Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 4849928, 12 pages
https://doi.org/10.1155/2022/4849928

Research Article

Global COVID-19 Epidemic Prediction and Analysis Based on Improved Dynamic Transmission Rate Model with Neural Networks

Yanyu Ding,[1] Jiaxing Li,[2] Weiliang Song,[2] Xiaojin Xie,[3] and Guoqiang Wang[3]

[1]*School of Chemistry and Chemical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China*
[2]*School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China*
[3]*School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China*

Correspondence should be addressed to Guoqiang Wang; guoq_wang@hotmail.com

Received 7 January 2022; Accepted 22 February 2022; Published 30 March 2022

Academic Editor: Wei Liu

The cross-regional spread of COVID-19 had a huge impact on the normal global social order. This paper aims to build an improved dynamic transmission rate model based on the conjugate gradient neural network predicting and analyzing the global COVID-19 epidemic. First, we conduct an exploratory analysis of the COVID-19 epidemic from Canada, Germany, France, the United States, South Korea, Iran, Spain, and Italy. Second, a two-parameter power function is used for the nonlinear fitting of the dynamic transmission rate on account of data-driven approaches. Third, we correct the residual error and construct an improved nonlinear dynamic transmission rate model utilizing the conjugate gradient neural network. Finally, the inflection points of the global COVID-19 epidemic and new outbreaks, as well as the corresponding existing cases are predicted under the optimal sliding window period. The experimental results show that the model presented in this paper has higher prediction accuracy and robustness than some other existing methods.

1. Introduction

With the acceleration of globalization and the development of science and technology, the population is moving rapidly, and the world has become interdependent and interconnected, making it possible for infectious diseases to spread rapidly. In early December 2019, the first case of a new type of coronavirus was reported in Wuhan, China, and it was named corona virus disease 2019 (COVID-19) on February 11, 2020. Subsequently, the COVID-19 epidemic spread rapidly from Hubei Province to many provinces across China and brought serious harm to the lives and health of Chinese, as well as to social and economic development. Until late February 2020, the COVID-19 epidemic prevention and control work in China had achieved a phased victory. Unfortunately, due to the lack of awareness regarding the COVID-19 epidemic in many overseas countries and inadequate prevention and control measures, the COVID-19 epidemic began to spread rapidly in Asia, the Middle East, Europe, and North America in late February and early March of 2020. In particular, the introduction of the concept of "herd immunity" by the British government has exacerbated the spread of the COVID-19 epidemic in Europe. Ferguson et al. [1] pointed out that if the British government does not change the current situation with respect to "herd immunity" approach, this wave of the COVID-19 epidemic will even cause 510,000 British deaths, and the number of deaths in the United States will be even greater, with possibly 2.2 million deaths. According to a report from Johns Hopkins University, as of January 24, 2022 (Beijing time), there were 351,983,072 confirmed cases and 5,614,569 deaths from the COVID-19 worldwide. More than 200 countries and regions in the world have confirmed the presence of the COVID-19 epidemic. More seriously, there have been 71,925,931 confirmed cases and 889197 deaths in the United States (USA). The top ten countries in terms of cumulative confirmed cases and deaths are shown in Figure 1.

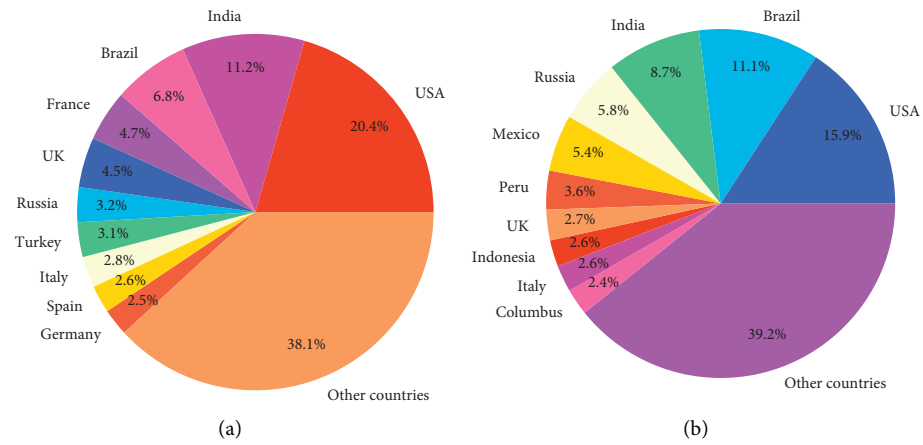

Figure 1: Top ten countries with cumulative confirmed cases and deaths.

Figure 1 shows that the United States is the country with the largest cumulative number of confirmed cases and deaths, accounting for approximately 20.4% and 15.9% of the global totals, respectively.

Statistics from Johns Hopkins University show that the cumulative number of confirmed cases worldwide exceeded 100 million cases on January 26, 2021, more than 200 million cases on August 4, 2021, more than 300 million cases on January 7, 2022. It is notable that it took 370 days from January 20, 2020 to reached 100 million cases. However, the increase rate of cumulative confirmed cases of the global COVID-19 epidemic has accelerated since Octobers 2020. It took 190 days to increase from 100 million to 200 million cases. More seriously, it only took 156 days to go from 200 million cases to 300 million cases. The details is displayed in Figure 2.

From Figure 2, it takes about 10–25 days for every 10 million increase from October 18, 2020, to December 12, 2021. Recently, with the new Omicron variant of COVID-19, the global spread of the COVID-19 epidemic has once again set a new record. It only took 14 days to increase from 270 million to 280 million cases, 7 days to increase from 280 million to 290 million cases, 4 days to increase from 290 million to 300 million cases, 4 days to increase from 300 million to 310 million cases, 3 days to increase from 310 million to 320 million cases, 4 days to increase from 320 million to 330 million cases, 3 days to increase from 330 million to 340 million cases, and 3 days to increase from 340 million to 350 million cases. However, there is no sign of improvement at present.

There are a lot of work have been studied on the transmission mechanism and prediction models of the COVID-19 epidemic, with the aim revealing the governing law of the epidemic, predicting changes and developmental trends, analyzing the causes and key factors of the epidemic, and seeking the best strategy for prevention and control. In particular, the SEIR model and its extensions are favoured by researchers. Wu et al. [2] used the SEIR model to predict the domestic and international spread of the COVID-19 epidemic in the short-term and long-term. The numerical results show that in order to avoid outbreaks in large cities with strong transport links in China, a large number of public health interventions must be implemented immediately at the population and individual levels. Caetano et al. [3] used an age-structured SEIR model to determine the impact of the implementation of past non-pharmaceutical interventions on the COVID-19 epidemic. Ghostine et al. [4] achieved encouraging results in small-scale short-term predictions by an extended SEIR model.

Based on the SEIR model, Wang et al. [5] constructed a complex network model for the spread of the COVID-19 epidemic in 15 cities with severe epidemics in Wuhan and the surrounding areas, focusing on the analysis of the possible time points at which the resumption of work in Wuhan and the surrounding areas could occur and the impact of resumption on the risk of secondary outbreaks. Yang et al. [6] used a modified SEIR model to predict the trend of the COVID-19 epidemic in China under public health interventions. Li et al. [7] studied the estimation of the scale of the COVID-19 epidemic in the United States based on air travel data from Wuhan. Liu et al. [8] used mathematical models to analyze the developmental trends of the COVID-19 epidemic in South Korea, Italy, France, and Germany. Hermanowicz [9] used a logistic model to conduct a systematic evaluation for predicting the growth of the COVID-19 epidemic in the United States based on 87-day data from China as of March 13, 2020 and 70-day data from the United States as of March 31, 2020. In the meantime, models based on data-driven statistics have also been widely used for the prediction and analysis of the COVID-19 epidemic, including function fitting [10, 11], machine learning [12–15], deep learning [16–18] and time series models [19, 20].

In particular, in view of the difficulty of accurately estimating the basic infection number R_0 in traditional infectious disease epidemiology, Huang et al. [21] proposed a data-driven concise and practical method for calculating the dynamic transmission rate of an epidemic to replace the basic infection number. Subsequently, Hu et al. [22, 23] used dynamic transmission rate model (DTRM) and dynamic growth rate model (DGRM) to predict and empirically analyze the domestic and global COVID-19 epidemics, respectively, and the experimental results show that the prediction accuracies of both models were greatly improved.

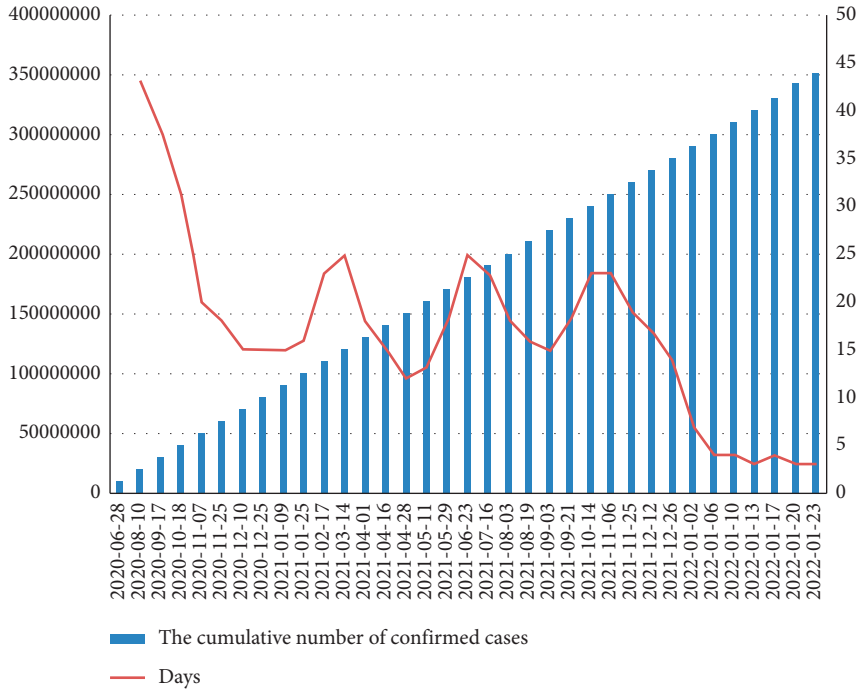

Figure 2: Global COVID-19 epidemic.

Compared with the SIR model [24], the SEIR model [2] and their extensions [5, 6], the DTRM based on a data-driven approach has a wider application range, higher prediction accuracy and stronger robustness. Xie et al. [25] used an nonlinear combinational DTRM (NCDTRM) based on support vector regression (SVR) to predict the COVID-19 epidemic in China. The experimental results demonstrate that NCDTRM effectively overcomes the shortcomings of insufficient information extraction and insufficient prediction accuracy of the single model. Xie et al. [26] presented an improved NCDTRM (INCDTRM) based on forecasting effective measure and SVR for analyzing and predicting the COVID-19 pandemic in eight countries. The experimental results reveal that INCDTRM has smaller prediction error and stronger generalization ability than the single prediction models, DGRM and NCDTRM that have been used previously.

Undoubtedly, the COVID-19 epidemic is a major public health emergency, that poses major challenge to the medical and health systems of all countries in the world and has a major impact on the global economic order. However, people are only concerned about the source, host, and spread of the COVID-19 epidemic, but the method of spreading, the pathogenic mechanism, the harmfulness, the lethality, the diagnosis and treatment plan, the treatment drugs, whether a given patient has sequelae after recovery, etc. have not been fully understood [27]. Therefore, in the face of the global COVID-19 epidemic, how to construct an effective and reasonable mathematical model for quickly, accurately and quantitatively evaluating the current stage of the epidemic, determining the effects of control measures, predicting future trends, and controlling the spread of the epidemic to avoid the collapse of the medical system has become a major issue and an urgent task for the government, the scientific community, and the public.

Existing research results show that when the sample size is not large enough, the DTRM model estimation parameter method has a large deviation [28]. Therefore, the paper aims to build an improved DTRM based on the conjugate gradient neural network (CGNN) (abbreviated as IDTRM-CGNN) to predict the inflection point of the COVID-19 epidemic in Canada, Germany, France, the United States, South Korea, Iran, Spain and Italy. The corresponding existing confirmed cases are also reported. The core idea of the IDTRM-CGNN is to use the CGNN to correct the residuals for improving the prediction accuracy. The empirical results show that our model has higher prediction accuracy and robustness than some other existing methods, and can provide scientific decision-making for the prevention and control of the global COVID-19 epidemic. Furthermore, these results can also provide a study and judgement of the effects of the epidemic control measures employed in different countries.

The rest of this paper is organized as follows: Section 2 briefly recalls some well-known results on the dynamic transmission rate, the inflection point of the COVID-19 epidemic, and the neural network based on the conjugate gradient method. Section 3, the IDTRM-CGNN is proposed to predict and analyze the global COVID-19 epidemic. Conclusions and remarks are made in Section 4.

2. Preliminary Knowledge

2.1. Dynamic Transmission Rate. In this subsection, we introduce the basic idea of the dynamic transmission rate, which plays an important role in the COVID-19 epidemic

predictions and analysis. For the details can be referred to [21–23, 25, 26].

Let $N(t)$ be the number of existing confirmed cases at time t. Then,

$$N(t) = L(t) - K(t) - D(t), \tag{1}$$

where $L(t)$, $K(t)$, and $D(t)$ are the numbers of cumulative confirmed cases, cumulative deaths and cumulative cures at time t, respectively.

It is well-known that the natural growth model is defined by

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = q(t)N(t), \tag{2}$$

where $q(t) \geq 0$ is the growth rate of the number of existing confirmed cases at time t.

Let

$$q(t) = g(t) - 1. \tag{3}$$

It follows from (2) and (3) that

$$\ln N(t) - \ln N(t_0) = \int_{t_0}^{t} g(x)\mathrm{d}x - (t - t_0). \tag{4}$$

Then, the number of existing confirmed cases is obtained by

$$N(t) = N(t_0)\exp\{a_t(t - t_0)\}, \tag{5}$$

where

$$a_t = \frac{\int_{t_0}^{t} g(x)\mathrm{d}x}{(t - t_0) - 1}. \tag{6}$$

For the sake of analysis, we introduce the dynamic transmission rate, i.e.,

$$c_t = 1 + a_t, \tag{7}$$

which was first studied in [21]. Then, we have

$$c_t = 1 + \frac{1}{t - t_0}\ln\frac{N(t)}{N(t_0)}. \tag{8}$$

This implies that

$$c_t = 1 + \frac{1}{k}\ln\frac{N(t)}{N(t-k)}, \tag{9}$$

where $k = t - t_0$ represents the sliding window period, see [22, 23, 25, 26] or Section 3.3 for details.

2.2. Inflection Point of the COVID-19 Epidemic. The inflection point of the COVID-19 epidemic in this paper refers to the moment when the number of existing confirmed cases reaches a peak within a certain period of time [22]. Therefore, the inflection point of the COVID-19 epidemic is the key point for the prevention and control of the COVID-19 epidemic, so it is an important factor for measuring whether the COVID-19 epidemic is under control. The following are the trend charts of the numbers of existing confirmed cases in Canada, Germany, France, the United States, South Korea, Iran, Spain and Italy from the outbreak of the COVID-19 epidemic to April 7, 2020, as shown in Figure 3.

Figure 3 shows that the first wave of the COVID-19 epidemic inflection point in South Korea arrived on March 12, 2020. This means that South Korea has achieved good momentum with regard to the prevention and control of the first wave of the COVID-19 epidemic. However, the numbers of existing confirmed cases in other countries are still increasing, and the first wave of the COVID-19 epidemic was not effectively controlled before April 7, 2020; that is, there was no inflection point in the COVID-19 epidemic.

2.3. Neural Network Based on Conjugate Gradient Method. Artificial neural networks (ANNs), also known as neural networks, are mathematical models that imitate the behaviour characteristics of animal neural networks and perform distributed parallel information processing. They are widely used in pattern recognition, signal processing, knowledge engineering, expert systems, robot control and other fields [29]. In view of the local convergence of neural networks and the difficulty of slow convergence speeds, the study of the CGNN has attracted the attention of many scholars [30]. In what follows, we briefly introduce the basic idea of the CGNN, which updates the weight and bias values according to the scaled conjugate gradient method.

Considering the neural networks were composed of an input layer, $(L-1)$ hidden layers, and an output layer. Let $x \in R^n$ and $y \in R^m$ be the input and output of the given neural network, respectively. Furthermore, we denote $x^{(0)} \in R^n$ be the initial input of the neural network. Then the output of the k-th hidden layer of the neural network is obtained from

$$\begin{cases} u^{(k)} = W^{(k)}x^{(k-1)} + b^{(k)}, \\ x^{(k)} = f_k\left(u^{(k)}\right), k = 1, 2, \ldots, L, \end{cases} \tag{10}$$

where $f_k(\cdot)$ is the activation function of the k-th hidden layer, $x^{(k-1)} \in R^{N_{k-1}}$ is the output of the $(k-1)$-th hidden layer, also is the input of the k-th hidden layer, $W^{(k)} = (w_{ij}^{(k)}) \in R^{N_j \times N_i}$ is the weight matrix between the $(k-1)$-th hidden layer and the k-th hidden layer, in whose $w_{ij}^{(k)}$ is the weight from the j-th neuron in the $(k-1)$-th hidden layer to the i-th neuron in the k-th hidden layer and N_k is the number of the neurons in the k-th hidden layers, $b^{(k)}$ is the bias vector of the k-th hidden layer.

The basic block diagram of the neural network is shown in Figure 4.

Suppose there are n train samples (x_i, y_i), where $x_i \in R^n$ and $y_i \in R^m$ are the input vector and the label vector, respectively. For the given neural network, the global error function, i.e., the sum of the squared differences of all the training samples, is defined by

$$E(w) \coloneqq E\left(W^{(k)}, b^{(k)}\right) = \frac{1}{2n}\sum_{i=1}^{n}\left\|x_i^{(L)} - y_i\right\|^2, \tag{11}$$

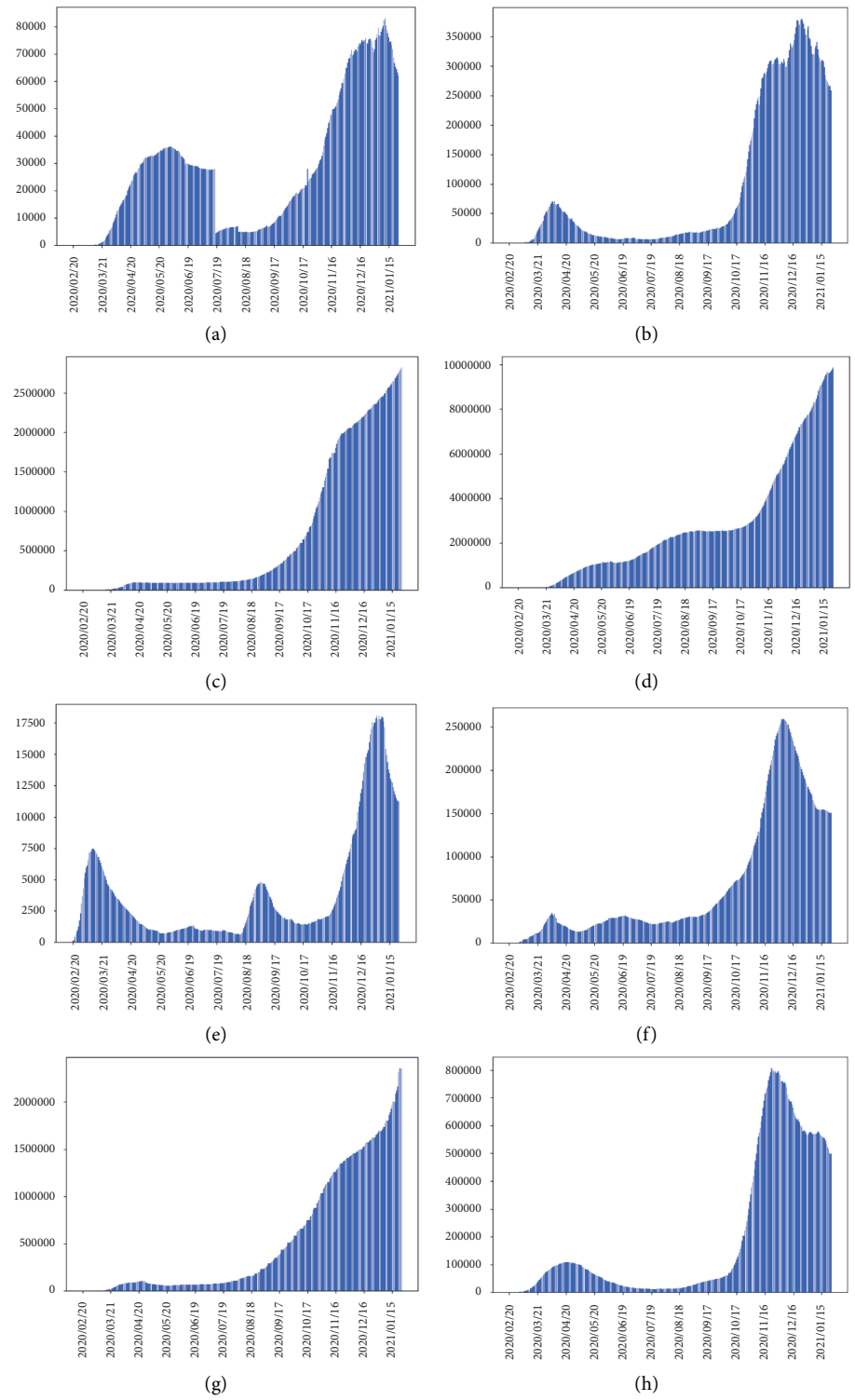

FIGURE 3: Trends of the numbers of existing confirmed cases in eight countries. (a) Canada (b) Germany (c) France (d) The United States (e) South Korea (f) Iran (g) Spain and (h) Italy.

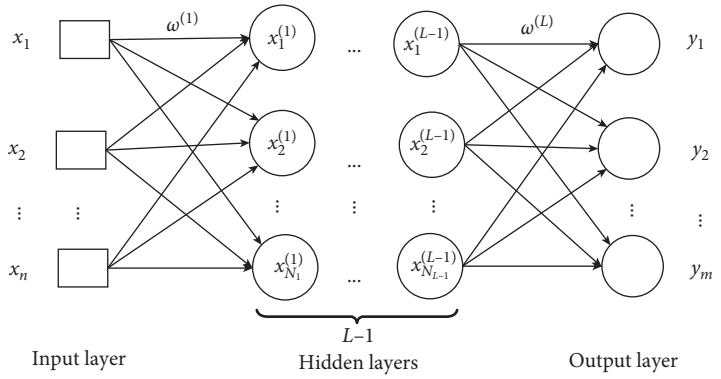

Figure 4: Diagram of the neural network.

where w is the set of all parameters of the neural network, including the weight matrix $W^{(k)}$ and biases $b^{(k)}$ are attached to the neural network, $x_i^{(L)}$ is the actual output in the output layer associated with the input vector x_i.

Mathematically, the problem of training a neural network can be expressed as finding the minimum w^* to minimize the global error function $E(w)$, that is,

$$\min E(w) = E\left(W^{(k)}, b^{(k)}\right). \tag{12}$$

Unfortunately, there is no guarantee that the objective function $E(w)$ is convex function with respect to $W^{(k)}$ and $b^{(k)}$. This easily leads to the risk of falling into local minima.

In general, one can apply the gradient descent method or other optimization methods to minimize the objective function. The conjugate gradient method is an efficient gradient descent algorithm for solving large-scale linear equations and nonlinear optimization problems [31]. It not only uses the first derivative information, but also overcomes the slow convergence of the steepest descent method and avoids the need to store and calculate the Hesse matrix for the Newton method for calculating its inverse matrix.

In this paper, we consider the neural network based on the scaled conjugate gradient method [30]. For the given neural network, the scaled conjugate gradient method generates a sequence of weights $\{w_k\}$ from the following iterative formula, i.e.,

$$w_{k+1} = w_k + \eta_k d_k,\ k = 0, 1, \ldots, \tag{13}$$

where w_0 is a given initial weight vector, $\eta_k > 0$ is the learning rate, and d_k is the descent search direction of the k th iteration defined by

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_{k-1} d_{k-1}, & k \geq 1, \end{cases} \tag{14}$$

where $g_k = \nabla E(w_k)$ and $\beta_{k-1} \in R$. In the literature, there have been proposed several choices for β_k which give rise to distinct conjugate gradient methods with quite different computational efficiency and theoretical properties [31].

The search direction of each iteration of the CGNN is the conjugate combination of the negative gradient direction and the search direction of the previous iteration, and it therefore has the advantages of low storage, few calculations, and high stability. The choice for β_k used in the scaled conjugate gradient is given by

$$\beta_{k-1} = \frac{\left|-g_k\right|^2 - g_k g_{k-1}}{d_{k-1}^T g_{k-1}}. \tag{15}$$

For a more detailed discussion of the CGNN, we refer to [30, 32–35] and the references within.

3. Global COVID-19 Epidemic Prediction and Analysis

The algorithms and models used in this paper are developed in Windows 10 with Python 3.6.0 and MATLAB R2018a, the SVR regression model, and Least Absolute Shrinkage and Selection Operator (LASSO) model are imported from the SVM class of sklearn python library and LASSO class of sklearn python library, respectively. The fitting methods and CGNN model are developed in MATLAB.

3.1. Optimal Fitting Function. It is well-known that the fitting function plays an important role in the accuracy of prediction. Some well-known fitting functions have been considered in the literature, including the four-parameter polynomial function, the normal distribution function, the three-parameter exponential function, the three-parameter hyperbolic function, the two-parameter power function and the four-parameter logical function [11, 22, 23, 25]. The details can be found in Table 1.

Through observation and experimentation, we finally choose the two-parameter power function $f_5(t)$ as the fitting function for fitting c_t. It is the same fitting function to the fitting function considered in [22].

3.2. Dataset. The COVID-19 data repository (https://github.com/CSSEGISandData/COVID-19, accessed on 24 January 2022) used in the study was obtained from the Johns Hopkins University Center for Systems Science and Engineering. In this paper, we consider the COVID-19 epidemic data from eight countries including Canada, Germany, France, the United States, South Korea, Iran, Spain and Italy. The starting and ending dates are shown in Table 2.

Table 1: Fitting functions.

Fitting function	References
$f_1(t) = \alpha_1 + \alpha_2 t + \alpha_3 t^2 + \alpha_4 t^3$	[23]
$f_2(t) = (1/\sqrt{2\pi}\,\sigma)\exp(-((t-\mu)^2/2\sigma^2))$	New
$f_3(t) = \alpha_1 \exp(\alpha_2 t) + \alpha_3$	[23]
$f_4(t) = (\alpha_1 + t/\alpha_2 + \alpha_3 t)$	[25]
$f_5(t) = ut^{v-1}$	[22]
$f_6(t) = (\alpha_1/(1-\alpha_2\exp(-\alpha_3 t)) + \alpha_4)$	[11]

Table 2: Date of COVID-19 outbreak data start and end.

Country	Start of dataset	End of dataset
Canada	2020/01/28	2020/04/07
Germany	2020/01/28	2020/04/07
France	2020/01/28	2020/04/07
The United States	2020/01/28	2020/04/07
South Korea	2020/01/28	2020/04/07
Iran	2020/02/19	2020/04/07
Spain	2020/02/01	2020/04/07
Italy	2020/01/31	2020/04/07

3.3. Optimal Sliding Window Period. To avoid inflection point prediction lag due to an excessively large sliding window period, we choose the optimal sliding window period from the integer set $\{1, \ldots, 7\}$, i.e., the maximum sliding window period does not exceed one week [22]. Based on the available global COVID-19 epidemic data, we calculate the dynamic transmission rate every k days, and use this value to replace the dynamic transmission rate for these k days. This strategy can reduce the volatility of the data, and overcome the lack of training data.

In this paper, we use the MAE and RMSE as the evaluation indicators for selecting the best sliding window period, and they are defined by

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|c_t - \widehat{c}_t\right|. \tag{16}$$

and

$$RMSE = \left[\frac{1}{N}\sum_{t=1}^{N}\left(c_t - \widehat{c}_t\right)^2\right]^{(1/2)}, \tag{17}$$

respectively, where $\widehat{c}_t$ is the predictive value of the dynamic transmission rate at time t, and N is the length of the time series of the predictive value.

The steps for selecting the optimal sliding window period are provided as follows:

Input: numbers of cumulative confirmed cases $L(t)$, cumulative deaths $D(t)$, and cumulative cures $K(t)$, $t = 1, 2, \ldots, N$

Output: optimal sliding window period k

Step 1: calculate the number of existing cases from Equation (1), $t = 1, \ldots, N$.

Step 2: calculate the dynamic transmission rate c_t using

$$c_t = 1 + \frac{1}{k}\ln\frac{N(T)}{N(T-k)}, \tag{18}$$

under different values of k, where $t \in [T-k, T-1]$, $T = sk$, $s = 1, 2, \ldots$, $k \in \{1, 2, \ldots, 7\}$.

Step 3: divide the data set c_t into the training data set and the testing data set, and the ratio of the training set to the testing set is 7 : 3.

Step 4: fit the training data set based on $f_5(t)$ under different sliding window periods $k, k = 1, 2, \ldots, 7$.

Step 5: calculate the predicted values obtained different sliding window periods $k, k = 1, 2, \ldots, 7$.

Step 6: calculate the MAE and RMSE of each predicted value under different sliding window periods $k, k = 1, 2, \ldots, 7$.

Step 7: calculate the average MAE and RMSE under different sliding window periods, and then select the optimal sliding window period k. Connect the training set and testing set to make it a fitting set, and refit the fitting set under the selected optimal sliding window period.

In what follows, we list the optimal fitting parameters of the fitting function $f_5(t)$, the optimal sliding window period and its evaluation indicators in eight countries as shown in Table 3.

3.4. Global COVID-19 Epidemic Prediction and Analysis Based on IDTRM. In order to better reveal the trend of c_t crossing the sliding window period, we consider the so-called IDTRM proposed in [22]. The calculation method of c_t in this paper is improved, as shown in (18). The value of the fitting parameter of the two-parameter power function $f_5(t)$ can reflect the severity of the epidemic in a given country to a certain extent, and can reflect the effect of the development and control of the COVID-19 epidemic [22]. According to the fitting curve expression obtained based on the optimal sliding window period, the solution of $\widehat{f}_5(t) = 1$ is the inflection point of the COVID-19 epidemic.

The estimated inflection points of the global COVID-19 epidemic based on IDTRM with the improved dynamic transmission rate c_t calculated from (18) is shown in Table 4.

From Table 4, through a comparison with the actual inflection point, the predicted result shows that the calculated dynamic transmission rate in this paper is suitable for the real situation, but the accuracy rate still needs to be improved.

For this purpose, we will use the CGNN to correct the residuals for improving the predictions accuracy, and to make the model more practical and instructive for the prediction of the global COVID-19 epidemic and new waves of the outbreaks. The details see the below.

3.5. Global COVID-19 Epidemic Prediction and Analysis Based on IDTRM_CGNN. To improve the prediction accuracy of IDTRM, we use IDTRM_CGNN to predict the global COVID-19 epidemic. In view of the good self-learning abilities of neural networks, we choose the CGNN to train the early residual set for obtaining the calibrated and predicted residual sets.

Table 3: Fitting parameters, optimal sliding window period and evaluation indicators.

Country	u	v	k	MAE	RMSE
Canada	1.3241	0.941	3	0.0241	0.0246
Germany	1.9939	0.8325	2	0.0450	0.0526
France	1.8440	0.8519	6	0.0115	0.0138
The United States	1.4116	0.926 4	5	0.0363	0.0376
South Korea	1.8040	0.8288	4	0.0220	0.0239
Iran	1.8880	0.8426	5	0.0228	0.0255
Spain	2.3148	0.7770	1	0.0150	0.0201
Italy	2.1015	0.8060	5	0.0182	0.0195

Table 4: Actual and estimated inflection points.

Country	Actual inflection point (A.I.P)	Estimated inflection point (E.I.P)
Canada	2020/05/31	2020/08/01
Germany	2020/04/08	2020/04/17
France	2020/04/16	2020/04/26
The United States	2020/05/31	2020/05/27
South Korea	2020/03/12	2020/03/19
Iran	2020/04/05	2020/04/16
Spain	2020/04/26	2020/04/08
Italy	2020/04/20	2020/04/07

The main steps of IDTRM_CGNN are as follows:

Input: numbers of cumulative confirmed cases $L(t)$, cumulative deaths $D(t)$, and cumulative cures $K(t)$, $t = 1, 2, \ldots, N$

Output: estimated inflection points and their corresponding existing confirmed cases

Step 1: calculate the number of existing confirmed cases $N(t)$ from equation (1), $t = 1, \ldots, N$.

Step 2: choose $f_5(t)$ as the optimal fitting function, see Section 3.1.

Step 3: calculate the dynamic transmission rate c_t using equation (18)

Step 4: choose the optimal sliding window period k^*, see Section 3.3.

Step 5: calculate the dynamic transmission rate $\widehat{c}_t$ by means of $f_5(t)$.

Step 6: calculate the residual of the dynamic transmission rate c_t from

$$\delta_t = c_t - \widehat{c}_t, \quad t = 1, 2, \ldots, N. \tag{19}$$

Step 7: predict the residual of the dynamic transmission rate δ_t' by means of the CGNN.

The residual sequence $\{\delta_t\}$ of the dynamic transmission rate is divided into the training set and the test set as the input layer variables of the neural network according to a ratio of 7 : 3; the hidden layer contains 2 layers in our experiments; the initial weight vector w_0 in the scaled conjugate gradient method is randomly initialized; the hidden layer activation function of the neural network selects the double curve tangent function, i.e.,

$$\tan \mathrm{h}x = \frac{\sin \mathrm{h}x}{\cos \mathrm{h}x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \tag{20}$$

The activation function of the output layer is an identity; the modified residual sequence $\{\delta_t'\}$ and the predicted residual sequence of the dynamic transmission rate are output. All the experiments are repeated 100 times to demonstrate the robustness of the method.

Step 8: calculate the corrected dynamic transmission rate using

$$c_t' = \widehat{c}_t + \delta_t', \tag{21}$$

which is equivalents to

$$\widehat{f}_5'(t) = \widehat{f}_5(t) + \delta_t'. \tag{22}$$

Step 9: predict the inflection point t^* using $\widehat{f}_5'(t) = 1$.

Step 10: predict the number of existing confirmed cases at time $t\,(t = N + 1, N + 2, \ldots)$ using

$$\widehat{N}(t) = \frac{1}{k}\sum_{i=1}^{k} \widehat{N}(t-i)\exp\{i(c_t' - 1)\}, \tag{23}$$

where $\widehat{N}(t)$ represents the estimated numbers of existing confirmed cases at time t. When $N(t-i)$ is unknown, $\widehat{N}(t-i)$ is used to instead of it [22].

To verify the robustness and generalization ability of the model, the residuals of the fitted curve and the dynamic transmission rate obtained are used to correct and predict the residuals in combination with the CGNN. The prediction results of the inflection point of the global COVID-19 epidemic based on IDTRM_CGNN are shown in Figure 5.

Figure 5 compares the prediction results regarding the inflection point of the COVID-19 epidemic before and after

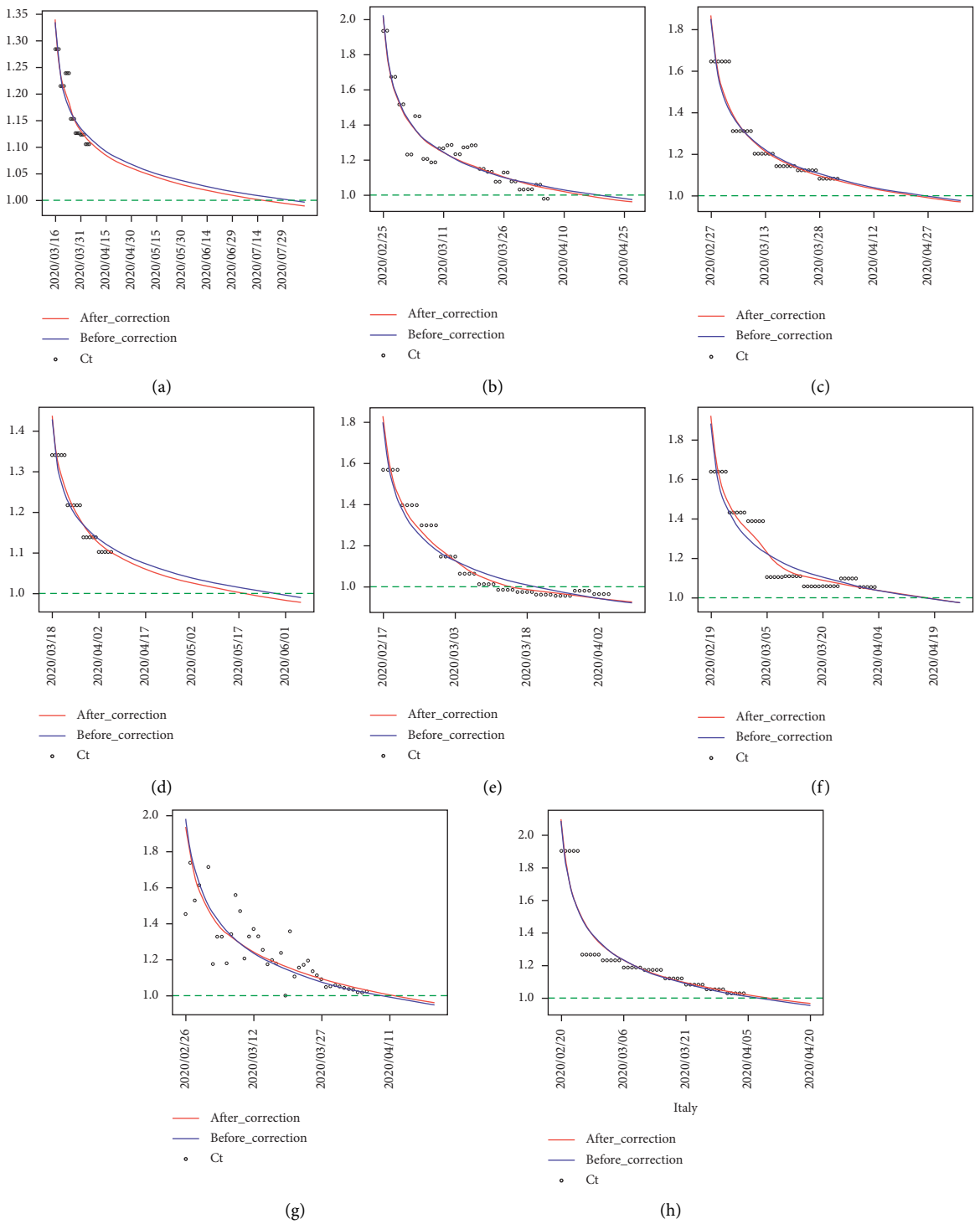

Figure 5: Prediction results based on IDTRM_CGNN. (a) Canada (b) Germany (c) France (d) The United States (e) South Korea (f) Iran (g) Spain and (h) Italy.

using the CGNN in Canada, Germany, France, the United States, South Korea, Iran, Spain and Italy. The overall prediction effect of the new model is better than those of some other existing methods, and it can effectively capture the fitting curve with smaller errors while maintaining the original trend. In order to demonstrate the advantages of the model, we use LASSO [36], SVR [37] and DTRM [22] to process the residuals, In addition, we also add the sliding window period steps to DTRM considered in [22], other unchanged, and the date of inflection point is calculated.

Table 5 provides the predictive inflection point of the COVID-19 epidemic based on LASSO, SVR, DTRM and IDTRM_CGNN. As can be seen from Table 5, our model has a good correction effect on the original curve which deviates

Table 5: The inflection point.

Country	A.I.P	IDTRM_CGNN	LASSO	SVR	DTRM [22]
Canada	2020/05/31	2020/07/16	2020/08/04	2020/07/26	2020/06/14
Germany	2020/04/08	2020/04/14	2020/04/17	2020/04/17	2020/04/21
France	2020/04/16	2020/04/22	2020/04/25	2020/04/24	2020/05/29
The United States	2020/05/31	2020/05/17	2020/05/27	2020/05/21	2020/05/08
South Korea	2020/03/12	2020/03/14	2020/03/18	2020/03/14	2020/03/20
Iran	2020/04/05	2020/04/16	2020/04/09	2020/04/17	2020/04/15
Spain	2020/04/26	2020/04/11	2020/04/13	2020/04/09	2020/04/16
Italy	2020/04/20	2020/04/10	2020/04/07	2020/04/07	2020/04/20

Table 6: Existing confirmed cases accumulated in 7 days.

Country	ALL(RECC)	ALL(PECC)	MAPE
Canada	108211	117414	8.62 %
Germany	451867	501200	11.42 %
France	608544	522295	14.06 %
The United States	3123671	2902818	7.11 %
South Korea	21416	18928	11.58 %
Iran	183693	247453	37.85 %
Spain	601975	599185	0.99 %
Italy	690617	631831	8.40 %

greatly. For example, the error of the inflection point of the COVID-19 epidemic in Canada has been corrected by half a month. In addition, other curves that deviated less sharply were also corrected to varying degrees, most of which were closer to the true inflection point. Compared with those of DTRM and other statistical models, the predicted inflection points of the COVID-19 epidemic based on IDTRM_CGNN in most countries are closer to reality, indicating that IDTRM after residual correction has higher accuracy and robustness.

In the meantime, IDTRM_CGNN is used to predict the daily number of confirmed cases for 7 days after April 7, 2020 (2020/04/08–2020/04/15), and the mean absolute percentage error (MAPE) is used as the performance index, which is defined by

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|N(t) - \widehat{N}(t)|}{N(t)} \times 100\%. \tag{24}$$

In general, we can divide the predictive ability of the model into four levels: high prediction (the error rate is between 0%–10%), good prediction (the error rate is between 10%–20%), feasible prediction (the error rate is between 20%–50%) and poor prediction accuracy (the error rate > 50%).

The real number of existing confirmed cases accumulated (ALL(RECC)) in 7 days, the predicted value of the number of existing confirmed cases (ALL(PECC)) in 7 days and the corresponding MAPE are shown in Table 6.

Table 6 shows that all MAPEs are within a suitable accuracy range. It is worth mentioning that our model achieves high prediction accuracies for Canada, the United States, Spain and Italy, in which their MAPEs are 8.62%, 7.11%, 0.99% and 8.40%, respectively. In addition, the average MAPE for the epidemic prediction in all countries is 10.81 %. This further validates the effectiveness of IDTRM_CGNN.

4. Conclusions and remarks

Aiming at the shortcomings of the traditional SIR and SEIR models in which the basic infection number is difficult to accurately estimate, this paper constructs IDTRM_CGNN to predict the inflection points and the corresponding number of existing confirmed cases of the COVID-19 epidemic in eight countries. The numerical results show that the model proposed in this paper has higher prediction accuracy and robustness. Furthermore, our model can also provide a certain reference value for countries around the world to effectively predict and grasp the development trend of the upcoming wave of the global COVID-19 epidemic.

However, the treatment of various complex influencing factors in this paper is relatively weak, and it is only applicable to countries with a monotonically decreasing trend of the dynamic transmission rate. Due to the characteristics of the fitting function, in cases with few data, IDTRM_CGNN is prone to overfitting.

The future research direction of this paper: Whether IDTRM_CGNN is universal for other infectious diseases is a topic worthy of continuing research. In addition, studying the use of rolling prediction technology to characterize the epidemic in real time, and performing online correction and updating of the dynamic transmission rate with the help of data assimilation methods are also future research directions of this article.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no potential conflicts of interest in this study.

Authors' Contributions

All the authors have seen and approved the manuscript to be published.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 11971302), the National Statistical Science Research Project of China (no. 2020LY067), and National Innovation and Entrepreneurship Training Program for College Students (no. 202110856037).

References

[1] N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, and A. C. Ghani, "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand," *British Medical Journal*, vol. 82, no. 5, pp. 201–157, 2020.

[2] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study," *The Lancet*, vol. 395, no. 10225, pp. 689–697, 2020.

[3] C. Caetano, M. L. Morgado, P. Patricio, J. F. Pereira, and B. Nunes, "Mathematical modelling of the impact of non-pharmacological strategies to control the COVID-19 epidemic in Portugal," *Mathematics*, vol. 9, no. 10, 2021.

[4] R. Ghostine, M. Gharamti, S. Hassrouny, and I. Hoteit, "An extended SEIR model with vaccination for forecasting the covid-19 pandemic in Saudi Arabia using an ensemble kalman filter," *Mathematics*, vol. 9, no. 6, 2020.

[5] Y. Xiao, X. Feng, Z. Xu, X. Wang, and S. Tang, "When will be the resumption of work in Wuhan and its surrounding areas during COVID-19 epidemic? A data-driven network modeling analysis," *Scientia Sinica Mathematica*, vol. 50, no. 7, pp. 969–978, 2020, In Chinese.

[6] Z. Yang, Z. Zeng, K. Wang, S. Wong, and J He, "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 2, pp. 165–174, 2020.

[7] D. Li, J. Lv, G. Botwin, J. Braun, and D. P. B. Mcgovern, "Estimating the Scale of COVID-19 Epidemic in the United States: Simulations Based on Air Traffic Directly from Wuhan, China," *MedRxiv*, 2020.

[8] Z. Liu, P. Magal, and G. Webb, "Predicting the number of reported and unreported cases for the COVID-19 epidemics in China, South Korea, Italy, France, Germany and United Kingdom," *Journal of Theoretical Biology*, vol. 509, Article ID 110501, 2021.

[9] S. W. Hermanowicz, "Simple Model for COVID-19 Epidemics-Back-Casting in China and Forecasting in the US," *MedRxiv*, 2020.

[10] H. Ankarali, S. Ankarali, H. Caskurlu, Y. Cag, and H. Vahaboglu, "A statistical modeling of the course of COVID-19 (SARS-CoV-2) outbreak: a comparative analysis," *Asia-Pacific Journal of Public Health*, vol. 32, no. 4, pp. 157–160, 2020.

[11] Z. Liu, "Uncertain growth model for the cumulative number of COVID-19 infections in China," *Fuzzy Optimization and Decision Making*, vol. 20, no. 2, pp. 229–242, 2021.

[12] D. Parbat and M. Chakraborty, "A python based support vector regression model for prediction of COVID19 cases in India," *Chaos, Solitons & Fractals*, vol. 138, Article ID 109942, 2020.

[13] C. Yesilkanat, "Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm," *Chaos, Solitons & Fractals*, vol. 140, Article ID 110210, 2020.

[14] S. F. Ardabili, A. Mosavi, P. Ghamisi, and F. Ferdinand, A. R. Varkonyi-Koczy, U. Reuter, T. Rabczuk, and P. M. Atkinson, COVID-19 outbreak prediction with machine learning," *Algorithms*, vol. 13, no. 10, 2020.

[15] F. Rustam, A. A. Reshi, A. Mehmood, S. Ullah, and G. S. Choi, "COVID-19 future forecasting using supervised machine learning models," *IEEE Access*, vol. 8, no. 1, Article ID 101489, 2020.

[16] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons & Fractals*, vol. 140, Article ID 110212, 2020.

[17] A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, "Deep learning methods for forecasting COVID-19 time-series data: a comparative study," *Chaos, Solitons & Fractals*, vol. 140, Article ID 110121, 2020.

[18] H. Abbasimehr and R. Paki, "Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization," *Chaos, Solitons & Fractals*, vol. 142, Article ID 110511, 2021.

[19] A. Bezerra and E. Santos, "Prediction the daily number of confirmed cases of COVID-19 in Sudan with ARIMA and holt winter exponential smoothing," *International Journal of Development Research*, vol. 10, no. 8, Article ID 39408, 2020.

[20] M. Maleki, M. R. Mahmoudi, M. H. Heydari, and K. -H. Pho, "Modeling and forecasting the spread and death rate of coronavirus (COVID-19) in the world using time series models," *Chaos, Solitons & Fractals*, vol. 140, Article ID 110151, 2020.

[21] N. E. Huang and F. Qiao, "A data driven time-dependent transmission rate for tracking an epidemic: a case study of 2019-nCoV," *Science Bulletin*, vol. 65, no. 6, pp. 425–427, 2020.

[22] Y. Hu, Y. Liu, L. Wu et al., "A dynamic transmission rate model and its application in epidemic analysis," *Operations Research Transactions*, vol. 24, no. 3, pp. 27–42, 2020, In Chinese.

[23] Y. Hu, J. Kong, L. Yang et al., "A dynamic growth rate model and its application in global COVID-19 epidemic analysis," *Acta Mathematicae Applicatae Sinica*, vol. 43, no. 2, pp. 452–467, 2020, In Chinese.

[24] Y. Zhang and J. Li, "Prediction and analysis of propagation of novel coronavirus pneumonia epidemic based on SIR model," *Journal of Anhui University of Technology*, vol. 37, no. 1, pp. 94–101, 2020, In Chinese.

[25] X. Xie, K. Luo, Y. Zhang et al., "Nonlinear combinational dynamic transmission rate model and COVID-19 epidemic analysis and prediction in China," *Operations Research Transactions*, vol. 25, no. 1, pp. 17–30, 2020, In Chinese.

[26] X. Xie, K. Luo, Z. Yin, and G. Wang, "Nonlinear combinational dynamic transmission rate model and its application in global COVID-19 epidemic prediction and analysis," *Mathematics*, vol. 9, no. 18, 2021.

[27] S. Layne, J. Hyman, D. Morens, and J. Taubenberger, "New coronavirus outbreak: framing questions for pandemic

prevention," *Science Translational Medicine*, vol. 12, no. 534, Article ID eabb1469, 2020.

[28] J. Guan, "The Parameter Estimation and Bias Correction of Short-Term Interest Rate Model," *Southwestern University of Finance and Economic*, Doctoral Dissertation, In Chinese, 2014.

[29] Z. Waszczyszyn, *Fundamentals of Artificial Neural Networks,* MIT Press, Cambridge, MA, USA, 1999.

[30] M. F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525–533, 1993.

[31] W. Sun and Y. Yuan, *Optimization Theory and Methods*, Springer US, Philadelphia, NY, USA, 2006.

[32] E. M. Johansson, F. U. Dowla, and D. M. Goodman, "Backpropagation learning for multilayer feed-forward neural networks using the conjugate gradient method," *International Journal of Neural Systems*, vol. 2, no. 4, pp. 291–301, 1991.

[33] X. Zhang, S. L. Broschat, and P. J. Flynn, "Inverse imaging of the breast using a conjugate gradient-neural network technique," *Journal of the Acoustical Society of America*, vol. 103, no. 5, pp. 2792-2793, 1998.

[34] J. Wang, W. Wu, and J. M. Zurada, "Deterministic convergence of conjugate gradient method for feedforward neural networks," *Neurocomputing*, vol. 74, no. 14-15, pp. 2368–2376, 2011.

[35] E. Ioannis and P. Panagiotis, "A new conjugate gradient algorithm for training neural networks based on a modified secant equation," *Applied Mathematics and Computation*, vol. 221, pp. 491–502, 2013.

[36] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[37] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.