



## SIRVD-DL: A COVID-19 deep learning prediction model based on time-dependent SIRVD

Zhifang Liao<sup>a</sup>, Peng Lan<sup>a</sup>, Xiaoping Fan<sup>b,\*</sup>, Benjamin Kelly<sup>c</sup>, Aidan Innes<sup>c</sup>, Zhining Liao<sup>c,\*\*</sup>

<sup>a</sup> School of Computer Science and Engineering, Central South University, Changsha, 410075, China

<sup>b</sup> Hunan University of Finance and Economics, Changsha, China

<sup>c</sup> Nuffield Health Research Group, Nuffield Health, Ashley Avenue, Epsom, Surrey, KT18 5AL, UK



### ARTICLE INFO

**Keywords:**

COVID-19

SIRVD Model

Vaccination

Deep learning LSTM

### ABSTRACT

COVID-19 is one of the biggest challenges that human beings have faced recently. Many researchers have proposed different prediction methods for establishing a virus transmission model and predicting the trend of COVID-19. Among them, the methods based on artificial intelligence are currently the most interesting and widely used. However, only using artificial intelligence methods for prediction cannot capture the time change pattern of the transmission of infectious diseases. To solve this problem, this paper proposes a COVID-19 prediction model based on time-dependent SIRVD by using deep learning. This model combines deep learning technology with the mathematical model of infectious diseases, and forecasts the parameters in the mathematical model of infectious diseases by fusing deep learning models such as LSTM and other time prediction methods. In the current situation of mass vaccination, we analyzed COVID-19 data from January 15, 2021, to May 27, 2021 in seven countries – India, Argentina, Brazil, South Korea, Russia, the United Kingdom, France, Germany, and Italy. The experimental results show that the prediction model not only has a 50% improvement in single-day predictions compared to pure deep learning methods, but also can be adapted to short- and medium-term predictions, which makes the overall prediction more interpretable and robust.

### 1. Introduction

COVID-19 is an infectious epidemic caused by a new Coronavirus virus(SARS-CoV-2). Since the outbreak of COVID-19 at the end of 2019, it has spread throughout the world. As of June 15, 2021, there are 176,275,966 confirmed cases and 3,812,436 deaths worldwide. COVID-19 has had a huge impact on the global economy, human health, and daily life. In response to this pandemic crisis, researchers and policy-makers around the world have been working hard to study and develop countermeasures and solutions against COVID-19 to control the epidemic and reduce its impact on human health and the economy [1]. There have been a large number of studies on COVID-19 by modeling the spread of the virus and predicting the number of people infected by the virus [2].

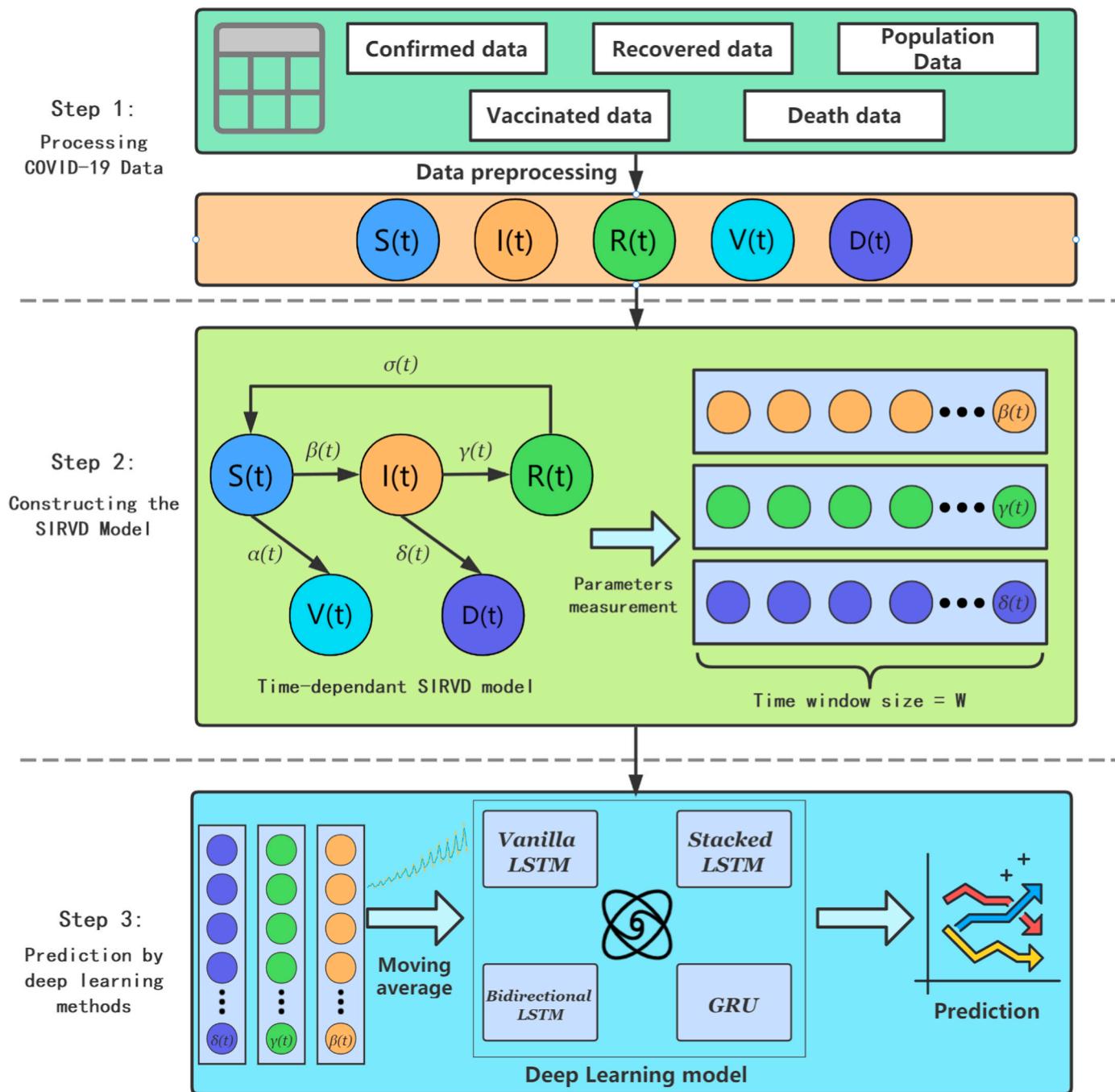
Since the mathematical theory of epidemiology proposed by Kermack in 1927, mathematical model of infectious diseases has been used as an important tool for epidemiological feature analysis and

transmission analysis [3]. Since the epidemic of COVID-19, several mathematical models have been used to predict the epidemic. Among them, the most used models are SIR(Susceptible, Infected, and Recovered) and SEIR(Susceptible, Exposed, Infected, and Recovered) models, and most studies are based on these two models for corresponding modifications and adjustments. Liao et al. [4] proposed a generalized adaptive SIR prediction model based on time window characterized by introducing a time window mechanism for dynamic data analysis, which can dynamically measure the basic infection number and exponential growth rate. Pushpendra et al. [5] proposed a generalized SIR (GSIR) model, which includes multi-day reported cases. The experimental evaluation on COVID-19 data in different countries shows that the model can provide continuous prediction and monitoring for the pandemic of COVID-19. Based on the SIR model, Cooper et al. [6] regarded the number of susceptible people as a variable, which provided a theoretical framework for the study of the spread of COVID-19 in the community. Hakimeh et al. [7] used Caputo derivatives to establish a

\* Corresponding author.

\*\* Corresponding author.

E-mail addresses: [zfliao@csu.edu.cn](mailto:zfliao@csu.edu.cn) (Z. Liao), [lanpeng5@qq.com](mailto:lanpeng5@qq.com) (P. Lan), [xpfan@mail.csu.edu.cn](mailto:xpfan@mail.csu.edu.cn) (X. Fan), [Benjamin.Kelly@nuffieldhealth.com](mailto:Benjamin.Kelly@nuffieldhealth.com) (B. Kelly), [Aidan.Innes@nuffieldhealth.com](mailto:Aidan.Innes@nuffieldhealth.com) (A. Innes), [zhining.liao@nuffieldhealth.com](mailto:zhining.liao@nuffieldhealth.com) (Z. Liao).



**Fig. 1.** The overall workflow of the SIRVD-DL predictive model.

Step 1: Data preprocessing. First, obtaining data sets of COVID-19 including the number of confirmed cases, recovered cases, deceased cases, and vaccination. Then performing data preprocessing and transform the data into the required format for the SIRVD-DL model.

Step 2: Constructing the SIRVD model. In this step, an existing SIRVD model is modified to adapt to the dynamic changes of time. Then we build the curve of parameter changes over time by measuring the model parameters.

Step 3: Deep learning model. Based on the measurement of model parameters obtained from the last step, the moving average method is used to smooth the data. Then a time series vector is constructed as the input. Model training and evaluation are carried out through deep learning methods such as LSTM and its variants to select the best model parameters. By using the model parameters, the SIRVD-DL model is built to predicting the number of COVID-19 infections.

fractional-order SIRD(Susceptible, Infected, and Deceased) mathematical model for the transmission of COVID-19. Through numerical simulation of different order derivatives, the second wave epidemic situation in Iran and Japan was predicted. Ramezani et al. [8] proposed a variant of the SEIRD(Susceptible, Infected, Recovered and Deceased) model to capture the non-linear behavior of the COVID-19 pandemic while accounting for asymptomatic infected individuals. Mathematical

models of infectious diseases can be used to evaluate the transmission process of epidemics, but the introduction of model parameters is based on many assumptions. Unknown parameters need to be estimated by model fitting, which makes the model more uncertain [9]. Moreover, for the long duration of COVID-19, the change of model parameters is related to many factors. It is a challenge for a single model to fit the real situation well [10]. In order to effectively predict COVID-19, more

**Table 1**

Sample data of global confirmed cases from Johns Hopkins University.

Province/State	Country/Region	Latitude	Longitude	1/22/20	1/23/20	...	5/30/21	5/31/21	...
Anhui	China	31.8257	117.2264	1	9	...	1004	1004	...
Beijing	China	40.1824	116.4142	14	22		1059	1059	
Chongqing	China	30.0572	107.8740	6	9		598	598	
Fujian	China	26.0789	117.9874	1	5		621	621	

**Table 2**

Our World In Data vaccination sample data.

location	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated
India	IND	2021-02-15	8516771	8418653	98118
India	IND	2021-02-16	8857341	8641002	216339
India	IND	2021-02-17	9186757	8844641	342116
India	IND	2021-02-18	9846523	9381591	464932
India	IND	2021-02-19	10449942	9693000	756942

advanced methods should be applied to consider various factors in the process of virus transmission.

As an emerging method, deep learning technology is widely used in the analysis and prediction of the COVID-19 epidemic. Vinay et al. [11] used the LSTM(long short-term memory) deep learning method for the first time to model the spread of infectious diseases in Canada to predict the severity of COVID-19. Smail et al. [12] used ARIMA(Autoregressive integrated moving average method), NARNN(Nonlinear autoregressive neural network), and LSTM methods to model confirmed cases of new coronary pneumonia, and the results of the study found that LSTM is the most accurate model. Khondoker et al. [13] studied four deep learning models: LSTM, GRU(Gated recurrent unit), CNN(Convolutional neural network), and MCNN(Multivariate convolutional neural network), the results show that the CNN is superior to other deep learning models in terms of verification accuracy and prediction consistency. Jayanthi et al. [14] used ARIMA, LSTM, Stacked LSTM, and prophet models to analyze and predict the global cumulative number of confirmed cases, death cases, and recovery cases. The results show that the Stacked LSTM algorithm has higher accuracy than other considered methods, and the error is less than 2%. From the above research, deep learning can provide reliable single-day prediction results. However, this kind of methods have two problems. The first one is that these methods by using pure numerical fitting cannot correctly capture the trend of epidemics in the spreading process [13]. The another one is that the temporal change pattern of the number of infected people is very simple. It is difficult for these methods to find right patterns of long-term changes, leading to predictions that are only valid for a short period. A model that can provide forecasts for longer periods is undoubtedly more meaningful for policy-making and strategic planning.

Considering that epidemic mathematical models have background knowledge of infectious diseases and deep learning technology has strong predictive capabilities, combining both technologies may improve the interpretability of deep learning methods and generate more robust predictions [15]. This paper proposes a COVID-19 prediction model – SIRVD-DL (Susceptible, Infected, Recovered, Vaccinated, and Deceased – Deep Learning) by using deep learning method based on time-dependent SIRVD to combine these two methods to make the model more explanatory and provide effective predictions. Firstly, this model modifies the existing SIRVD model of vaccination status to enable it to dynamically measure model parameters. The model parameters are smoothed by using the moving average method to deal with data noise and anomalies and predicted by the deep learning method. Finally, the predicted parameters are substituted into the SIRVD model to obtain the final predicted number of infections. We have conducted related numerical experiments and are interested in addressing the following three important research questions.

- RQ1: Is the time-dependant SIRVD-DL model reasonable and effective?
- RQ2: How is the prediction performance of the proposed SIRVD-DL model? What is the improvement compared to single deep learning methods? How does the model perform in the short and medium-term forecasts?
- RQ3: Is the moving average method proposed in this paper effective for data smoothing? What is the difference between using this method and not using this method?

The rest of the paper is organized as follows: in the second section, we propose the SIRVD-DL prediction model. In the third section, we conducted numerical experiments and analyzed the experimental results to illustrate the effectiveness of our model. Then, in section 4, we made some discussions and suggestions. Finally, the last section is the summary of the paper.

## 2. Method

The SIRVD-DL prediction model is mainly divided into three parts: data preprocessing, constructing the SIRVD model, and deep learning model. The overall workflow of the model is shown in Fig. 1.

SIRVD-DL aims to evaluate the changes in the parameters of the epidemic to build the best model to predict the development trend of the epidemic. In the rest of this section, we will describe the contents of each part in detail.

### 2.1. Data preprocessing

#### 2.1.1. Data sources

Since the outbreak of COVID-19, many governments have released various public data sources and incorporated real-time observation data for the latest analysis and predictions. This article collects relevant research data from two data sources. One is a time-series data collected by the Johns Hopkins University System Science and Engineering Center (CSSE), which is widely used by many researchers, including the global cumulative confirmed cases, cumulative cured cases, and cumulative deaths. The data set provides variables such as country, province, longitude, latitude, and the number of cases corresponding to the date. Table 1 shows some sample data of global confirmed cases.

The second data source is from the website - *Our World In Data*. It is updated daily and includes data on confirmed cases, deaths, hospitalizations, testing, and vaccinations as well as other variables of potential interest. The data to be used in this article is the COVID-19 vaccination data [16] collected from official reports by the “Our World” data team. Table 2 shows some sample data of vaccination data. In the table, location represents the name of the country (or region within a country),

iso\_code is a three-letter country code, the date is the date of the observation, total vaccination is the total number of doses administered, and people\_vaccinated is the total number of people who received at least one vaccine Dose, people\_fully\_vaccinated is the total number of people who received all doses prescribed by the vaccination protocol.

### 2.1.2. Data transforming

After obtaining the above two data sets, we need to transform the data into the format required by the SIRVD model. The infected cases  $I(t)$  can be obtained by the formula:

$$I(t) = C(t) - (R(t) + D(t)) \quad (1)$$

In order to get the change of the number of susceptible people  $S(t)$ , we can use the following formula:

$$S(t) = N - (I(t) + R(t) + D(t) + V(t)) \quad (2)$$

Where  $S(t)$ ,  $I(t)$ ,  $R(t)$ ,  $V(t)$ , and  $D(t)$  are the functions of the changes of the population status with time, respectively representing the number of susceptible, infected, recovered, vaccinated, and deceased individuals in the population when the total population number is  $N$  at time  $t$ .  $C(t)$ ,  $R(t)$ ,  $D(t)$ , and  $V(t)$  can be obtained from the above-mentioned two datasets. The population  $N$  can be obtained from the data source provided by *Our World In Data*. It is all based on the last revision of the United Nations World Population Prospects. After preprocessing the data, the next step is to construct a time-dependant SIRVD model.

## 2.2. The time-dependant SIRVD model

### 2.2.1. The basic SIRVD epidemic model

The basic SIRVD model [17] is based on the epidemic model SIR. The SIRVD model describes the interaction of the virus with the host during transmission, and divides the population into 5 types: susceptible, infected, recovered, vaccinated, and deceased. SIRVD adds two new states based on the SIR model, vaccinated and deceased. Among them, vaccinated means individuals in the population who have been vaccinated against the disease and have the ability not to be infected, while deceased means the individual who died because of the disease [18]. The SIRVD model modified based on SIR can be represented by the following ordinary differential equations:

$$\frac{dS(t)}{dt} = -\frac{\beta I(t)S(t)}{N} + \sigma R(t) - \alpha S(t) \quad (3)$$

$$\frac{dI(t)}{dt} = \frac{\beta I(t)S(t)}{N} - \gamma I(t) - \delta I(t) \quad (4)$$

$$\frac{dR(t)}{dt} = \gamma I(t) - \sigma R(t) \quad (5)$$

$$\frac{dV(t)}{dt} = \alpha S(t) \quad (6)$$

$$\frac{dD(t)}{dt} = \delta I(t) \quad (7)$$

$$N = S(t) + I(t) + R(t) + V(t) + D(t) \quad (8)$$

Among them,  $\beta$  is the rate of infection, which is that the transmission from susceptible to infected;  $\gamma$  is the recovery rate, which is the transmission from infected to recovered;  $\delta$  is the death rate, which is the transmission from infected to deceased;  $\alpha$  is the vaccination rate, which is the transmission from susceptible to vaccinated; finally,  $\sigma$  is the susceptibility rate, which is the transmission from recovered to susceptible. It should be noted that the transmission process of the virus is described by the nonlinear term  $\frac{\beta I(t)S(t)}{N}$ , which represents the number of people transferred from the susceptible  $S$  to the infected  $I$  per unit time.

In the basic SIRVD model, the five parameters of infection rate  $\beta$ ,

recovered rate  $\gamma$ , death rate  $\delta$  vaccination rate  $\alpha$ , and susceptibility rate  $\sigma$  are assumed to be constant, which ignores their time-dependent characteristics because they are constantly changing in the process of virus transmission. In order to accurately and effectively predict the development trend of the disease, we propose a time-dependent SIRVD model, in which the five parameters are all functions of time  $t$ . This time-dependant SIRVD model can better track the spread of the disease, control, and predict future trends.

### 2.2.2. The time-dependent SIRVD epidemic model

The time-dependant SIRVD model regards the infection rate  $\beta$ , the recovery rate  $\gamma$ , the death rate  $\delta$ , the vaccination rate  $\alpha$ , and the susceptibility rate  $\sigma$  as functions that change with time  $t$ :  $\beta(t)$ ,  $\gamma(t)$ ,  $\delta(t)$ ,  $\alpha(t)$ , and  $\sigma(t)$ . The rewritten differential equations are as follows:

$$\frac{dS(t)}{dt} = -\frac{\beta(t)I(t)S(t)}{N} + \sigma(t)R(t) - \alpha(t)S(t) \quad (9)$$

$$\frac{dI(t)}{dt} = \frac{\beta(t)I(t)S(t)}{N} - \gamma(t)I(t) - \delta(t)I(t) \quad (10)$$

$$\frac{dR(t)}{dt} = \gamma(t)I(t) - \sigma(t)R(t) \quad (11)$$

$$\frac{dV(t)}{dt} = \alpha(t)S(t) \quad (12)$$

$$\frac{dD(t)}{dt} = \delta(t)I(t) \quad (13)$$

Among them, the five variables( $S(t)$ ,  $I(t)$ ,  $R(t)$ ,  $V(t)$ , and  $D(t)$ ) still satisfy Eq. (8). If we assume that the total population  $N$  is constant, then the sum of the increase or decrease of the state of each population is 0.

$$\frac{dS(t)}{dt} + \frac{dI(t)}{dt} + \frac{dR(t)}{dt} + \frac{dV(t)}{dt} + \frac{dD(t)}{dt} = 0 \quad (14)$$

Since the data of COVID-19 is updated daily, we can change Eqs. 9–13 into difference equations.

$$S(t+1) - S(t) = -\frac{\beta(t)I(t)S(t)}{N} + \sigma(t)R(t) - \alpha(t)S(t) \quad (15)$$

$$I(t+1) - I(t) = \frac{\beta(t)I(t)S(t)}{N} - \gamma(t)I(t) - \delta(t)I(t) \quad (16)$$

$$R(t+1) - R(t) = \gamma(t)I(t) - \sigma(t)R(t) \quad (17)$$

$$V(t+1) - V(t) = \alpha(t)S(t) \quad (18)$$

$$D(t+1) - D(t) = \delta(t)I(t) \quad (19)$$

Similarly, the five variables( $S(t)$ ,  $I(t)$ ,  $R(t)$ ,  $V(t)$ , and  $D(t)$ ) still satisfy Eq. (8). During the spread of COVID-19, the reinfection rate can be assumed to be approximately equal to zero because the human body produces antibodies against the virus [19]. Therefore, Eq. (15) can be rewritten as:

$$R(t+1) - R(t) = \gamma(t)I(t) - \sigma(t)R(t) \quad (20)$$

That is, the expression of  $\gamma(t)$  can be obtained:

$$\gamma(t) = \frac{R(t+1) - R(t)}{I(t)} \quad (21)$$

In the same way, the other two time-dependent parameters  $\alpha(t)$  and  $\delta(t)$  can be easily derived from the above difference equations Eq. (16) and Eq. (17).

$$\alpha(t) = \frac{S(t+1) - S(t)}{V(t)} \quad (22)$$

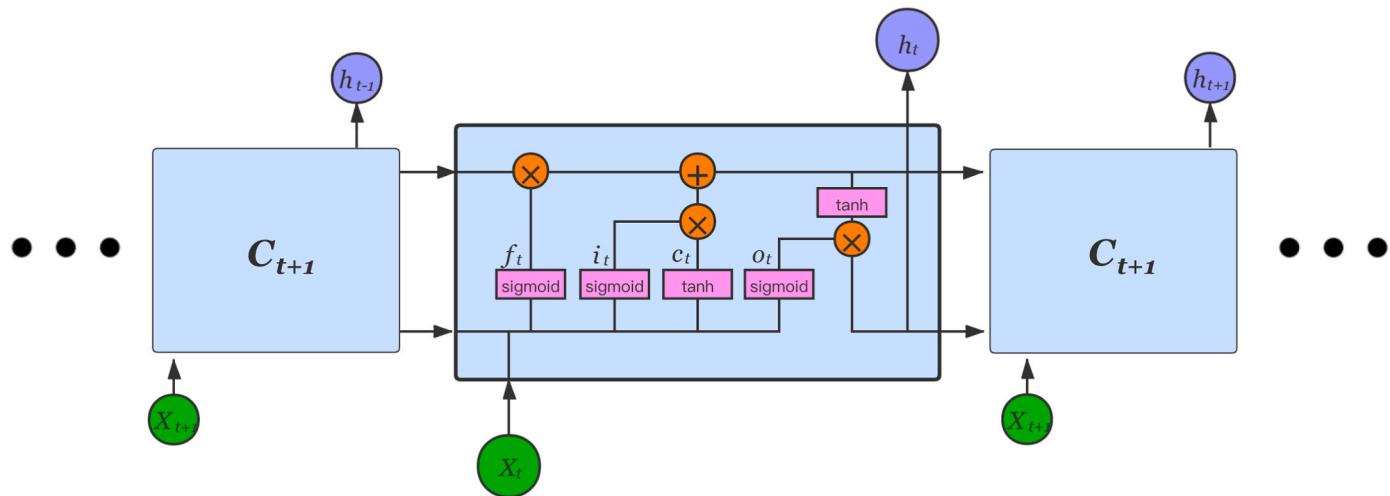


Fig. 2. Vanilla LSTM structure.

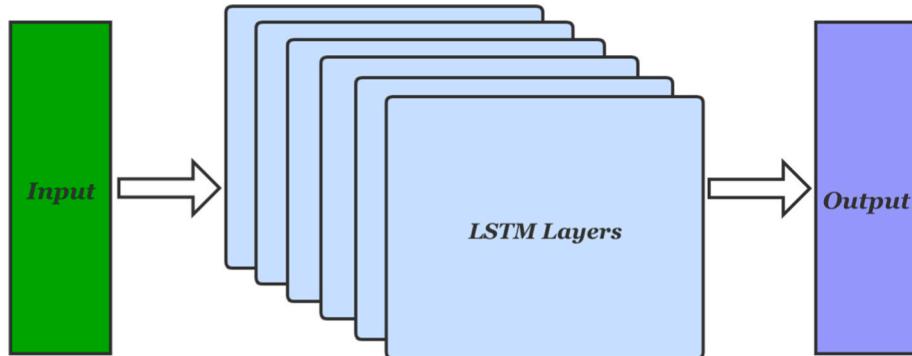


Fig. 3. Stacked LSTM structure.

$$\delta(t) = \frac{D(t+1) - D(t)}{I(t)} \quad (23)$$

After getting the rate of recovery and the rate of death, bring into Eq. (14), we can get the time-dependant parameter  $\beta(t)$  which is expressed by Eq. (24).

$$\beta(t) = \frac{(I(t+1) - I(t) + R(t+1) - R(t) + D(t+1) - D(t)) \times N}{I(t) \times S(t)} \quad (24)$$

### 2.3. Deep learning prediction model

Through the time-dependant SIRVD model, the measurement and evaluation values of the parameters can be obtained and arranged in time series. There is a problem that the curve of a time series of data - the number of infected people does not have a related sequence pattern [13]. To solve the problem, our model firstly predicts the estimated parameters to discover the development trend of the epidemic and then builds the time-dependant SIRVD model by using the values of predicted parameter. Deep learning technology is an effective time series prediction method [20]. In the next section, we will introduce some specific deep learning time series prediction methods, the prediction process of SIRVD-DL, and the evaluation metrics of the model.

#### 2.3.1. Deep learning time series prediction model

Deep learning (DL) has a good performance in time series data analysis and prediction, and can automatically learn the time correlation and structure of data, such as seasonality, periodicity, and trend [14]. The recurrent neural network is a kind of neural network which is

specially used to deal with sequence-structure data. It contains different time distributions of hidden states, which makes it possible to store a lot of information about the past. They are most commonly used in predictive applications because of their ability to process variable-length sequence data [21]. But the recurrent neural network is not able to overcome the vanishing gradient or gradient explosion, and because the only previous time step involves the hidden layer activation function, it can only store short-term memory, long-term historical information can not be passed [22]. In order to solve this problem, Hochreiter et al. proposed a long and short-term memory neural network LSTM. The model includes a memory storage unit and three logic gated units to control the flow of data, and can filter out which values to store, forget or erase, and solve the problem of vanishing gradient or gradient explosion in the long-term dependence of the recurrent neural network. There are many variants of LSTM, which are divided into Vanilla LSTM, Stacked LSTM, Bidirectional LSTM, and GRU according to their structure, which will be described in detail below.

#### 2.3.2. Vanilla LSTM

Vanilla LSTM is the most basic LSTM and the most widely used [22]. Fig. 2 shows the architecture of Vanilla LSTM. The rounded rectangle represents a Cell state, the green circle represents the input of the time series, and the purple circle represents the output of the hidden layer. The output of the Cell state at time  $t$  and the output of the hidden layer are linked to the next cell state, and the data flow in the cell State. The cell state contains a memory block of hidden units, which are used to control the flow of information from input to output ports. The first sigmoid function is the forget gate, which forgets the previous cell state

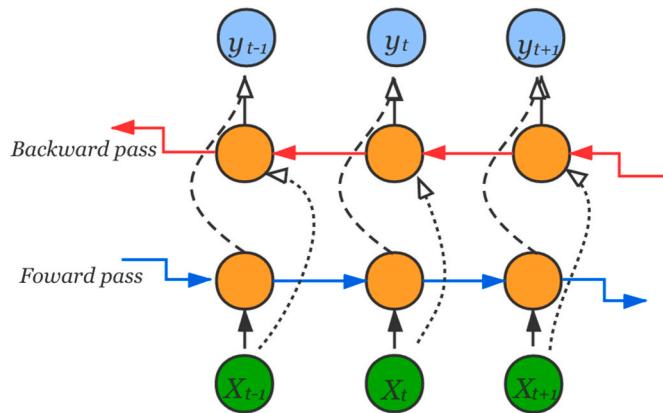


Fig. 4. Bidirectional LSTM structure.

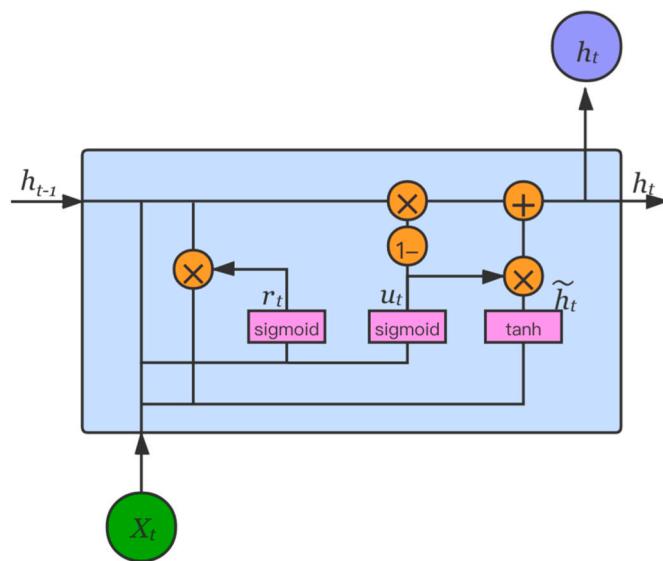


Fig. 5. GRU structure.

Table 3

The number of layers, the number of units, and the total number of parameters for each deep learning method.

Method	Parameter	Values
Vanilla LSTM	Number of layers	2
	Number of units	16
	Total Params	1297
BiDirectional LSTM	Number of layers	2
	Number of units	32
	Total Params	2595
Stacked LSTM	Number of layers	3
	Number of units	64
	Total Params	10705
GRU	Number of layers	3
	Number of units	64
	Total Params	8225

information. The next sigmoid and the first tanh function are input gates, which represent information stored in the cell state or information that should be forgotten. The last sigmoid function is the output gate, which determines the information passed to the next hidden state. Eq. 25–29 are the specific mathematical form of LSTM:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (25)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (26)$$

$$c_t = i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) + f_t c_{t-1} \quad (27)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (28)$$

$$h_t = o_t \tanh(c_t) \quad (29)$$

Among them,  $W$  represents the weight matrix,  $b$  represents the bias matrix, they exist in the forget gate  $f_t$ , input gate  $i_t$ , cell state  $c_t$  and output gate  $o_t$ .

### 2.3.3. Stacked LSTM

Stacked LSTM [23], also known as a deep LSTM, is an extension of the vanilla LSTM we described above. In Stacked LSTM, there are multiple hidden layers containing multiple memory cells. Multi-layer superposition increases the depth of the neural network. Each layer has some information and passes it to the next layer to form an accurate model with a higher level and deeper representation. The stacked LSTM structure is shown in Fig. 3. For each time step, it provides a separate output instead of providing a single output for all time steps.

### 2.3.4. Bidirectional LSTM

The standard RNN only processes the input in one direction and processes the information that it has in the future. This problem is solved by implementing the bidirectional topology of LSTM. Bidirectional LSTM [24] extracts complete time information at time  $t$  by considering past and future information. This method divides the hidden neurons of the standard RNN into a forward state and a backward state. The neurons in the forward state are not connected to the neurons in the backward state, and the neurons in the backward state are not connected to the neurons in the forward state. The basic structure of the three time steps of the two-way LSTM expansion is shown in Fig. 4. Without a reverse state, this structure is similar to a standard one-way RNN. With this structure, there is no need to include additional delays as in standard RNNs.

### 2.3.5. GRU

GRU(gated recurrent unit) can be described as a variant of LSTM [25], which is similar to LSTM. GRU is mainly used to solve the vanishing gradient problem in typical RNNs, thereby improving the learning of long-term dependencies in the network. Fig. 5 shows the structure of GRU. The GRU block also uses the tanh and sigmoid functions to calculate the necessary values. But unlike the LSTM block, the GRU block does not have a separate storage unit. There is no separate forget gate for this type of block, and the input/update gate is responsible for controlling the flow of information. Due to the difference in these two structures, it has fewer parameters and a simpler design, which ultimately makes it more computationally efficient and easier to train. In addition to the update gate, the GRU block has a reset gate. In a GRU block, four values are calculated: update gate, reset gate, candidate activation, and output activation. Each gate and candidate activation has its own weight and bias, and the current block input and the previous activation value are used as inputs for calculating these values. In the first step, the sigmoid function is used to calculate the value of the gate. Eq. 30–33 are the specific mathematical form of GRU:

$$u_t = \sigma(W_{xu}x_t + W_{hu}h_{t-1} + b_u) \quad (30)$$

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \quad (31)$$

$$\tilde{h}_t = \tanh(W_{xh}x_t + W_{hh}(r_t \odot h_{t-1}) + b_h) \quad (32)$$

$$h_t = u_t \odot \tilde{h}_t + (1 - u_t) \odot h_{t-1} \quad (33)$$

In this paper, Vanilla LSTM, Stacked LSTM, Bidirectional LSTM, and

**Table 4**

Hyperparameter combination for training.

Hyperparameter	Possible values
Epochs	{1000, 2000, 3000}
Batch size	{64, 128, 256}
Learning rate	{0.00001, 0.01, 0.1, 0.5}
Optimizer	{Adam, SGD, ADAGRAD}

GRU are all implemented using the Keras package (An open-source software library that provides a Python interface for artificial neural networks.) and the loss function is unified using MSE. Table 3 shows the number of layers, the number of units, and the total number of parameters for each method.

### 2.3.6. Predicting process

Based on the time-dependent SIRVD model, the model parameters in the form of time series can be predicted. Given the historical data of a period  $\{C(t), R(t), D(t), V(t), 0 \leq t \leq T - 1\}$ ,  $t$  represents the specific day, and the date range is  $T$ . The input of the time-dependant SIRVD model can be obtained through Eqs. (1) and (2):  $\{S(t), I(t), R(t), V(t), D(t), 0 \leq t \leq T - 1\}$ , then use Eq. (21-24) obtains the specific evaluation values of the model parameters:  $\{\beta(t), \gamma(t), \delta(t), \alpha(t), \sigma(t), 0 \leq t \leq T - 2\}$ .

We use the estimated model parameters  $\beta(t)$ ,  $\gamma(t)$ , and  $\delta(t)$  as the prediction targets of the deep learning method. Since the model parameters are constantly changing during the development of the epidemic, in order to measure this change, we construct a time window of size  $w$  as a dimension of the input time-series data into the deep learning model. Let the model parameter to be predicted as  $x$ , and  $x$  is one of the items in the set  $\{\beta(t), \gamma(t), \delta(t)\}$ , then the time series can be obtained as  $(x_0, x_1, x_2, \dots, x_{T-2})$ . Generally, the model parameters eval-

uated by SIRVD will have data noise, which will affect the subsequent use of deep learning methods to extract temporal features. To deal with this problem, a moving average with a small sliding window is used to smooth the model parameters. It divides the model parameter curve into two parts: baseline and residuals. Specifically, let the length of this small sliding window as  $W$  and the step length as 1. For each point  $x_t$ , the corresponding point on the baseline is  $x_t^*$ , and the value of  $x_t^*$  is the average value of  $(x_{t-W+1}, x_{t-W+2}, \dots, x_t)$ . The difference between  $x_t$  and  $x_t^*$  is called residual. Baseline B and residuals R can be calculated by the following formula.

$$x_t^* = \frac{1}{W} \sum_{i=1}^W x_{t-i+1} \quad (34)$$

$$B = (x_W^*, x_{W+1}^*, \dots, x_{T-2}^*) \quad (35)$$

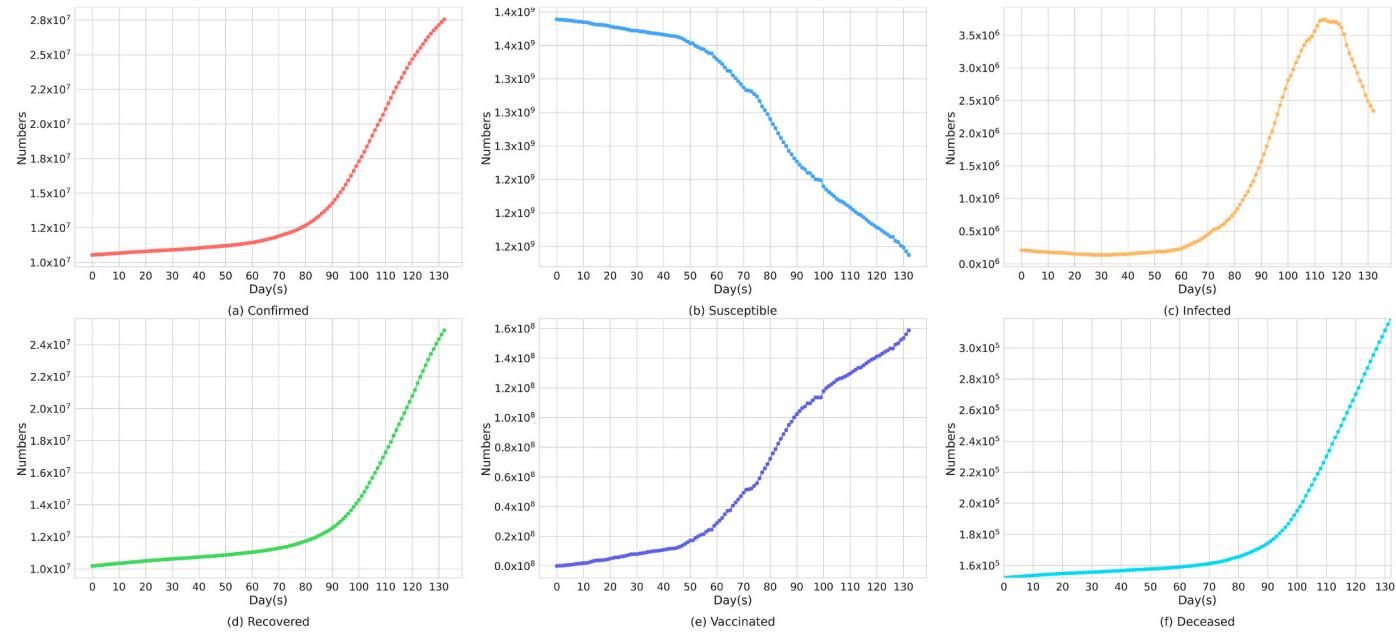
$$R = (x_W - x_W^*, x_W - x_{W+1}^*, \dots, x_{T-2} - x_{T-2}^*) \quad (36)$$

By using the moving average method, most of the noise data are eliminated while the baseline retains its underlying shape. The residual contains random noise and is not considered as model input. Thus, the shape of a single piece of input data can be obtained as  $w \times 1$ . In the time range of length  $T$ , let the output dimension of the model be  $d$ , and the input data latitude that can be constructed is  $(T - w - d) \times w$ . Finally, the optimal hyperparameters are obtained by training and evaluating the model and then predicting  $\hat{\beta}(t)$ ,  $\hat{\gamma}(t)$ , and  $\hat{\delta}(t)$ . The predicted number of infections can be obtained by using Eq. (16). The prediction algorithm is shown in ALGORITHM 1.

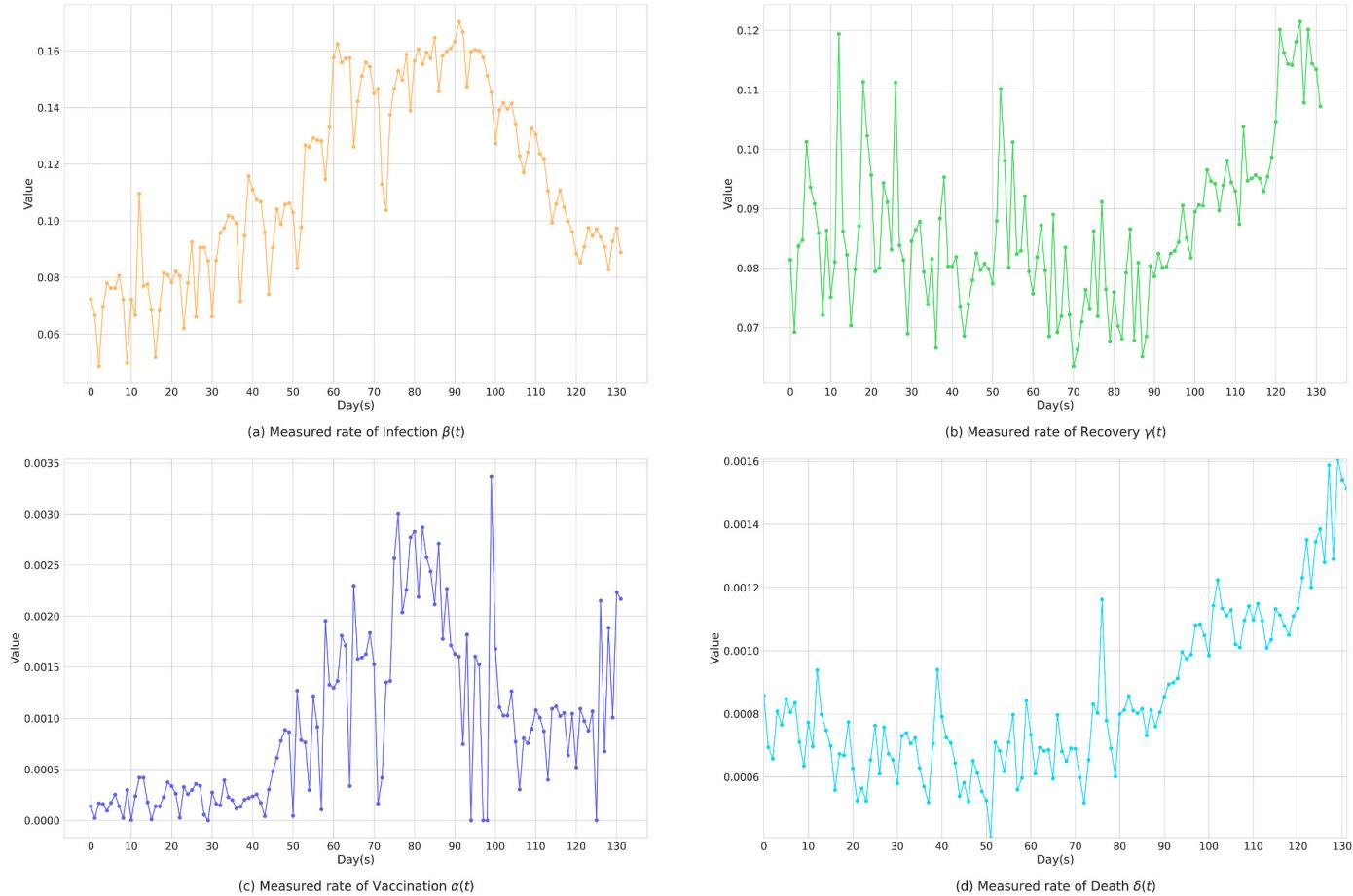
#### ALGORITHM 1. SIRVD-DL prediction model

**Input:**  $\{C(t), R(t), D(t), V(t), 0 \leq t \leq T - 1\}$ ,  $N$ : the total population,  $w$ : the time window size,  $n$ : the feature number,  $d$ : the prediction days.

- 1: Data preprocessing:
- 2: Get the input of SIRVD:  $\{S(t), I(t), R(t), V(t), D(t), 0 \leq t \leq T - 1\}$  using Eq. (1-2)
- 3: While  $0 \leq t \leq T - 1$  do
- 4: calculate  $\{\beta(t), \gamma(t), \delta(t), \alpha(t), \sigma(t), 0 \leq t \leq T - 2\}$  using Eq. (21-24)
- 5: Min-max normalization
- 6: While  $W \leq t \leq T - 2$  do
- 7: Moving average:  $x_t^* = \frac{1}{W} \sum_{i=1}^W x_{t-i+1}$
- 8: Extract the baseline  $\beta^*(t)$ ,  $\gamma^*(t)$  and  $\delta^*(t)$
- 9: end while
- 10: Define  $X$  as the deep learning model input,  $|X| = (T - w - d) \times w$
- 11: While  $0 \leq t \leq T - w - d$  do
- 12: Define  $x$  as the feature vector
- 13:  $x = (x_{t-w+1}^*, x_{t-w+2}^*, \dots, x_t^*)$
- 14: Append  $X$  with  $x$
- 15: end while
- 16: Define  $Y$  as the object value,  $|Y| = (T - w - d) \times d$
- 17: Split the train set and test set by  $X, Y$
- 18: Train and evaluate the model using deep learning methods
- 19: Get the predicted model parameters:  $\hat{\beta}(t), \hat{\gamma}(t), \hat{\delta}(t)$
- 20: Compute the predicted values using Eq. (16)
- 21: **Output:** the predicted values  $\hat{I}(t)$



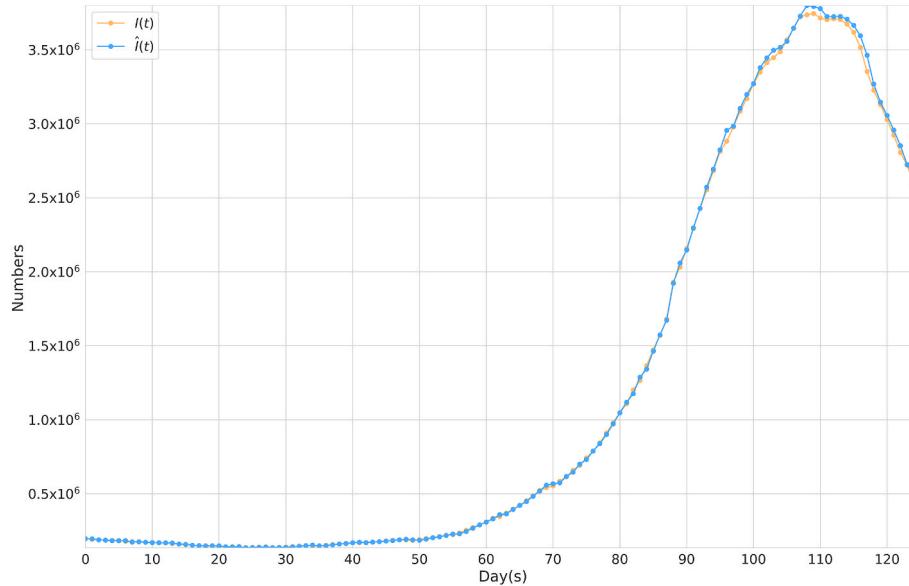
**Fig. 6.** The data distribution of the SIRVD model. (a) Confirmed (b) Susceptible (c) Infected (d) Recovered (e) Vaccinated (d) Deceased.



**Fig. 7.** The measured model parameters using the time-dependent SIRVD model. (a) The measured rate of Infection  $\beta(t)$  (b) Measured rate of Recovery  $\gamma(t)$  (c) Measured rate of Vaccination  $\alpha(t)$  (d) Measured rate of Death  $\sigma(t)$

**Table 5**  
Single-day prediction results of the test set.

Method	Evaluation metric values			
	RMSE	NRMSE	R2	MAPE
Vanilla LSTM	151582.32	0.049222	0.922392	3.292124
Stacked LSTM	92068.48	0.029897	0.958642	2.149824
BiDirectional LSTM	145205.95	0.047152	0.918766	3.659723
GRU	89787.23	0.029156	0.959125	1.893425
SIRVD-DL	<b>38518.55</b>	<b>0.012373</b>	<b>0.995</b>	<b>0.92</b>



**Fig. 8.** Single-day prediction using SIRVD-DL in India.

$$R^2 = \frac{\left( \sum_{i=1}^N ((x_i - \bar{x})(y_i - \bar{y})) \right)^2}{\sum_{i=1}^N (x_i - \bar{x})^2 \times \sum_{i=1}^N (y_i - \bar{y})^2} \quad (39)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left( \frac{|\hat{y}_i - y_i|}{y_i} \right) * 100\% \quad (40)$$

### 2.3.7. Model evaluation metrics

By comparing actual and predicted values, the accuracy of the model can be evaluated. This study uses four evaluation metrics to make fair and effective comparisons. They are: root mean square error (*RMSE*), normalized root mean square error (*NRMSE*), determination coefficient (*R*<sup>2</sup>), and average absolute percentage error (*MAPE*). The calculation method is shown in Eq. (37-40).

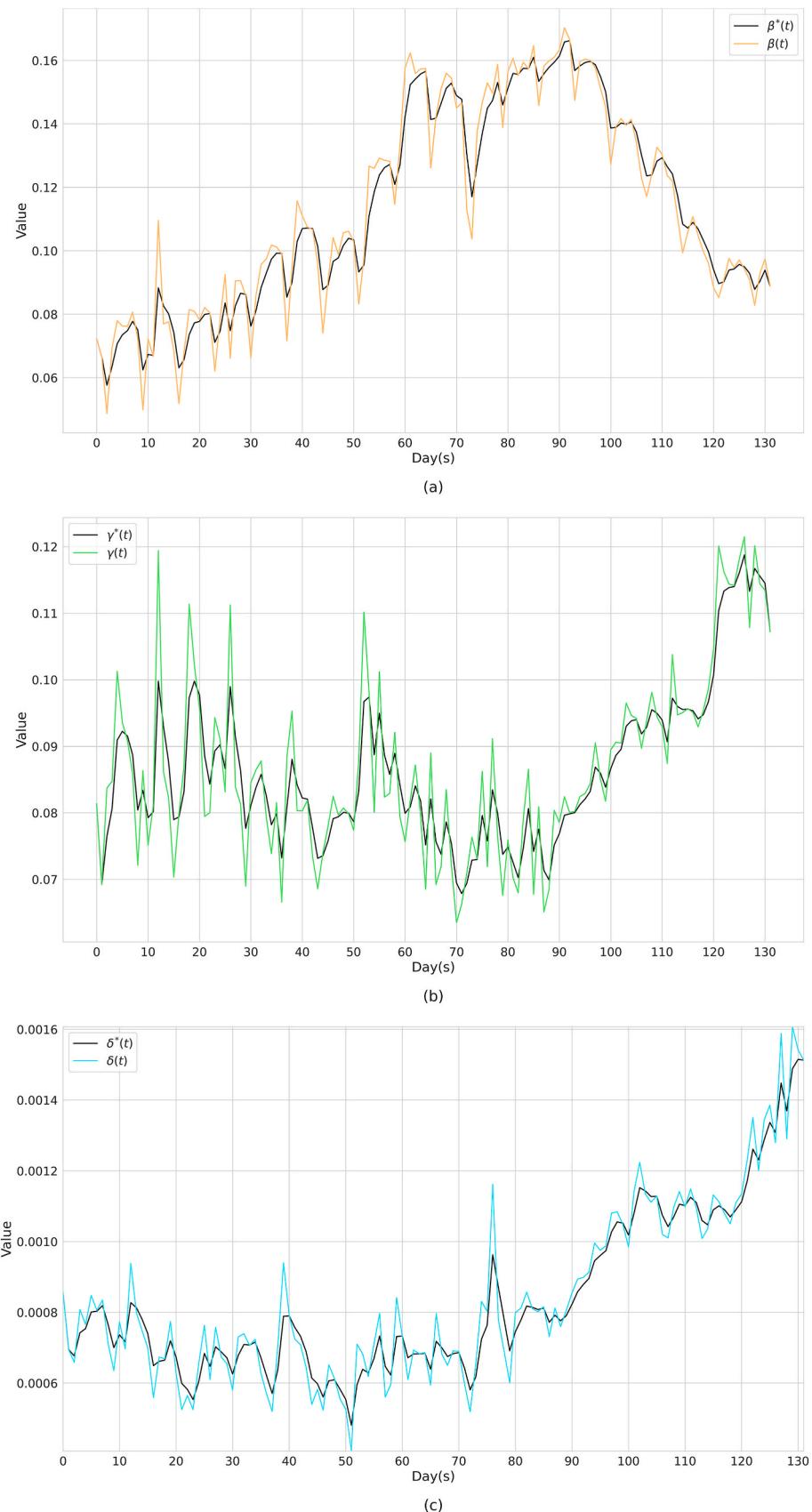
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \hat{y}_i - y_i \right)^2} \quad (37)$$

$$NRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N \left( \hat{y}_i - y_i \right)^2}}{\bar{y}} \quad (38)$$

**Table 6**

The prediction performance in 3-day, 7-day, 14-day, 21-day, and 28-day between SIRVD-DL and other models.

Days	Metrics	BiDirectional LSTM	GRU	Stacked LSTM	Vanilla LSTM	SIRVD-DL
3-day	R2	0.82	0.825	0.82	0.849	0.999
	MAPE%	5.51	5.48	6.23	4.58	2.51
7-day	R2	0.594	0.459	-0.028	0.783	0.993
	MAPE%	8.92	12.4	18.3	6.5	5.07
14-day	R2	-6.57	-12.54	-51.84	-3.25	0.936
	MAPE%	31.71	46.2	98.22	24.09	10.93
21-day	R2	-22.37	-31.64	-68.21	-11.75	0.718
	MAPE%	53.01	62.05	85.82	34.88	17.89
28-day	R2	-225.94	-188.63	-198.27	-229.79	0.29
	MAPE%	89.29	79.21	85.95	86.53	26.57



**Fig. 9.** Extract the baseline of model parameters using the moving average method. The baseline  $x_t^*$  preserves the underlying shape of original data  $x_t$ . (a)  $\beta^*(t)$  and  $\beta(t)$  (b)  $\gamma^*(t)$  and  $\gamma(t)$  (c)  $\alpha^*(t)$  and  $\alpha(t)$  (d)  $\delta^*(t)$  and  $\delta(t)$



**Fig. 10.** The residual of model parameters using the moving average method. (a)  $\beta(t) - \beta^*(t)$  (b)  $\gamma(t) - \gamma^*(t)$  (c)  $\alpha(t) - \alpha^*(t)$  (d). $\delta(t) - \delta^*(t)$

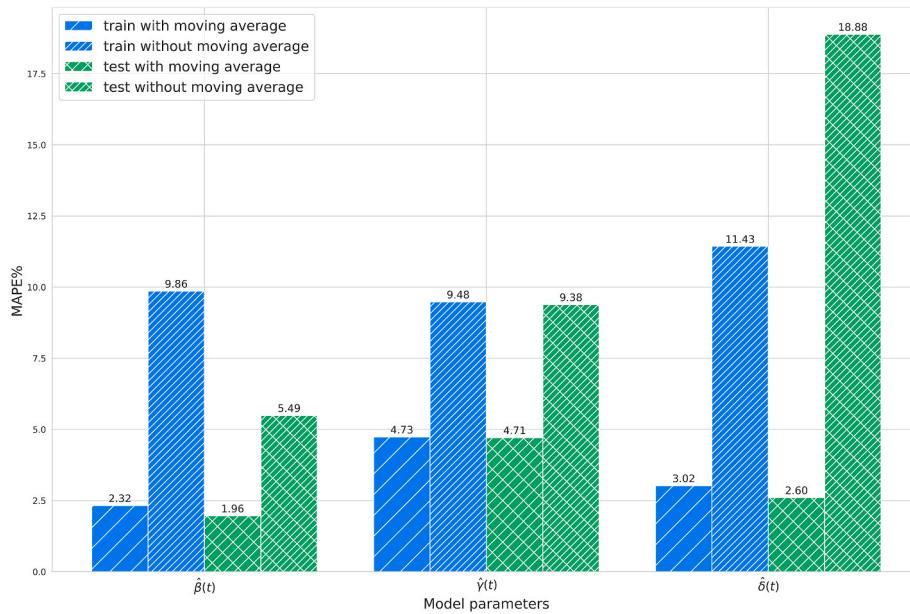


Fig. 11. The performance of the model on the training set and test set with and without moving average for data smoothing.

values that define the architecture of the deep learning model. The correct value of hyperparameters is the key to achieving high-quality models. In order to determine the best combination of hyperparameters, this paper uses a grid search algorithm that takes a set of possible parameters for each adjusted hyperparameter. Then, after determining each possible hyperparameter combination [31], each combination is used to train the deep learning model. In order to avoid the possibility of errors due to the initial random setting of the weights, each set of hyperparameters is used for three pieces of training, and then each implemented model is evaluated. The hyperparameter combinations are shown in Table 4.

The data set used in our experiment is India's historical data from January 15, 2021, to May 27, 2021. The data is divided into a training set and a test set to train and test our prediction model. The evaluation metrics of the model are in Section 2.3.3.

### 3.2. Experiment and result analysis

#### 3.2.1. RQ1: Is the time-dependant SIRVD model reasonable and effective?

As shown in Fig. 6, the relevant data set in India includes the cumulative number of confirmed cases, the number of susceptible people, the number of infected people, the number of recovered people, the number of vaccinated people, and the number of deaths. We applied the time-dependant SIRVD model to this data set and obtained the changes in model parameters from January 15, 2021, to May 26, 2021, through evaluation, as shown in Fig. 7, including infection rate, recovery rate, vaccination rate, and death rate. It should be noted that the susceptibility rate  $\sigma(t)$  is not included here. Since the human body will produce antibodies to prevent future re-infection of the COVID-19 virus [19], the susceptibility rate  $\sigma(t)$  is assumed to be equal to zero. It can be seen from Fig. 7(a) that the infection rate is at its maximum when Day = 90. Compared with the middle position of the rising period in Fig. 6(c), this is the fastest rising time, which shows that the model parameters measurement is valid. In Fig. 7(b), the recovery rate has a trough between Day = 60 and Day = 90. This is because a large number of infections leads to insufficient medical resources, which greatly reduces the success rate of recovery. Similarly, in the same interval, the vaccination rate in Fig. 7(c) reached a peak. The death rate in Fig. 7(d) gradually increased after the peak of infection, indicating that the number of death cases would reach the peak.

In general, the model parameters evaluated based on the time-

dependant SIRVD model are a good measure of real-time changes in the development of the epidemic. And, the most important thing is that it can be found from Fig. 7 that each model parameter obtained by the measurement fluctuates up and down with time series periodicity, which is the basis for the subsequent parameter prediction.

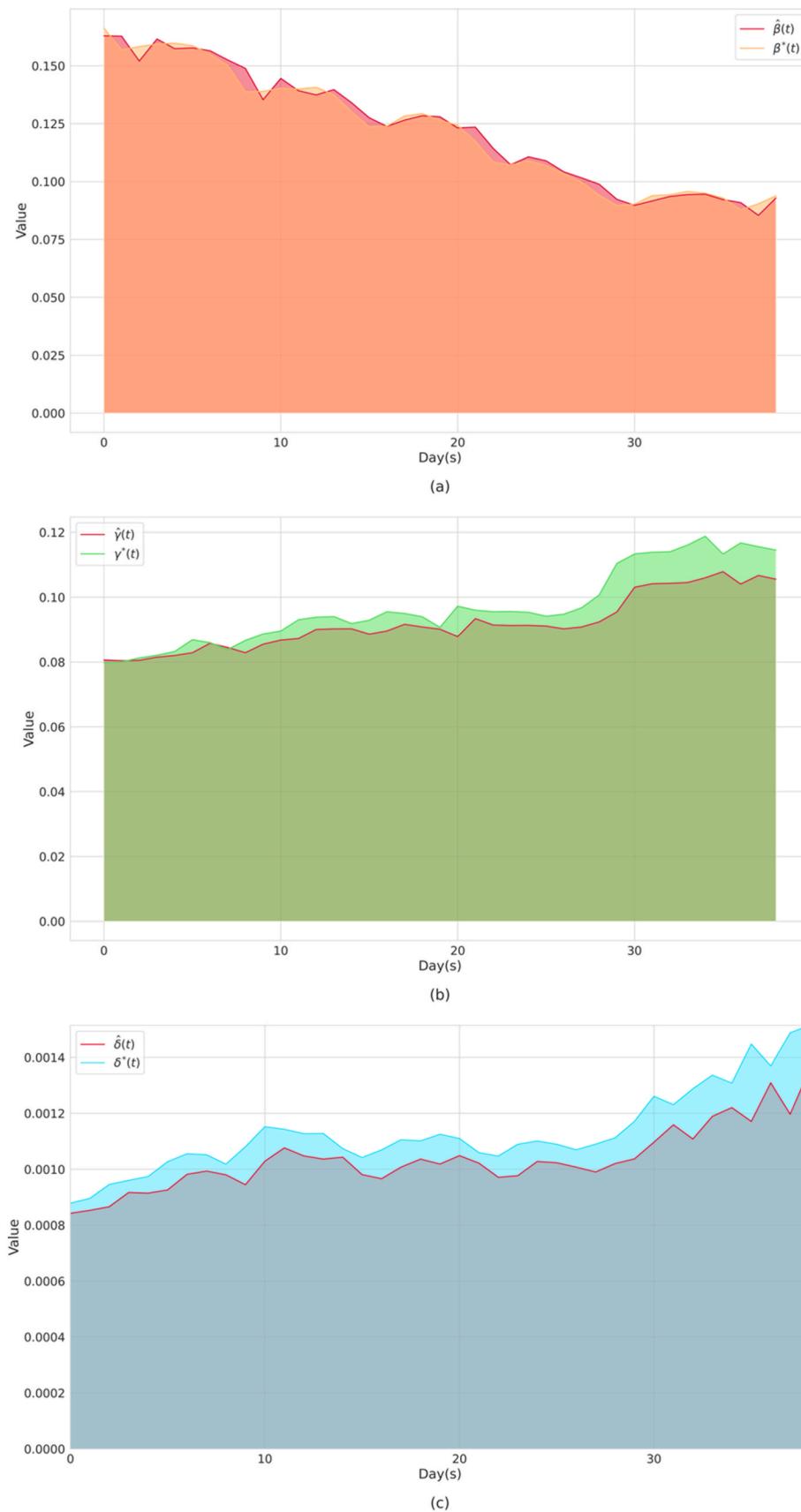
#### 3.2.2. RQ2 how is the prediction performance of the proposed SIRVD-DL model? What is the improvement compared to single deep learning methods? How does the model perform in the short and medium-term forecasts?

In order to verify the effectiveness of the proposed SIRVD-DL model prediction, we compared it with the method of using deep learning to predict in existing studies. Among them, Stacked LSTM and bi-directional LSTM use the methods of Arora et al. [32], Gru and vanilla LSTM use the methods of Nabi et al. [13]. Table 5 shows the comparison of SIRVD-DL and Vanilla LSTM, Stacked LSTM, BiDirectional LSTM, and GRU four prediction models on single-day prediction. The test data set is India's data from April 18, 2021, to May 27, 2021.

It can be seen from Table 5 that SIRVD-DL is the best based on each evaluation metrics in single-day forecasting, with the lowest RMSE of 385128.55, NRMSE of 0.012373, R2 of 0.995, and MAPE of 0.92. MAPE is within one percent, which is an improvement of 51% compared to the best existing single deep learning prediction methods. Fig. 8 shows the difference between the predicted number of infections and the actual number of infections of the SIRVD-DL model.

At the same time, in order to test the effect of the model in the short-term and medium-term prediction, we did 3-day, 7-day, 14-day, 21-day, and 28-day experiments. The experimental results are shown in Table 6. It can be seen from the table that SIRVD-DL predicted a MAPE of 2.51% on the 3rd, 5.07% on the 7th, 10.93% on the 14th day, 17.89 on the 21st day, and 26.57% on the 28th day. Compared with other methods, SIRVD-DL performs better in short-term and medium-term prediction. As the number of forecast days increases, the coefficient of determination R2 is always maintained at a relatively high level, which shows that the prediction of the model is effective. It is worth noting that with the increase in the number of prediction days, R2 appears to be a negative number in other methods. This is because we used the `sklearn.metrics.r2_score` function in Sklearn to calculate. A negative number indicates that the prediction error of the fitting function is greater than the mean function, which indicates that the prediction performance of the model is not good.

In general, the SIRVD-DL model performs well in single-day forecasts



**Fig. 12.** Model parameters ( $\beta^*(t)$ ,  $\gamma^*(t)$  and  $\delta^*(t)$ ) after data smoothing and predicted model parameters ( $\hat{\beta}(t)$ ,  $\hat{\gamma}(t)$  and  $\hat{\delta}(t)$ ).

**Table 7**

The performance of SIRVD-DL on the data sets of Argentina, Brazil, South Korea, Russia, the United Kingdom, France, Germany, and Italy.

	RMSE	NRMSE	R2	MAPE
Argentina	4544.43	0.022	0.994	1.55
Brazil	23888.52	0.0217	0.989	1.7
South Korea	194.99	0.026	0.993	1.85
Russia	2192.7	0.006	0.999	0.414
United Kingdom	2979.57	0.001	0.999	0.04
France	37669.02	0.009	0.999	0.368
Germany	7222.95	0.033	0.99	2.3
Italy	4640.2	0.009	0.997	0.71
Average	10416.05	0.016	0.995	1.12

and short-term and medium-term predictions, which shows that the proposed model is effective. Moreover, the short-term and medium-term predictions have very important guidance and reference significance for helping governments balance the load of medical infrastructure and to implement such control measures as mandatory implementation or lifting of blockades.

### 3.2.3. RQ3 is the moving average method proposed in this paper effective for data smoothing? What is the difference between using this method and not using this method?

In order to predict the number of infections, we first used the moving average method to smooth the parameters evaluated by the SIRVD model, which is shown in Fig. 9. The baseline extracted by our algorithm eliminates most of the anomalies and noises while preserving its underlying shape. This processing method makes the changes of model parameters more periodic and more suitable for time series prediction. As shown in Fig. 10, it is the residual error after model parameter extraction. The residual error contains random noise, so prediction is not considered. After the data is smoothed, we apply standardization again to get a standardized baseline. Such a baseline serves as the input of our deep learning model.

We have carried out comparative experiments using moving average for data smoothing and not using moving average respectively. As shown in Fig. 11, the x-axis shows the three predicted model parameters  $\hat{\beta}(t)$ ,  $\hat{\gamma}(t)$  and  $\hat{\delta}(t)$ . The blue bars are the performance results of the two methods on the training set. The green bars are the performance results on the test set. It can be clearly found that after the moving average is used for processing, the predicted MAPE has a significant drop, which shows that the moving average method has a greater improvement in the prediction of the improved model parameters. Fig. 12 shows the actual parameter prediction effect.

## 4. Discussion

COVID-19 has changed the daily lives of most people around the world. Researchers across the world have done much work on the transmission mode of COVID-19 that has an important guiding role in helping medical workers and government decision-makers. A large number of studies have proposed different COVID-19 prediction models, among which artificial intelligence-based methods are currently the most interesting to the global scientific community [20]. Although the use of artificial intelligence methods for prediction has high accuracy, it is difficult to use deep learning technology to discover the time change pattern of the number of infected people [13]. In order to solve this problem, We proposed the SIRVD-DL model to combine deep learning methods with mathematical models of infectious diseases to predict the parameters of the infectious disease model. The advantage of the SIRVD-DL prediction model is that the introduction of an infectious disease mathematical model makes the overall prediction more interpretable and robust. From our experimental results, it can be found that compared with other methods [32], SIRVD-DL not only improves the single-day forecast by 51% but is also suitable for short-term and

medium-term forecasting. This is more practical for policy formulation and strategic planning.

In order to demonstrate the universality of our method, we also used data from eight countries including Argentina, Brazil, South Korea, Russia, the United Kingdom, France, Germany, and Italy as verification. The end data for COVID-19 data in 8 countries is the same, which is May 27, 2021. The start date of COVID-19 data for Argentina is December 29, 2020, Brazil is January 16, 2021, South Korea is February 25, 2021, Russia is December 15, 2020, and the United Kingdom is May 2021. On the 26th, France, Germany, and Italy are all on December 27, 2020. The prediction results of these eight countries using SIRVD-DL are shown in Table 7. From the table, we can see that SIRVD-DL performs well on the data sets of other countries. The average MAPE is within 1.5% and the coefficient of determination is 0.995. This shows that our model is extensive and can be used in other countries and regional COVID-19 predictions.

It should be noted that our model still has some limitations. One limitation is the accuracy of the data. The data we use is official statistics, and there will be some differences from the actual data in the real world. At the same time, the model does not consider the situation of asymptomatic infections and survivors being infected again. In addition, the reinfection rate of vaccination and the rate of vaccine efficacy are not considered in this paper because they are difficult to obtain and may be inaccurate, but they will impact on the model. Another limitation of our research is that the deep learning method we use for parameter prediction may not be optimal, and there may be better ways to solve the model and predict the parameters. We will improve our model in future work.

## 5. Conclusion

We proposed the SIRVD-DL model to establish a virus transmission model and predict the trend of COVID-19. The model combines the mathematical model of infectious diseases and deep learning technology to overcome the problem that single deep learning methods cannot capture development trends. It can reflect the epidemic model parameters in real-time under the current situation of large-scale vaccination and predict the epidemic trend of infectious diseases. Experimental numerical results show that the SIRVD-DL model can evaluate model parameters such as infection rate, recovery rate, and death rate. The single-day prediction accurate rate improves 51% compared to the best existing single deep learning prediction methods. At the same time, the model performs very well in short-term and medium-term predictions. The single-day prediction error is 2.51%, the 7-day prediction error is 5.07%, the 14-day prediction error is 10.93, and the 28-day prediction error is within 30%.

## Data availability

Publicly available datasets can be found here: <https://github.com/CSSEGISandData/COVID-19>. The vaccination dataset is available at: <https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations>. And our code and experimental data are publicly available at: <https://github.com/Rambo5555/SIRVD-DL>.

## Author contributions

Zhifang Liao conceived the experiments, Xiaoping Fan collected the data, Peng Lan and Zhining Liao conducted the experiments, Benjamin Kelly and Aidan Innes analyzed the results. All authors reviewed the manuscript.

## Funding

This work was supported by Special Funds for the Construction of an Innovative Province of Hunan, No:2020GK2028.

## Declaration of competing interest

The authors declare no competing interests.

## References

- [1] D. Rafiq, A. Batool, M. Bazaz, Three months of COVID-19: a systematic review and meta-analysis, *Rev. Med. Virol.* 30 (2020) e2113.
- [2] C. Giesen, L. Diez-Izquierdo, C.M. Saa-Requejo, I. Lopez-Carrillo, C.A. Lopez-Vilela, A. Seco-Martinez, M.T.R. Prieto, E. Malmierca, C. Garcia-Fernandez, C. E. Surveillance, Epidemiological characteristics of the COVID-19 outbreak in a secondary hospital in Spain, *Am. J. Infect. Contr.* 49 (2021) 143–150.
- [3] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. Lond. - Ser. A Contain. Pap. a Math. Phys. Character* 115 (1927) 700–721.
- [4] Z. Liao, P. Lan, Z. Liao, Y. Zhang, S.J.S.R. Liu, TW-SIR: Time-Window Based SIR for COVID-19 Forecasts, vol. 10, 2020.
- [5] P. Singh, A. Gupta, Generalized SIR (GSIR) Epidemic Model: an Improved Framework for the Predictive Monitoring of COVID-19 Pandemic, *ISA transactions*, 2021.
- [6] I. Cooper, A. Mondal, C.G. Antonopoulos, A SIR model assumption for the spread of COVID-19 in different communities, *Chaos, Solitons & Fractals* 139 (2020) 110057.
- [7] H. Mohammadi, S. Rezapour, A. Jajarmi, On the Fractional SIRD Mathematical Model and Control for the Transmission of COVID-19: the First and the Second Waves of the Disease in Iran and Japan, *ISA transactions*, 2021.
- [8] S.B. Ramezani, A. Amirlatifi, S. Rahimi, A novel compartmental model to capture the nonlinear trend of COVID-19, *Comput. Biol. Med.* 134 (2021), 104421 - 104421.
- [9] I. Holmdahl, C. Buckee, Wrong but useful—what covid-19 epidemiologic models can and cannot tell us, *N. Engl. J. Med.* 383 (2020) 303–305.
- [10] J. Panovska-Griffiths, Can Mathematical Modelling Solve the Current Covid-19 Crisis? Springer, 2020.
- [11] V.K.R. Chimmula, L. Zhang, Time series forecasting of COVID-19 transmission in Canada using LSTM networks, *Chaos, Solitons & Fractals* 135 (2020) 109864.
- [12] İ. Kirbaş, A. Sözen, A.D. Tuncer, F.Ş. Kazancıoğlu, Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches, *Chaos, Solitons & Fractals* 138 (2020) 110015.
- [13] K.N. Nabi, M.T. Tahmid, A. Rafi, M.E. Kader, Forecasting COVID-19 cases: a comparative analysis between recurrent and convolutional neural networks, *Res. Phys.* 24 (2021) 104137.
- [14] J. Devaraj, R.M. Elavarasan, R. Pugazhendhi, G. Shafiuallah, S. Ganeshan, A. K. Jeysree, I.A. Khan, E. Hossain, Forecasting of COVID-19 cases using deep learning models: is it reliable and practically significant? *Res. Phys.* 21 (2021) 103817.
- [15] A. Fokas, N. Dikaios, G. Kastis, Mathematical models and deep learning for predicting the number of individuals reported to be infected with SARS-CoV-2, *J. R. Soc. Interface* 17 (2020) 20200494.
- [16] E. Mathieu, H. Ritchie, E. Ortíz-Ospina, M. Roser, J. Hasell, C. Appel, C. Giattino, L. Rodés-Guirao, A global database of COVID-19 vaccinations, *Nat. Hum. Behav.* (2021) 1–7.
- [17] J. Farooq, M.A. Bazaz, A novel adaptive deep learning model of Covid-19 with focus on mortality reduction strategies, *Chaos, Solitons & Fractals* 138 (2020) 110148.
- [18] T. Usherwood, Z. LaJoie, V. Srivast, A Model and Predictions for COVID-19 Considering Population Behavior and Vaccination, 2021.
- [19] A.L. Phelan, COVID-19 immunity passports and vaccination certificates: scientific, equitable, and legal challenges, *Lancet* 395 (2020) 1595–1598.
- [20] J. Musulin, S. Baressi Šegota, D. Štifanić, I. Lorencin, N. Andelić, T. Šušteršić, A. Blagojević, N. Filipović, T. Čabov, E. Markova-Car, Application of artificial intelligence-based regression methods in the problem of COVID-19 spread prediction: a systematic review, *Int. J. Environ. Res. Publ. Health* 18 (2021) 4287.
- [21] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, 2014 arXiv preprint arXiv:1412.3555.
- [22] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (1997) 1735–1780.
- [23] A. Graves, A.-r. Mohamed, G. Hinton, Speech Recognition with Deep Recurrent Neural Networks, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Ieee, 2013, pp. 6645–6649.
- [24] M. Schuster, K.K. Paliwal, Bidirectional recurrent neural networks, *IEEE Trans. Signal Process.* 45 (1997) 2673–2681.
- [25] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation, 2014 arXiv preprint arXiv:1406.1078.
- [26] T.E. Oliphant, A Guide to NumPy, Trelgol Publishing USA2006.
- [27] W. McKinney, pandas: a foundational Python library for data analysis and statistics, *Python High Performance Sci. Comput.* 14 (2011) 1–9.
- [28] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, Tensorflow: A System for Large-Scale Machine Learning, 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), 2016, pp. 265–283.
- [29] A. Gulli, S. Pal, Deep Learning with Keras, Packt Publishing Ltd, 2017.
- [30] G. Van Rossum, F.L. Drake, The python Language Reference Manual, Network Theory Ltd, 2011.
- [31] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, *J. Mach. Learn. Res.* 13 (2012).
- [32] P. Arora, H. Kumar, B.K. Panigrahi, Prediction and analysis of COVID-19 positive cases using deep learning models: a descriptive case study of India, *Chaos, Solitons & Fractals* 139 (2020) 110017.