantitative research to understand the model’s influence on expectations better. The SEIR model can simulate the changes in the number of various populations during the spread of infectious diseases. The values show the parameters. In addition, the infection rate  $\beta$  of healthy people becoming latent, the infection rate of exposed people becoming infected  $\alpha$ , and the recovery rate of infected people becoming recovered  $\gamma$  need to be considered. **Table 6** demonstrates that Pakistan’s populous strongly influences the spread of COVID-19  $\alpha$  and  $\gamma$  parameter impact. From the perspective of the prediction curve and the error, the influence of different initial values of the conversion rate  $\alpha$  from the exposed person to the infected person on the curve prediction effect is studied. Thus, a total of seven parameters are adjusted.

**Table 6** shows that the weighted error is the smallest when the conversion rate of an exposed person becomes an infected person,  $\alpha = 1/4$ , which is 3.67 and has the best effect. As time increases, the larger the  $\gamma$ , the better the prediction effect. With the development of the epidemic, the cure rate is growing, so as time increases, the prediction effect with a larger  $\gamma$  value will be better. In the actual prediction process, the value of  $\gamma$  can be increased every 6 days on the additional test set (for example, from 1/20, 1/18, and 1/14 until close to 1). On March 20, there was a significant increase in sickness transmission. Between March and September 2020, the number of confirmed cases and deaths in Punjab, Baluchistan, Sindh, KPK, and AJK is shown in **Figures 2–6**. According to new data, the transmission of the illness has dramatically changed since March. Validate the model parameter through a comparison method (i.e., RMSE, MAE). The result of confirmed cases in the log10 base vs. the last 7 days of training data. The reason for using the last 7 days of training data is because it shows major growth in confirmed cases, as shown in **Figure 6**.

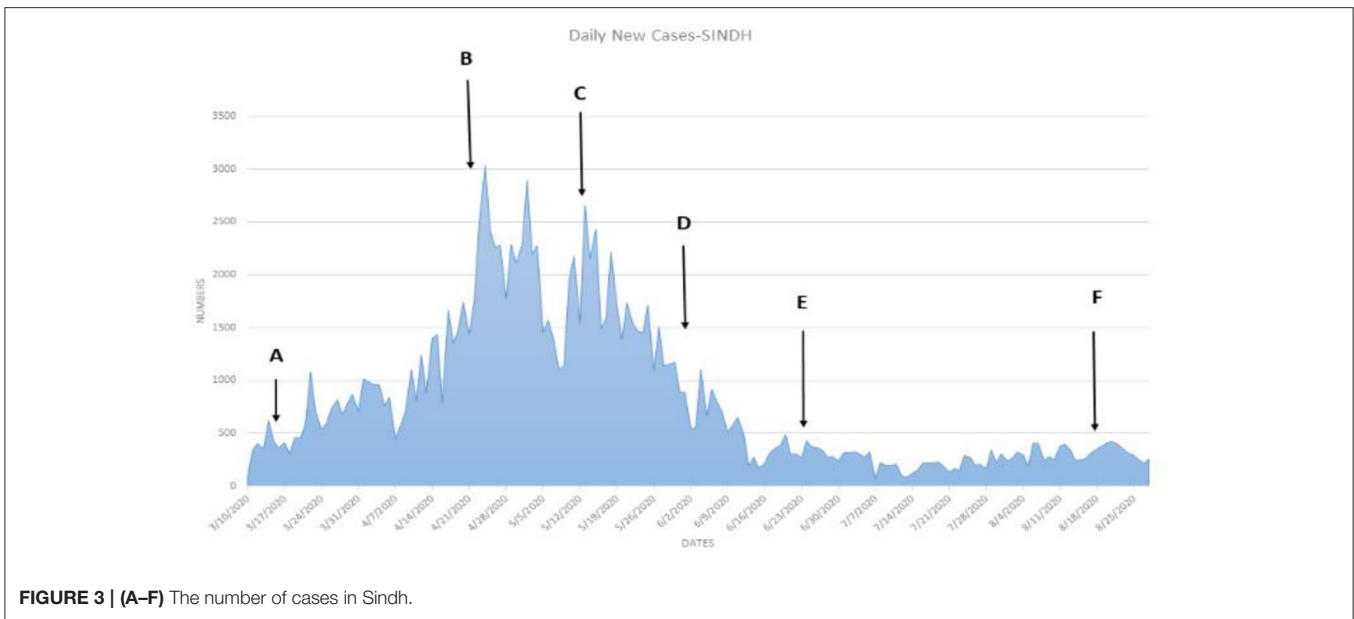
**Figures 7A–E** depicts the most recent projections for each area of Pakistan for the second half of December 2020. Predicted numbers of deaths, recovered cases, and confirmed cases all agree perfectly with the data. They match the information perfectly during the first 6 days, but by the end of the following week, they are more likely to underestimate or misunderstand the observed attributes. This framework’s potential to achieve astonishing multiweek expectation levels is demonstrated mathematically in the results to deal with the pandemic’s current state and projecting out to future times is shown in **Figure 8**.

**Figure 8** depicts the anticipated outcome based on the use of two models. By adjusting the SEIR model’s coefficients repeatedly, the repeated organization helps clarify the framework’s evolution and possible modifications.

COVID-19 is contagious and the rate at which the illness spreads from one infected person to another is 2.67; the study



**FIGURE 2 | (A-F)** The number of cases in Punjab.



**FIGURE 3 | (A-F)** The number of cases in Sindh.

needs to know the maximum rate at which the illness may spread. The study used 7-day preparatory information in all models because there is no big trend in Pakistan before walk 2020. SEIR model boundary distinguishes changes in the pandemic development and adjusts the SEIR model coefficient and results to remedy the expectation. The forecast bend, MAE, and RMSE showed that the proposed model in the research is superior to the SEIR model in the investigation. It is possible to conduct this review in different countries, allowing the model to be used in different countries.

Although it does not change drastically over time, the SEIR model boundary benefits from several intervening strategies, as seen in **Figure 9**. Similarly, the SEIR model

relies on the number of vulnerable people in the population, and a day-by-day estimate of the framework’s relevance for making immediate forecasts began on January 10, followed by estimates on January 14, January 18, and January 20. The osmosis cycle’s evaluated status and bounds guide these hypotheses.

**Figure 9** also aids in comprehending the Pakistani government’s plans for dealing with COVID-19 in each of the country’s regions. A time-series graph depicts COVID-19 inoculation in **Figure 10**, it becomes more apparent that there is a positive vertical pattern in the information. People who provide basic forms of assistance, such as specialists and healthcare workers, need to have their safety and wellbeing protected by

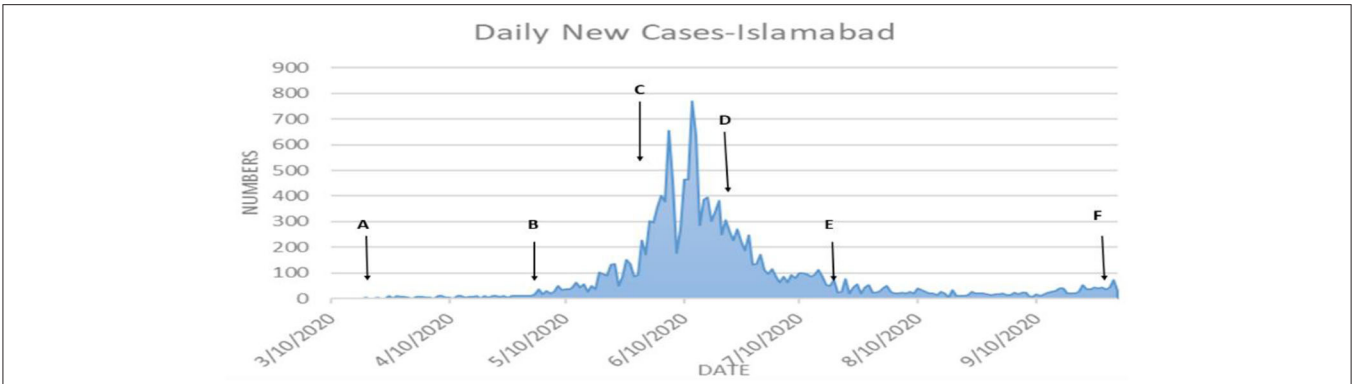


FIGURE 4 | (A-F) The number of cases in Islamabad.

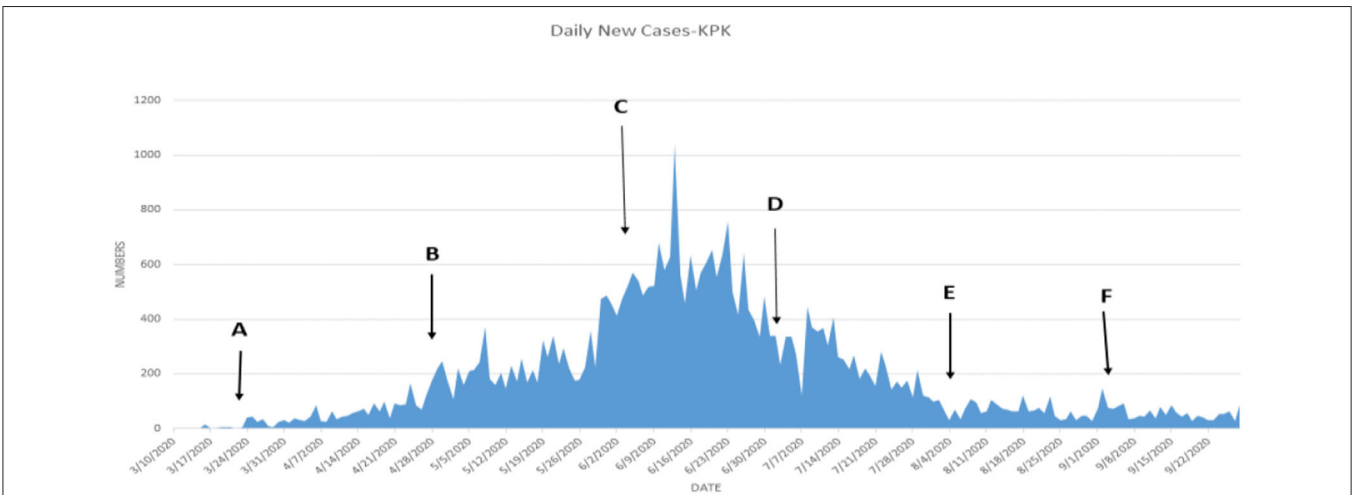


FIGURE 5 | (A-F) The number of cases in KPK.

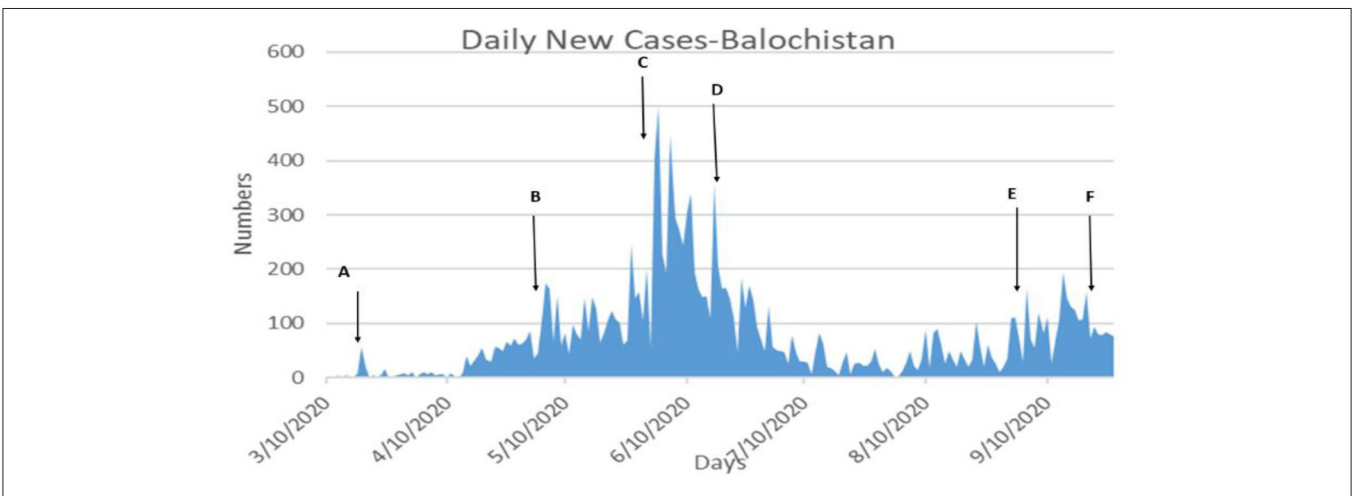
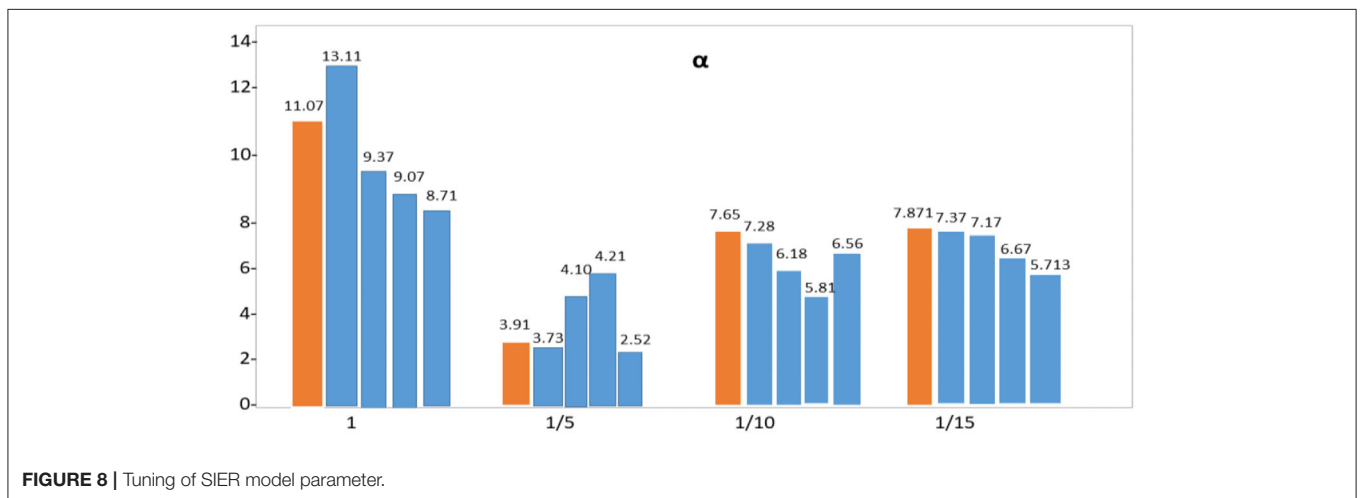
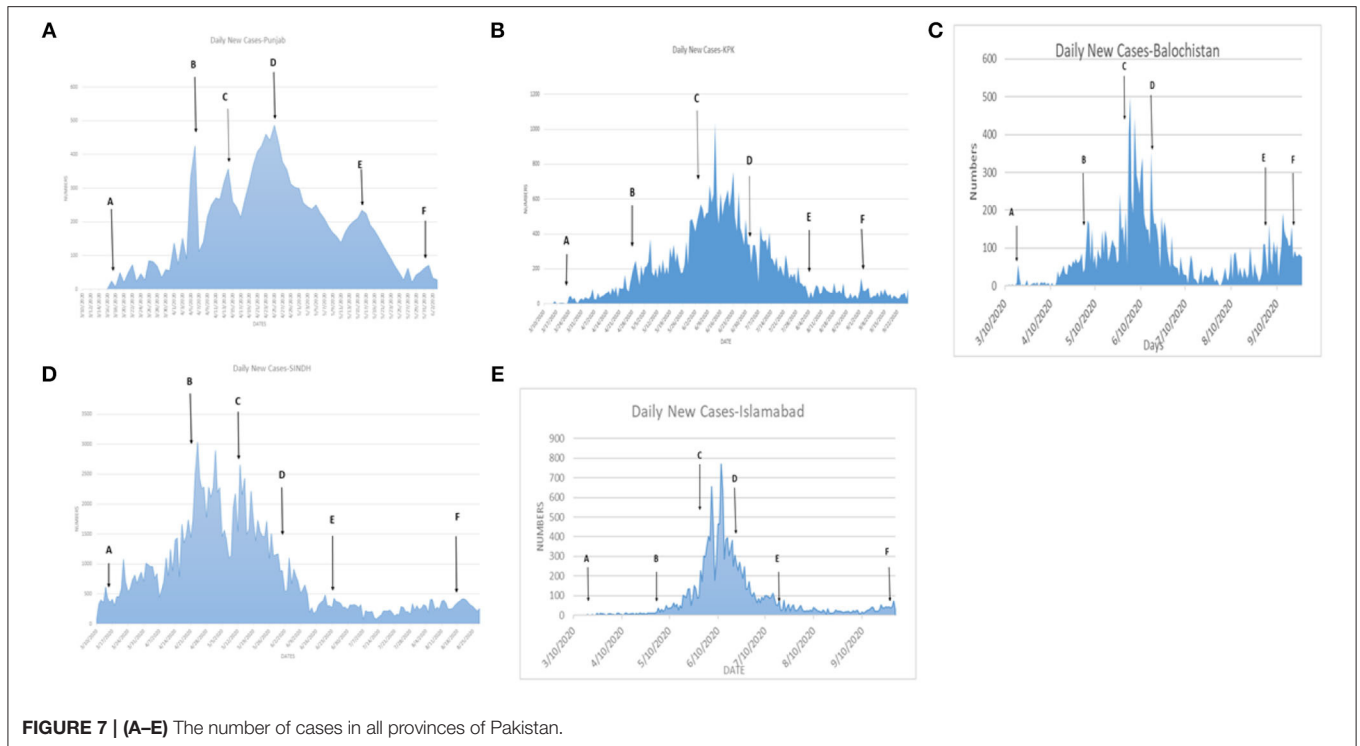


FIGURE 6 | (A-F) The number of cases in Baluchistan.





nationally recognized clinical standards. People’s indifference and gatherings could lead to a significant increase in the number of cases in the future. The public authority must be even more vigilant and implement harsher measures to avoid a climax. In addition, clinical office layouts across the country must be compelled to be modernized.

**Figure 11** represents the actual reported, exposed, infectious, and the estimated COVID-19 cases; **Figure 12** left graphs show the size, and the right graphs show the vaccination graphs.

Vaccination is proven effective in the stated case, and with the passage of time and awareness, the volume of vaccination is increasing.

## DISCUSSIONS

In this study, the SEIR model is used to assess COVID-19’s effects in Pakistan’s Punjab and Sindh provinces and the rest of the country. Compared to other places, the situation in these two towns is dire, and the government of Sindh should take specific measures to protect residents from the spread of COVID-19. Because it considers the interactions between small groups of people spread across Pakistan’s many regions, the SEIR model generates new aspects and variables. Others may add classes to the model and update the results if unexpected arrangement executions, such as lockdowns, clever lockdowns, and travel boycotts between cities, are carried out. Even under

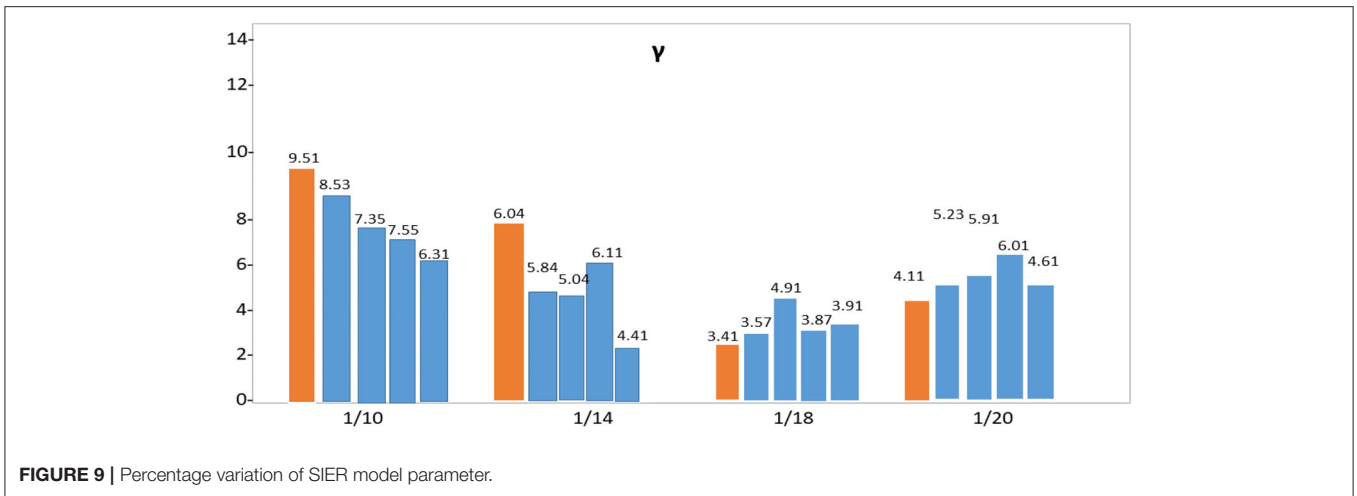


FIGURE 9 | Percentage variation of SIER model parameter.

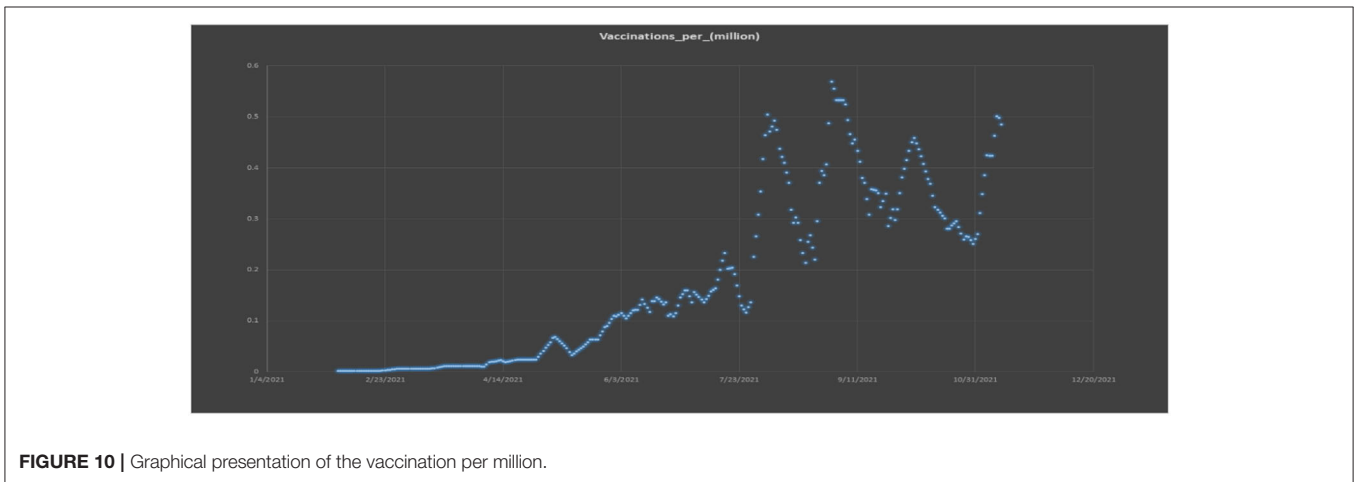


FIGURE 10 | Graphical presentation of the vaccination per million.

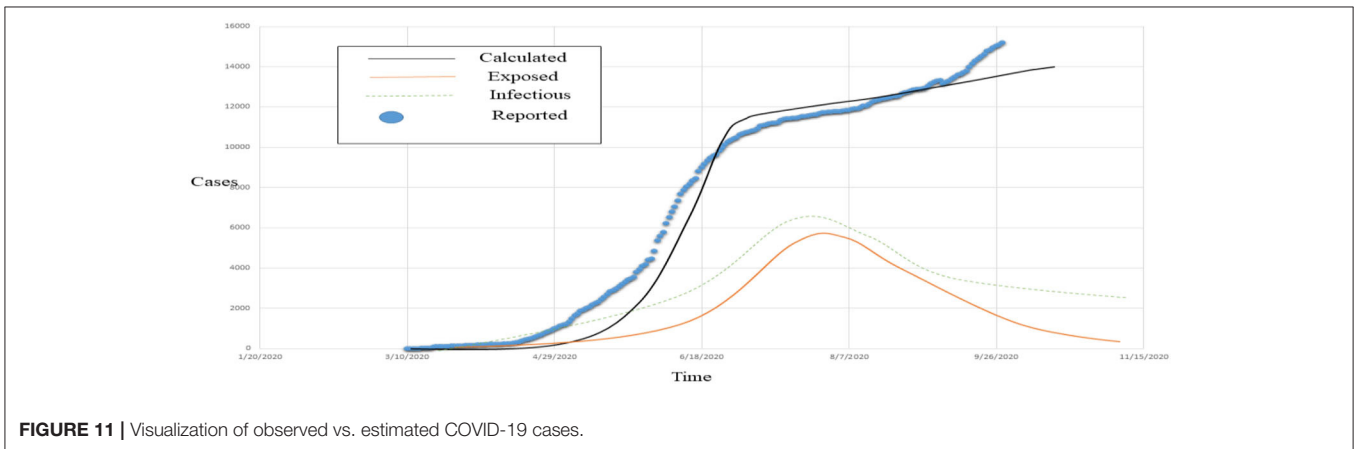
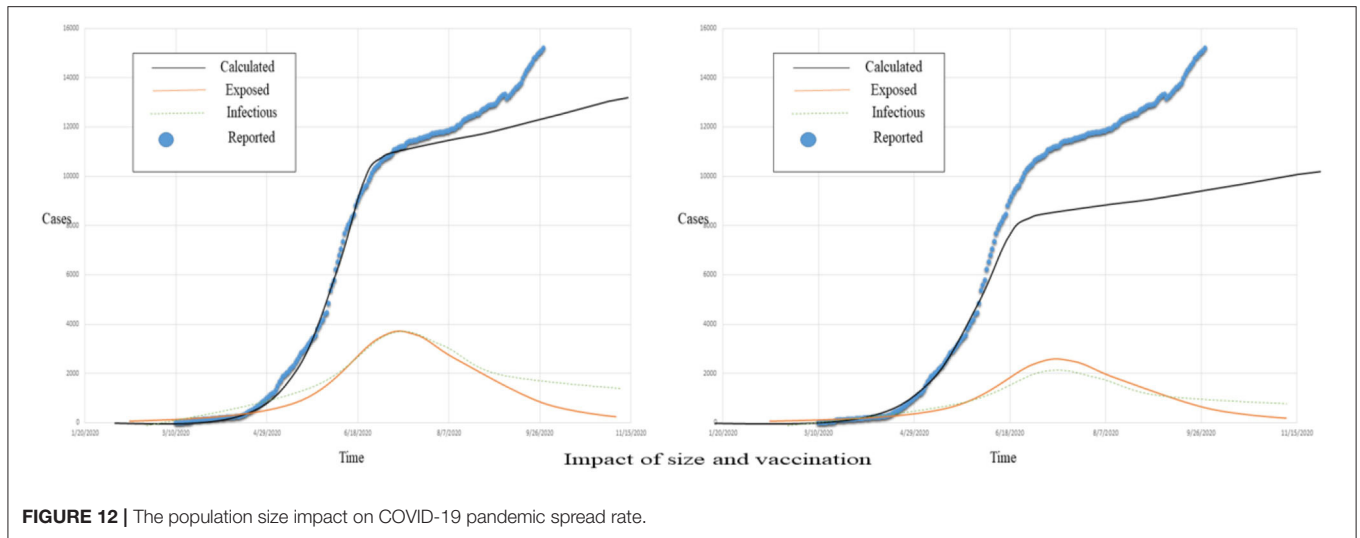


FIGURE 11 | Visualization of observed vs. estimated COVID-19 cases.

different models, the advantages can still be quite subtle and arbitrary.

The pandemic model predicts that more people will contaminate than in the days leading up to and following the outbreak. Additional precise data on boundaries will be obtained

after the end of infection and may use entire information to anticipate the improvement of (and  $R_0$ ) with time. Social and familial seclusion are effective strategies, according to these findings. There is no way to prepare for a pandemic in a medical care system that is not designed to handle normal



**FIGURE 12 |** The population size impact on COVID-19 pandemic spread rate.

scenarios. As the number of people exposed to pollution grows, illness outbreaks can occur within days. The number of setbacks might be significant if the spread and dominance of illnesses are widespread. One additional lockdown is achieved with a certain space intelligent arrangement where the dispersion is much wider, and it may have a significant impact on preventing this calamity. After a few days of lockdown, educational facilities are allowed to reopen under the authority of approved standard operating procedures. Cooperatives in high-risk areas have been protected, and the public authority has provided help for everyday use goods such as food and water.

## CONCLUSIONS

COVID-19 epidemic is simulated and predicted using an enhanced SEIR model with seven phases of infection, including vaccination. The number of deaths, the number of recovered cases, and the number of ongoing cases are analyzed using genuine data. Two weeks into the future, the revised status and parameters are utilized to make short-term projections of the COVID-19 pandemic to assess the system's applicability. The final analysis focused on vaccination's role in disease transmission. As a final step, we hope that our findings will assist policymakers to design psychological treatments that will lessen the psychosocial effects of COVID-19 while also benefiting the most disadvantaged populations that are more likely to suffer from poor health as a result of this pandemic. Based on the validation findings, we can infer that the suggested hybrid model has a reasonable capacity to forecast and decent performance. Long-term projections of the outbreak's dynamics are particularly valuable for healthcare personnel and government authorities. Logistic models cannot predict the number of confirmed diagnoses, but the hybrid model in the research does. The SEIR model can forecast the number of confirmed diagnoses, but it takes a huge quantity of data and repeated revisions to be

modified. Experiment results show that the suggested model in this study is superior to the SEIR alone, as demonstrated by the prediction curve, MAE and RMSE values.

## Limitations of SIR and SIER Models

A fresh direction and new way of thinking have been supplied for the COVID-19 pandemic due to the growth in vast amounts of data and the development of computerized reasoning. The premise of artificial intelligence considers the links between objects rather than the substance of the causal relationship. Because we did not want to keep making conclusions based on existing examples of data, we combined the SIR and SEIR restrictions. When the two methods and enhancements are combined, amazing outputs and unmatched concepts result. The creamer model overwhelms the fundamental model's inability to foresee the amount of validated examination. It must incorporate a considerable volume of data into SEIR's model, which can already predict the number of examinations conducted.

## Future Suggestions

These results may also apply to other countries with similar socioeconomic profiles as COVID-19 affects all the countries, and social distancing measures are more or less similar globally. Furthermore, given the recent increase in COVID-19 cases in Pakistan, there is a risk of compliance with procedures for preventative precautions. Following these guidelines and disseminating the right information requires complex awareness campaigns and educational interventions that focus on safe health practices and proper evidence-based information about this disease.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary

material, further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

WZ: conceptualizing, writing, drafting, and data and methodology. YS: conceptualizing, writing, drafting-original draft, and data and methodology. YL:

conceptualizing and writing. WG: data and methodology. All authors contributed to the article and approved the submitted version.

## FUNDING

The authors thank the financial support from the Natural Science Foundation of China (Grant Number: 71973073).

## REFERENCES

- Pakistan AMo. *Administrative Map of Pakistan*. Islamic Republic of Pakistan (2020). Available online at: <https://www.nationsonline.org/one-world/map/pakistan-administrative-map.html>
- Ali I, Shah SA, Siddiqui N. *Pakistan Confirms First Two Cases of COVID-19, Govt says "no need to panic"*. DAWCOM N. (2020). p. 26. Available online at: <https://www.dawn.com/news/1536792>
- Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet*. (2020) 395:912–20. doi: 10.1016/S0140-6736(20)30460-8
- Diekmann O, Heesterbeek H, Britton T. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton Series in Theoretical and Computational Biology. Princeton, NJ: Princeton University Press (2013).
- Qi C, Zhu YC, Li CY, Hu YC, Liu LL, Zhang DD, et al. Epidemiological characteristics and spatial-temporal analysis of COVID-19 in Shandong Province, China. *Epidemiol Infect*. (2020) 2020:141–8. doi: 10.1017/S095026882000151X
- Zou Y, Pan S, Zhao P, Han L, Wang X, Hemerik L, et al. Outbreak analysis with a logistic growth model shows COVID-19 suppression dynamics in China. *PLoS ONE*. (2020) 15:e0235247. doi: 10.1371/journal.pone.0235247
- Bagal DK, Rath A, Barua A, Patnaik D. Estimating the parameters of the susceptible-infected-recovered model of COVID-19 cases in India during lockdown periods. *Chaos Solitons Fractals*. (2020) 140:110154. doi: 10.1016/j.chaos.2020.110154
- Alsayed A, Sadir H, Kamil R, Sari H. Prediction of epidemic peak and infected cases for COVID-19 disease in Malaysia, 2020. *Int J Environ Res Public Health*. (2020) 17:4076. doi: 10.3390/ijerph17114076
- Iffat Idress. *Areas and Population Groups in Pakistan Most Exposed to Combined Effects of Climate Change, Food Insecurity and COVID-19*. GSDRC, University of Birmingham (2021).
- Mourtzoukou EG, Falagas ME. Exposure to cold and respiratory tract infections. *Int J Tuberc Lung Dis*. (2007) 11:938–43.
- Minoza JMA, Bongolan VP, Rayo JF. *COVID-19 Agent-Based Model with Multi-objective Optimization for Vaccine Distribution*. arXiv preprint arXiv:2101.11400 (2021).
- Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new COVID-19 associated with human respiratory disease in China. *Nature*. (2020) 579:265–9. doi: 10.1038/s41586-020-2008-3
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel COVID-19-infected pneumonia. *N Engl J Med*. (2020) 382:1199–207. doi: 10.1056/NEJMoa2001316
- WOM (2021). Available online at: <https://www.worldometers.info/COVID-19/country/pakistan/>
- Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med*. (2020) 26:855–60. doi: 10.1038/s41591-020-0883-7
- Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with covid-19 in Wuhan, China: a retrospective cohort study. *Lancet*. (2020) 395:1054–62. doi: 10.1016/S0140-6736(20)30566-3
- Minoza JMA, Sevilleja JE, de Castro R, Caoili SE, Bongolan VP. *Protection After Quarantine: Insights From a Q-SEIR Model With Nonlinear Incidence Rates Applied to COVID-19*. (2020). doi: 10.1101/2020.06.06.20124388
- Choudhuri A, Khatun M, Bandyopadhyay A, Dasgupta U, Chowdhury S, Bisoi A, et al. *Assessment of Worldwide COVID-19 Transmission Landscape for Predicting Its Upcoming Severity Along With a Clinical Update for Its Prevention*. (2020). doi: 10.22541/au.160347794.45407178/v1
- Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT. New machine learning method for image-based diagnosis of COVID-19. *PLoS ONE*. (2020) 15:e0235187. doi: 10.1371/journal.pone.0235187
- Alzahrani SI, Aljamaan IA, Al-Fakih EA. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *J Infect Public Health*. (2020) 13:914–9. doi: 10.1016/j.jiph.2020.06.001
- Blyuss KB, Kyrlychko YN. Effects of latency and age structure on the dynamics and containment of COVID-19. *J Theor Biol*. (2021) 513:110587. doi: 10.1016/j.jtbi.2021.110587
- Sun Y, Razaq A. Composite fiscal decentralisation and green innovation: Imperative strategy for institutional reforms and sustainable development in OECD countries. *Sustain Develop*. (2022) 1–14. doi: 10.1002/sd.2292
- Shaobo H, Yuexi P, Kehui S. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dyn*. (2020) 101:1667–80. doi: 10.1007/s11071-020-05743-y
- Rahmadani E, Lee H. ODE-based epidemic network simulation of viral Hepatitis A and kernel support vector machine based vaccination effect analysis. *J Korean Inst Intell Syst*. (2020) 30:106–12. doi: 10.5391/JKIS.2020.30.2.106
- Rusliza A, Budin H. Stability analysis of mutualism population model with time delay. *Int J Math Comput Phys Electr Comput Eng*. (2012) 6:151–5. doi: 10.5281/zenodo.1085667
- Diekmann O, Heesterbeek JAP, Roberts MG. The construction of next generation matrices for compartmental epidemic models. *J R Soc Interface*. (2010) 7:873–85. doi: 10.1098/rsif.2009.0386
- Anonim. *Situasi Kasus Indonesia*. (2020). Available online at: <https://covid19.kemkes.go.id/> (accessed April 16, 2022).
- Bongolan VP, Minoza JMA, de Castro R, Sevilleja JE. Age-stratified infection probabilities combined with a quarantine-modified model for COVID-19 needs assessments: model development study. *J Med Internet Res*. (2021) 23:e19544. doi: 10.2196/19544
- Clifford S, Pearson CA, Klepac P, Van Zandvoort K, Quilty BJ, CMMID COVID-19 working group, et al. Effectiveness of interventions targeting air travellers for delaying local outbreaks of SARS-CoV-2. *J Travel Med*. (2020) 27:taaa068. doi: 10.1093/jtm/taaa068
- Chandra R, Shaurya G, Rishabh G. Evaluation of deep learning models for multi-step ahead time series prediction. *IEEE Access*. (2021) 9:83105–23. doi: 10.1109/ACCESS.2021.3085085
- Celeste JJ, Bongolan VP. School reopening simulations with COVID-19 agent-based model for Quezon City, Philippines. *Int Arch Photogramm Remote Sens Spat Inf Sci*. (2021) 46:85–90. doi: 10.5194/isprs-archives-XLVI-4-W6-2021-85-2021
- Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in china under public health interventions. *J Thorac Dis*. (2020) 12:165–74. doi: 10.21037/jtd.2020.02.64

33. Castillo Ossa LF, Chamoso P, Arango-López J, Pinto-Santos F, Isaza GA, Santa-Cruz-González C, et al. A hybrid model for COVID-19 monitoring and prediction. *Electronics*. (2021) 10:799. doi: 10.3390/electronics10070799
34. Ma N, Ma W, Li Z. Multi-model selection and analysis for COVID-19. *Fractal Fract*. (2021) 5:120. doi: 10.3390/fractalfract5030120
35. Tamang SK, Singh PD, Datta B. Forecasting of COVID-19 cases based on prediction using artificial neural network curve fitting technique. *Glob J Environ Sci Manag*. (2020) 6:53–64. doi: 10.22034/GJESM.2019.06.SI.06
36. Rao F, Mandal PS, Kang Y. Complicated endemics of an SIRS model with a generalized incidence under preventive vaccination and treatment controls. *Appl Math Model*. (2019) 67:38–61. doi: 10.1016/j.apm.2018.10.016
37. Shan C, Zhu H. Bifurcations and complex dynamics of an SIR model with the impact of the number of hospital beds. *J Differential Equations*. (2014) 257:1662–88. doi: 10.1016/j.jde.2014.05.030
38. Nawaz SA, Li J, Bhatti UA, Bazai SU, Zafar A, Bhatti MA, et al. A hybrid approach to forecast the COVID-19 epidemic trend. *PLoS ONE*. (2021) 16:e0256971. doi: 10.1371/journal.pone.0256971
39. Tang B, Wang X, Li Q, Bragazzi NL, Tang S, Xiao Y, et al. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J Clin Med*. (2020) 9:462. doi: 10.3390/jcm9020462
40. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis*. (2020) 20:553–8. doi: 10.1016/S1473-3099(20)30144-4

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhao, Sun, Li and Guan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.