

An intelligent forecast for COVID-19 based on single and multiple features

Yilei Wang¹ | Yiting Zhang¹  | Xiujuan Zhang¹ |
Hai Liang² | Guangshun Li¹ | Xiaoying Wang³

¹School of Computer Science, Qufu Normal University, Rizhao, China

²School of Computer Science, Guilin University of Electronic Technology, Guilin, China

³The Smart Hospital R & D Center, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

Correspondence

Xiaoying Wang, The Smart Hospital R & D Center, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510000, China.

Email: wxying@mail.sysu.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 62072273

Abstract

It is urgent to identify the development of the Corona Virus Disease 2019 (COVID-19) in countries around the world. Therefore, visualization is particularly important for monitoring the COVID-19. In this paper, we visually analyze the real-time data of COVID-19, to monitor the trend of COVID-19 in the form of charts. At present, the COVID-19 is still spreading. However, in the existing works, the visualization of COVID-19 data has not established a certain connection between the forecast of the epidemic data and the forecast of the epidemic. To better predict the development trend of the COVID-19, we establish a logistic growth model to predict the development of the epidemic by using the same data source in the visualization. However, the logistic growth model only has a single feature. To predict the epidemic situation in an all-round way, we also predict the development trend of the COVID-19 based on the Susceptible Exposed Infected Removed epidemic model with multiple features. We fit the data predicted by the model to the real COVID-19 epidemic data. The simulation results show that the predicted epidemic development trend is consistent with the actual epidemic development trend, and our model performs well in predicting the trend of COVID-19.

KEYWORDS

COVID-19, crawler data, logistic growth model, SEIR model, visual analysis

1 | INTRODUCTION

In December 2019, patients with unexplained pneumonia appeared successively in China and other countries around the world. On January 7 of the following year, the expert team initially named the pathogen of this unexplained viral pneumonia as the Corona Virus Disease 2019 (COVID-19).¹ The transmission route and infectivity of the virus are stronger than SARS. Further it is equally infectious even during the asymptomatic incubation period.² As of April 17, 2021, the cumulative number of confirmed cases of new coronary pneumonia in China has exceeded 110,000, and the cumulative number of confirmed cases overseas has reached 170 million. The number of deaths due to COVID-19 in the world has exceeded 3 million. The raging epidemic not only affects the global economic system, but also threatens people's lives. At this time, real-time monitoring of epidemic data is crucial. However, real-time broadcast of epidemic data in text form will inevitably provoke a certain amount of reading pressure on readers, and visualization of data meets the public's demand for intuitive information. For example, China's Baidu, Sina, and other websites displayed a large amount of real-time epidemic data on the website in the form of a table at the first time. This development has changed the situation of large amounts of data and a strong sense of "stacking" in the past.^{3,4}

The epidemic continues to be spreading. Compared with the explosive growth of confirmed cases at the beginning of the epidemic, the current situation of the epidemic has gradually eased. But in China and the world, the epidemic is still a long-term focus. In terms of epidemic prevention and control, visual analysis is not sufficient. Mathematical models are utilized to predict the trend of the epidemic and establish a global epidemic warning mechanism.

2 | RELATED WORK

2.1 | Feature selection

Although a large amount of epidemic data related to the COVID-19 epidemic can be crawled on the Internet, there are still a large number of features that are irrelevant or redundant to the establishment of the model, which leads to excessive training time or overfitting. To avoid similar situations and further improve the accuracy of the model, we use feature selection to reduce the number of features and select the truly relevant features to simplify the model. According to the form of feature selection, feature selection methods can be divided into three categories: filter, wrapper, and embedded.^{5,6}

In recent years, many researchers in academia have also proposed improved feature selection methods, such as Warda M. Shabana, Asmaa H. Rabieb, Ahmed I. Salehb, et al. proposed in July 2020. A hybrid method that combines the feature selection of wrappers and filters: hybrid feature selection methodology. It extracts the most informative features from chest CT images of COVID-19 patients and nonpatients, facilitating disease detection at an early stage and immediate isolation.⁷ In 2016, Sannasi Ganapathy, Pandi Vijayakumar, Palanichamy Yogesh, et al. developed an intelligent new feature selection algorithm based on Conditional Random Field to optimize the number of features.⁸ In 2019, U. Kanimozhi, S. Ganapathy, D. Manjula, et al. proposed a new model based on fuzzy time rules. The model takes into account the opinions of patients, relatives, and experts through questionnaire surveys and interactions for feature selection and classification, to identify the most important

features.⁹ In August 2021, A. Narin applied particle swarm optimization algorithm and ant colony algorithm to the feature selection method, which is beneficial to radiologists as a decision support system.¹⁰

2.2 | Prediction model

Wang and Zhang proposed a new PatchShuffle stochastic pooling neural network related to COVID-19 virus to help doctors diagnose COVID-19 cases more accurately, which is very important for us to understand the characteristics of the COVID-19 virus.¹¹ Habibzadeh and Stoneman drew a bird's-eye view of the global epidemic at the beginning of the epidemic to gain a deeper understanding of the COVID-19.¹² However, as the epidemic continues to spread, the latest data is needed to map the global epidemic. Leung et al. conducted a visual analysis of the spread of the epidemic in China,¹³ but did not analyze the situation of the epidemic abroad. In May 2021, Rokaya Rehouma, Michael Buchert, et al. used machine learning for image segmentation and classification to identify patients with COVID-19 and many ML modules have achieved remarkable predictive results using data sets with limited sample sizes.^{14–16}

Scholars have also tried to use various methods to explore and analyze the development of the COVID-19. In January 2020, Almeshal et al.¹⁷ used zoning and logistic models to predict the spread of COVID-19 in Kuwait. Chen et al.¹⁸ also established a logistic growth model to predict the epidemic trend in the United States. Read et al.¹⁹ used the Susceptible Exposed Infected Removed (SEIR) model to predict the epidemic trend. They anticipated that as of February 4, the number of infected people in Wuhan will reach 190,000. Due to the lack of data at the beginning of the epidemic, the forecast results clearly overestimate the progress of the epidemic. Pandey et al.²⁰ analyzed the outbreak situation in India as of March 30, 2020, and established an SEIR model to predict the number of cases in India in the next 2 weeks. In April 2020, Muhammad Dur-e-Ahmad, Mudassar Imran, and others used the SEIR model to fit the epidemic data of multiple countries to estimate the basic reproduction number R_0 , and conducted a sensitivity analysis on all parameters that affect the R_0 value.²¹ At the same time, López and Rodo²² established an improved SEIR model to simulate the epidemic situation in Spain and Italy. Annas et al.²³ also analyzed the stability of the SEIR model in Indonesia. In November 2020, Radulescu et al.²⁴ studied the spread of the COVID-19 in the community and established an SEIR infectious disease model to predict the development of the epidemic. At the same time, in other fields, there are also cases of using models to predict. Gourav Kumar and Uday Pratap Singh established a hybrid time series econometric model to predict stock prices.²⁵ As the epidemic continues, the COVID-19 has mutated, and the mutant strain is more contagious. However, many of the epidemic data used for visual analysis and modeling prediction are available at the beginning of the epidemic, and the results obtained are no longer applicable to the current epidemic situation. The research area is too small and the prediction results will be biased. So, the latest data are needed to visually analyze and model the epidemic data again.

Many of the current works have separated the visual analysis of the modeling and forecasting, and they have all been implemented very well. However, few scholars combine visualization with modeling prediction. Chen et al.²⁶ combined visual analysis with modeling and prediction. They studied how the virus spreads around Wuhan, and concentrated on the analysis of the spread of the epidemic. The goal is not the right real-time development of the epidemic situation. Therefore, the model prediction should be consistent with the data used in the graphical analysis, so as to better reflect the development trend of the epidemic.

2.3 | Motivations and contributions

In summary, only visual analysis can play a real-time monitoring role in the development of the epidemic, but it cannot predict the development of the next stage of the epidemic. On the other hand, although only modeling and forecasting can make a rough simulation of the future trend of the epidemic, it ignores the readers' need to grasp the current situation of the epidemic. Therefore, as the epidemic continues to spread today, the combination of visual analysis and modeling prediction is essential for real-time control of the epidemic situation and forecast of the epidemic tendency. To solve this problem, this article combines visual analysis and modeling prediction to forecast the development trend of the Novel Coronavirus Pneumonia Epidemic. The contributions of this article are as follows.

1. We use Python crawler technology to crawl the real-time epidemic data of Tencent News. In view of the diversified characteristics of epidemic data, distinct functions in the Python library are used to visualize the data from multiple angles, realizing real-time monitoring and analysis of the epidemic situation in China and the world.
2. We use the Plotly function in Python to build a virus spread model and draw a simulated virus spread map to help readers understand the process of COVID-19 virus spread in the population. Then, define the curve function in the logistic growth model according to the single characteristics of the patient (e.g., single patient type: diagnosed), and import the crawled epidemic data into the function. The nonlinear least-squares method was used to fit and predict the development trend of China's epidemic in the early stage and the intermediate stage of the epidemic. The fitting results show that the current epidemic situation in China has come to an end, and there will be no major fluctuations and turning points, which is consistent with the actual situation of the current epidemic situation in China.
3. To predict the development trend of the epidemic more comprehensively, the characteristics of patient types are refined and further divided into four types: susceptible, latent, confirmed, and recovered. Use MATLAB to establish an SEIR epidemic model and modify the parameters of the differential equation to simulate the trend of various population changes in this epidemic. In addition, by establishing a model of the relationship between infection rate and contact distance, we draw an image of the COVID-19 infection rate that changes with distance, and use the infection rate obtained by modeling as a parameter of the SEIR model to enhance the prediction model Accuracy. The purpose is to enhance people's awareness of the prevention of further coronary pneumonia, so as to better prevent and control the spread of the epidemic. The results of the modeling show that with the improvement of medical standards, the development trend of the epidemic will have a tendency to ease.

3 | SYSTEM STRUCTURE

3.1 | Data set and experimental setup

Data set: This article uses Python's crawler technology to crawl the epidemic module of Tencent News, obtain real-time epidemic data from countries around the world and provinces in China, and store the data locally. In total, we collected epidemic information from 185 countries around the world, and we also obtained epidemic information from 34 provinces in China. The deployment timestamps of the collected epidemic information range from January 28, 2020 to May 1, 2021. Then

call the Seaborn function and Plotly function in the Python library to visually analyze the epidemic data, presenting the trend of the epidemic in two ways: dynamic and static.

Experimental setup: All experiments were conducted using a machine⁵ with 64 GB of memory and an Intel(R) Core(TM) i5-7200U CPU with 12 cores clocked at 2.7 GHz.

3.2 | System structure model

We first obtained real-time epidemic information from Tencent News, and then realized the visual analysis of the epidemic data from both static and dynamic aspects. Then, a single-feature model (logistic growth model) and a multifeature model (SEIR epidemic model) are established to fit and predict the future development trend of the epidemic.^{27,28} The system architecture of the main functions of this article is shown in Figure 1.

4 | VISUAL ANALYSIS

The epidemic data comes from the website of Tencent News. The specific process is as follows. (1) Use a browser to open Tencent News Network, and use the browser's "censorship element" to view the source code and "network" feedback messages.²⁹ (2) Use Python to compile the code, send a request to the website and obtain the real-time JSON data of Tencent's epidemic situation,^{30,31} and the output results are sorted by the names of provinces and countries in

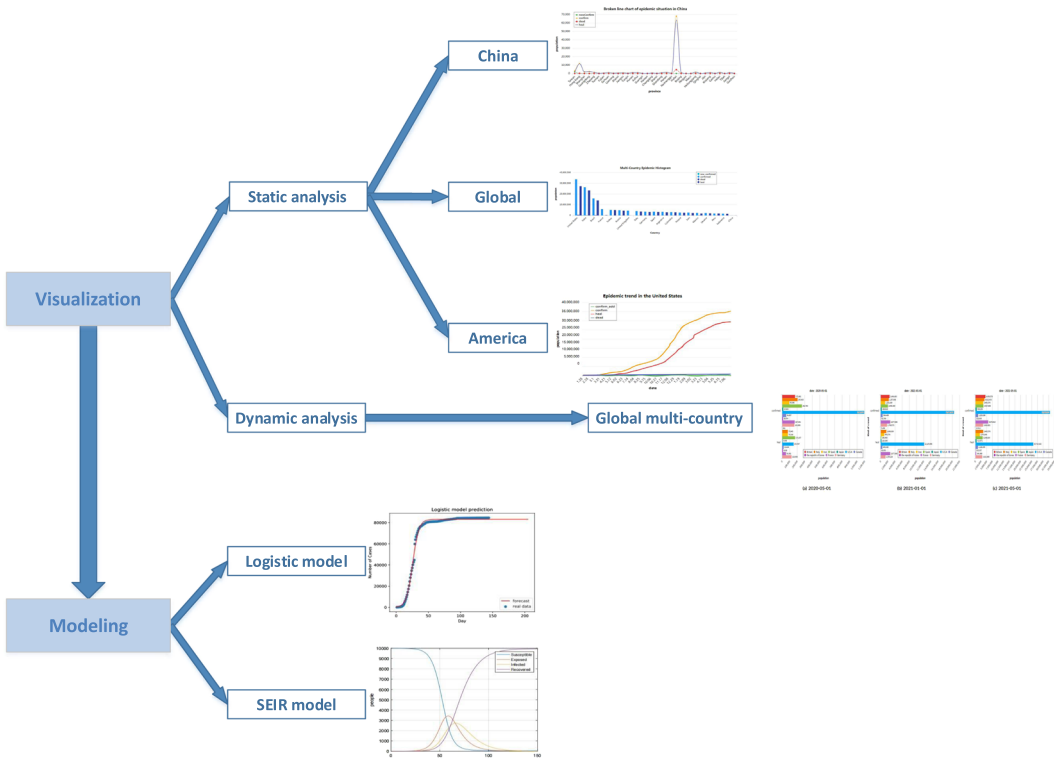


FIGURE 1 System architecture diagram. SEIR, Susceptible Exposed Infected Removed. [Color figure can be viewed at wileyonlinelibrary.com]

China according to the number of confirmed cases.³² (3) Analyze and clean the captured epidemic data and store it in a CSV file, named after the current date.³³

We use the Seaborn function in the Python library to visually analyze the crawled data.^{34,35} Figure 2 shows the distribution of the suspected number, confirmed number, death number, and recovery number in various provinces in China. From Figure 2, we can clearly and intuitively observe the latest situation of the Novel Coronavirus Pneumonia Epidemic in various provinces, and it is found that the number of confirmed cases and deaths in Hubei Province is much higher than those of other provinces in China. It is the province that has suffered the most damage from the Novel Coronavirus Pneumonia Epidemic in China. In addition, the epidemic situation in Hong Kong, Taiwan, and other places is not optimistic.

To understand the development of the foreign epidemic situation, Figure 3 shows a visualization legend of the number of confirmed cases, the number of suspected infections, the number of deaths, and the number of cured people in each country. Because there are too many

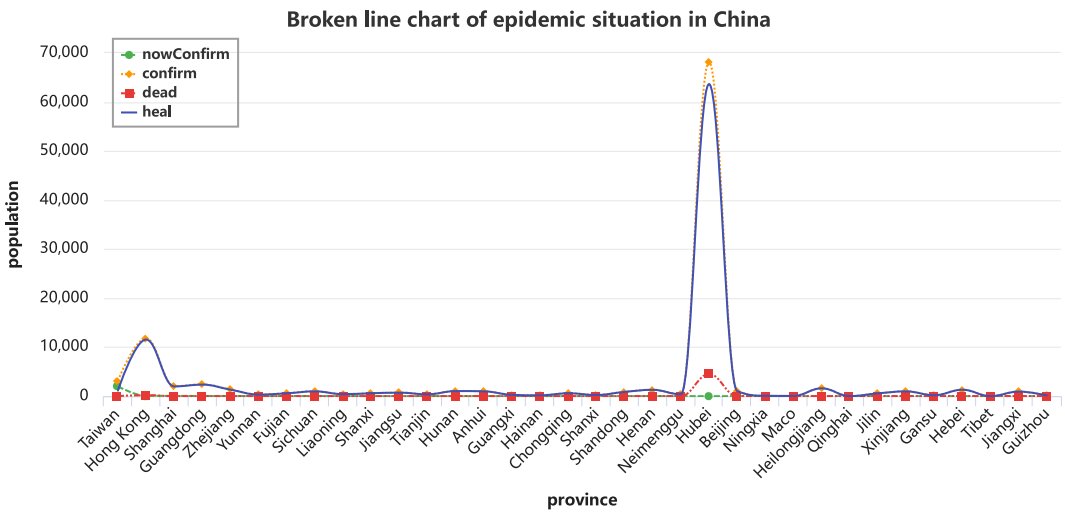


FIGURE 2 Visual analysis of China's epidemic situation [Color figure can be viewed at wileyonlinelibrary.com]

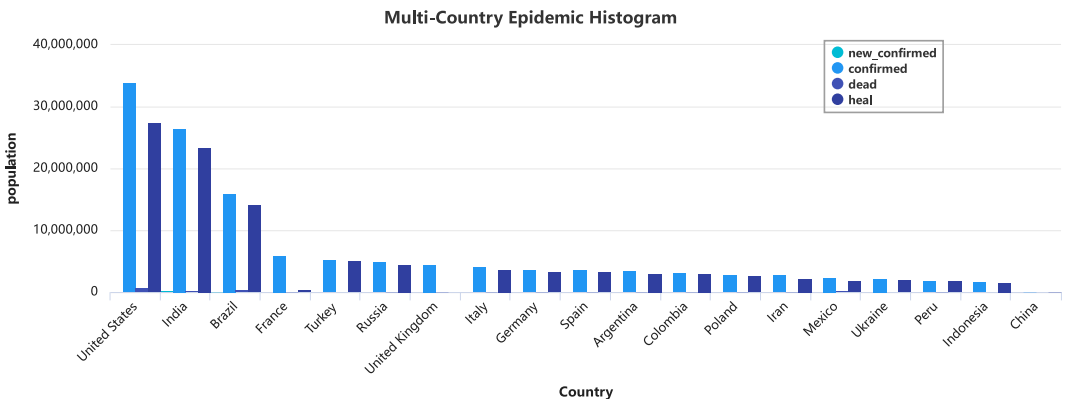


FIGURE 3 Multicountry epidemic histogram [Color figure can be viewed at wileyonlinelibrary.com]

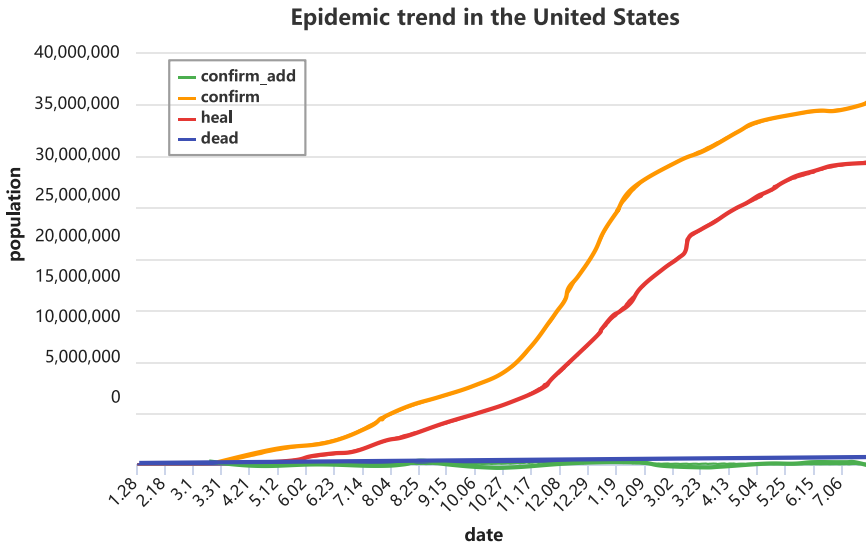


FIGURE 4 US epidemic trend chart [Color figure can be viewed at wileyonlinelibrary.com]

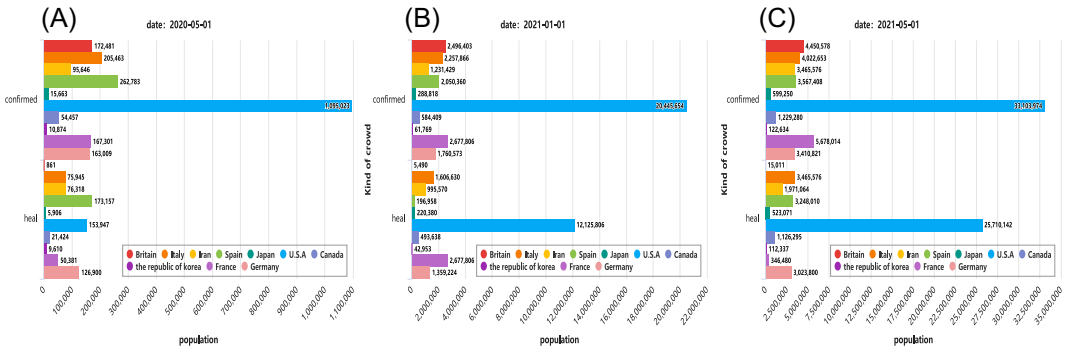


FIGURE 5 Dynamic change of epidemic situation in many countries. (A) 2020-05-01, (B) 2021-01-01, and (C) 2021-05-01. [Color figure can be viewed at wileyonlinelibrary.com]

countries in the world, only the Seaborn function is called to draw the 20 countries with the largest number of confirmed cases in the world.^{36,37}

It can be seen from Figure 3 that many overseas countries have been affected by the Novel Coronavirus Pneumonia Epidemic to varying degrees. Among them, the number of people diagnosed and killed by the new crown pneumonia in the United States is much higher than those in other countries. Countries such as Brazil and India have also been hit by the epidemic. To understand the data of the Novel Coronavirus Pneumonia Epidemic in the United States, Figure 4 shows the trend of the epidemic in the United States. It plots the growth curve of the latest confirmed number, the total number of confirmed cases, the number of cured, and the number of deaths over time.

It is easy to see that the United States has become the most severely hit by the COVID-19 virus in the world, and the number of confirmed cases in the United States is still rising. The global epidemic is still spreading, and there is no sign of wishing to stop.

Nowadays, most visualizations of epidemics are displayed in static form. To facilitate readers to observe the dynamic trends of epidemics gradually changing over time, Figure 5A–C shows the

dynamic changes of epidemics in multiple countries. Three dates were intercepted to show the changes in the epidemic situation in multiple countries.

By observing the dynamic trends of the number of confirmed and cured people in many countries, it is not difficult to find that although the number of cured people is also increasing, it is far less rapid than the increase in the number of confirmed people. The COVID-19 virus is still spreading at an uncontrollable rate, endangering the lives of people in all countries around the world.

5 | PROPOSED MODEL

The spread of the epidemic is wide, it lasts for a long time, and has a great impact on countries around the world. The development trend of the epidemic is still an issue that requires long-term focus. Although visual analysis can show the real-time situation of the development of the epidemic, it cannot predict the development trend of the Novel Coronavirus Pneumonia Epidemic. Therefore, we use the plot function in Python to draw a simulated virus diffusion map to help readers understand the process of COVID-19 virus spreading among people. And then we use the crawled real-time epidemic data to establish the logistic growth model and the SEIR infectious disease model. On the basis of the single- and multifeatures of the epidemic, a reasonable fitting and prediction of the future progress of the epidemic can be made to provide a reference for formulating epidemic prevention measures. The logistic growth model can fit and predict the confirmed data, while the SEIR model can roughly simulate the changes in various groups of people in the model by modifying the parameters in the model.^{38,39} By observing the two modeling results, the changes in the data can be well reflected, and readers can further grasp the future trend of the progress of the epidemic, so as to better prevent and control the epidemic.

5.1 | Diffusion of simulated COVID-19 virus in population

First, we use different colors to mark distinct people. Green represents healthy people and red represents infected people. A movement function is put in place to simulate the free movement of people. Choosing an object-oriented approach and treating each person as an independent object.^{40,41} Traverse all objects and calculate the relative distance of the current object to everyone else. Factors such as distance, infection rate, and so forth change the current status of the object's neighbors.^{42,43} When the distance between the target and the patient is less than the

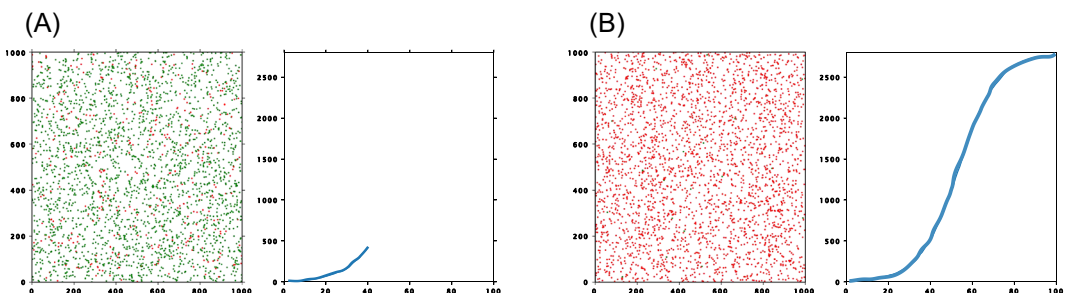


FIGURE 6 Simulate the spread of the virus: (A) prestage and (B) late stage [Color figure can be viewed at wileyonlinelibrary.com]

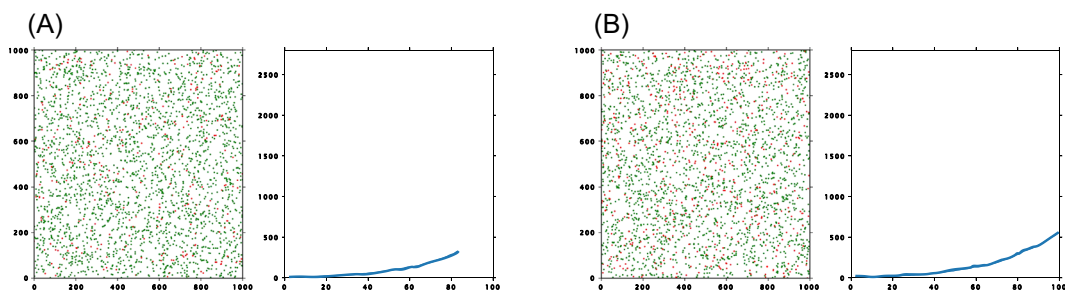


FIGURE 7 Virus spread after simulated isolation: (A) prestage and (B) late stage [Color figure can be viewed at wileyonlinelibrary.com]

transmission radius, there is a definite chance of infection. Figure 6A,B shows the simulated virus spread. The left part of the pictures shows the spread of the virus in the plane, and the x - and y -axes represent the size of the area. The right part represents the growth process of the number of infected people, the x -axis is the number of iterations, and the y -axis is the total number of infected people.

It can be seen from the figure that in the natural state, COVID-19 virus spreads rapidly among the population through people's free movement, and the number of infections increased exponentially in the middle of the epidemic.

After adding isolation measures, the infection radius of the virus is reduced, the infection probability is decreased, and the transmission speed is greatly slowed down. The simulation results show that isolation is an effective way to hamper the further spread of the virus. The simulated virus spread after isolation is shown in Figure 7A,B.

5.2 | Forecast of logistic growth model based on single feature

The logistic growth model is also known as the retarded growth model.⁴⁴ The model is typically used in fields such as describing populations, the growth of infectious diseases, and the forecast of commodity sales. The specific algorithmic idea of the model is as follows: the first enter the crawled epidemic data, automatically calculate the number of days corresponding to the input data, use the abscissa to indicate the number of days, the ordinate to indicate the number of cases, and use the input data as scattered points. The diagram prints out. Then define the S-curve function formula and use the nonlinear least-squares method to fit. Finally, enter the total amount of days that need to be predicted, get the fitted data, and print out the chart.

On the basis of the data of confirmed cases in the real epidemic, we established a logistic growth model to predict the development trend in the next few decades. The specific process is as follows: crawl Tencent News, get the daily number of confirmed diagnoses of the epidemic in China, then define the curve function according to the formula of the logistic growth model, and use the nonlinear least-squares method to fit and predict the number of confirmed diagnoses after 61 days.⁴⁵ Figure 8A,B, respectively, shows the development trend forecast of the early and midterm epidemics.

From Figure 8A, we can clearly see that because China did not know much about the way the virus spreads and the types of the virus at the beginning of the epidemic, it was unable to control the spread of the epidemic well. At the same time, during the Spring Festival travel period, the large passenger volume of railways and the scattered direction of people's traffic

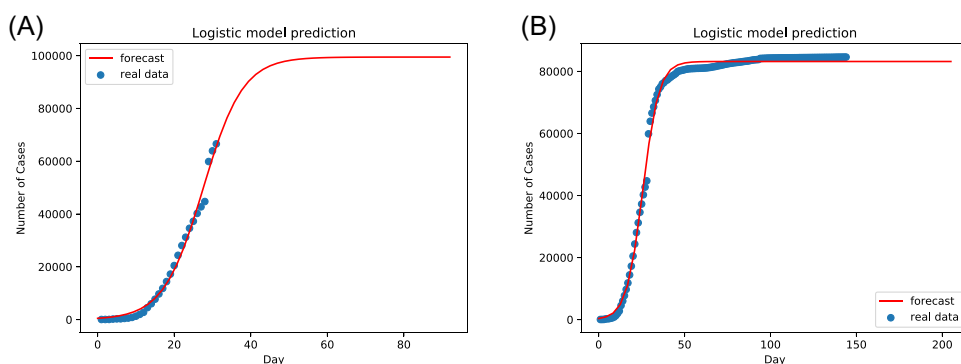


FIGURE 8 US epidemic trend chart: (A) prestage and (B) late stage [Color figure can be viewed at wileyonlinelibrary.com]

activities also led to a rapid increase in the number of confirmed diagnoses about a month after the outbreak.

It can be seen from Figure 8B that Chinese researchers have continuously deepened the research on the transmission mode of the new coronavirus, and the medical level has been continuously improved. At present, the Chinese people's awareness of prevention has increased, and the number of diagnosed daily is increasing slowly, and the peak value has also decreased a lot compared with the value predicted at the beginning of the epidemic. Judging from the fitting results, the current epidemic situation in China has come to an end, and there will be no major fluctuations and turning points.

5.3 | Forecast of SEIR model based on multiple features

5.3.1 | Establish SEIR model

Although the logistic model based on a single feature can predict the development trend of the epidemic well, its reference indicators are limited to confirmed cases. However, in the development of the epidemic, there are many factors that affect the development of the epidemic. In addition to the number of confirmed cases, the number of other patients (such as susceptible, latent, and recovery) also play a key role in the development of the epidemic. Therefore, it is necessary to consider the prediction model of the development trend of the Novel Coronavirus Pneumonia Epidemic under multiple characteristics.

The classic SEIR model^{46,47} is a classic infectious disease model invented at the beginning of the last century. The model can roughly reflect the development process of infectious diseases, and its core is the differential equation. First of all, the model divides the population within the epidemic range of infectious diseases into four categories: category *S*, susceptible; category *E*, latent; category *I*, diagnosed; category *R*, migrant.^{20,48} The schematic diagram of the SEIR model is shown in Figure 9.

1. The total number of people in the survey area, N , remains unchanged, that is, life, death, or migration is not considered.
2. Susceptible persons (Type *S*) and confirmed persons (Type *I*) will become latent persons (Type *E*) after effective contact, and latent persons (Type *E*) will become diagnosed persons (Type *I*) after the average incubation period; confirmed persons (Type *I*) can be cured and

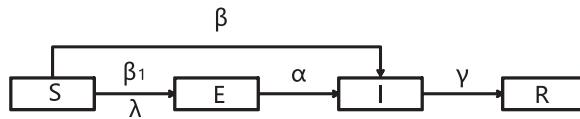


FIGURE 9 Schematic diagram of the SEIR model. SEIR, Susceptible Exposed Infected Removed.

become a convalescent (Type *R*) after being cured. The convalescent (Type *R*) is no longer susceptible to life-long immunity.²

3. Record the proportions of *S*, *E*, *I*, and *R* groups on day *t* as *s(t)*, *e(t)*, *i(t)*, *r(t)*, and the numbers are, respectively, *S(t)*, *E(t)*, *I(t)*, *R(t)*. When the initial date *t* = 0, the initial values of the proportions of various groups of people are *S*₀, *E*₀, *I*₀, *R*₀.
4. β is the probability of a susceptible person being infected by an infected person.
5. β_1 is the probability that a susceptible person is infected by a latent person.
6. α is the probability of a latent person being transformed into an infected person, the reciprocal of the average incubation period *Y*.
7. λ is the average number of susceptible persons effectively contacted by each sick person every day.
8. ν is the probability of an infected person recovering.

On the basis of the above conditions, the mechanism of the change in the number of people can be obtained, which can be expressed in Equation (1).

$$\begin{aligned}
 \frac{dS}{dt} &= -\frac{\lambda(\beta I + \beta_1 E)S}{N}, \\
 \frac{dE}{dt} &= \frac{\lambda(\beta I + \beta_1 E)S}{N} - \alpha E, \\
 \frac{dI}{dt} &= \alpha E - \gamma I, \\
 \frac{dR}{dt} &= \gamma I.
 \end{aligned}
 \tag{1}$$

The iterative is in Equation (2).

$$\begin{aligned}
 S_n &= S_{n-1} - \frac{\lambda(\beta I_{n-1} + \beta_1 E_{n-1})S_{n-1}}{N}, \\
 E_n &= E_{n-1} + \frac{\lambda(\beta I_{n-1} + \beta_1 E_{n-1})S_{n-1}}{N} - \alpha E_{n-1}, \\
 I_n &= I_{n-1} + \alpha E_{n-1} - \gamma I_{n-1}, \\
 R_n &= R_{n-1} + \gamma I_{n-1}.
 \end{aligned}
 \tag{2}$$

5.3.2 | Establish infection rate model

To solve the differential equation in the previous section, we need to know *S*₀, *I*₀ (initial value), infection rate β , β_1 , and recovery rate γ . To find the infection rate β , it is necessary to establish the relationship between it and the distance. Use the point *P*(*x_i*, *y_i*) on the plane coordinate system to describe the individual. Assuming that the distance between any two bodies *P*(*x_i*, *y_i*)

and $Q(x_j, y_j)$ is D_{ij} , if Q is a virus carrier, the infection rate of Q causing P to infect the virus is P_{ij} . The relationship between the two is shown in Equation (3).

$$P_{ij}(D_{ij}) = \begin{cases} (1 + W_{ij})P_{0i} & \text{if } D_{ij} \leq D_{0i}, \\ (1 + W_{ij})P_{0i}\alpha_i^{-(D_{ij}-D_{0i})} & \text{if } D_{ij} > D_{0i}. \end{cases} \quad (3)$$

Among them, D_{0i} is the critical distance of individual $P(x_i, y_i)$. W_{ij} reflects the intimacy of individual P and Q , and its value is 0.1. P_{0i} is the infection rate of P within the critical distance when $W_{ij} = 0$. It is assumed that the infection rate of an individual within the critical range is determined and defined as the basic infection rate. α_i is a factor greater than 1. The distance referred to here is a generalized distance. It does not refer to the actual distance between two individuals in reality. It can be considered as a comprehensive distance between two individuals in many aspects, including the space–time relationship between individuals, intimacy relationship, resistance ability, and so forth.

When $D_{0i} = 5$, $\alpha_i = 2$, $W_{ij} = 0$, and $W_{0i} = 0.5$, according to the above relationship, Python can be used to plot the COVID-19 virus infection rate in Figure 10.

It can be seen from Figure 10 that the basic infection rate of COVID-19 virus is 0.038 in the natural state ($W_{ij} = 0.5$). After the isolation measures were taken, W_{ij} was reduced to 0, and the basic infection rate was also reduced to 0.028. The basic infection rate in the natural state is taken as the infection rate of the SEIR model, that is, $\beta = 0.038$.

6 | RESULTS AND DISCUSSION

In this article, the initial value of the infected population is set to 1, and the initial value of the susceptible population is set to $N - 1$, where N is the total population, and the initial value of the restored population is set to 0. Since the incubation period of the disease is 14 days at most, $\gamma = \frac{1}{14}$ is taken. $\beta = 0.038$, $\beta_1 = 0.02$. As a result, simulation modeling is performed on MATLAB, and the generated result is shown in Figure 11.

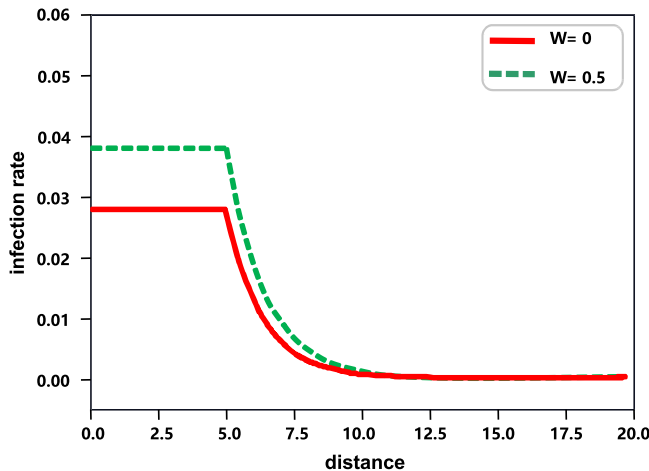


FIGURE 10 Trends in COVID-19 infection rates with distance. COVID-19, Corona Virus Disease 2019. [Color figure can be viewed at wileyonlinelibrary.com]

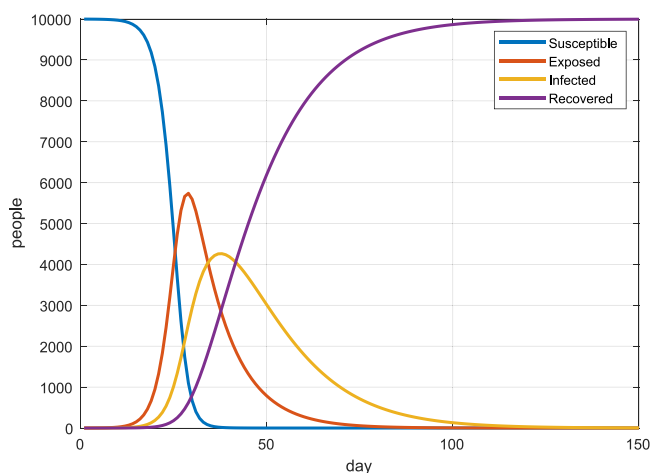


FIGURE 11 SEIR model predicts COVID-19 development (in natural state). COVID-19, Corona Virus Disease 2019; SEIR, Susceptible Exposed Infected Removed. [Color figure can be viewed at wileyonlinelibrary.com]

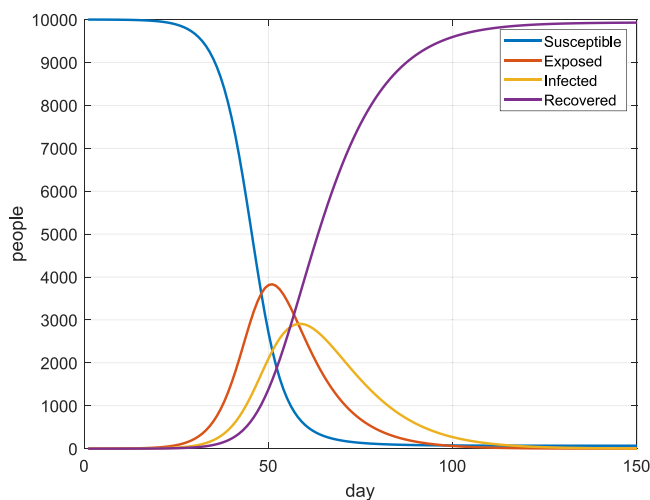


FIGURE 12 SEIR model predicts COVID-19 development (improved medical standards). COVID-19, Corona Virus Disease 2019; SEIR, Susceptible Exposed Infected Removed. [Color figure can be viewed at wileyonlinelibrary.com]

It can be found from Figure 11 that in the natural state, the overall propagation is basically a normal distribution. Over time, the recovered population will gradually be higher than the susceptible population, and the total number will always be the same. But without intervention, it can be observed that the peak of transmission accounts for a large proportion of the total population. With the national medical team and the special medicine developed, the cure rate will be significantly improved. Assuming that γ rises to 0.1, the result is shown in Figure 12.

By observing Figure 12, we can find that not only the proportion of infected people has been greatly reduced, but also the time when the proportion of recovered patients surpassed those

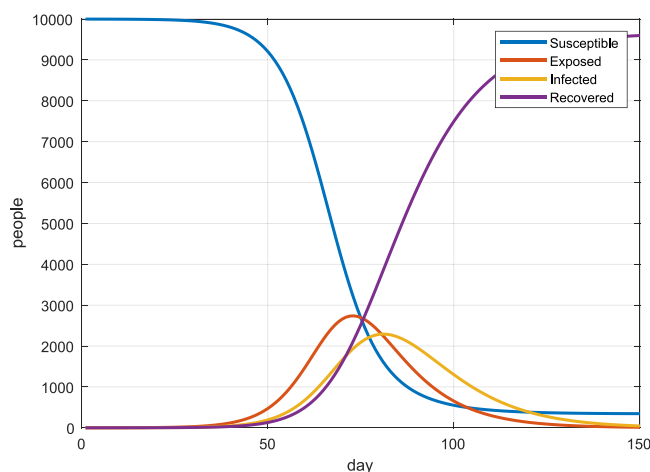


FIGURE 13 SEIR model predicts COVID-19 development (increased isolation measures). COVID-19, Corona Virus Disease 2019; SEIR, Susceptible Exposed Infected Removed. [Color figure can be viewed at wileyonlinelibrary.com]

infected has also been greatly advanced, and the development trend of the epidemic has gradually eased.

From Figure 13, we can clearly see that after the addition of isolation measures, the peaks of the susceptible population and the confirmed population have decreased, indicating that isolation is an effective way to suppress the spread of the virus. According to the results of the modeling, it is not difficult to see that taking quarantine measures in advance, increasing the intensity of quarantine, and timely follow-up of medical resources are the key factors to reduce the number of confirmed cases and the number of latent persons. Therefore, the reference significance and value of the model can be reflected here, and it can also reflect the foresight and correctness of our early decision-makers.

But the SEIR model is a one-way model, the number of susceptible people is constantly being imported into the number of infected people, and the number of infected people is also being imported into the number of recovered people, so the number of susceptible people and the number of infected people will drop to 0 in the end. At the same time, everyone is going to be the recovery, and that is the limitation of the SEIR model.

In the SEIR model, there are initial conditions such as the initial values s_0 , e_0 , i_0 , r_0 , and so forth of the proportion of various groups of people. Next, we will show the influence of these initial conditions on the trend of the SEIR model in the form of a graph. Considering the actual situation, there are no people who have recovered at the initial stage of the epidemic, and the proportion of latent people is often higher than that of confirmed cases. We assume $e_0/i_0 = 2$ and $r_0 = 0$. Investigate the spread of the epidemic in different i_0 .

It can be seen from Figure 14 through the simulation of different initial conditions of the proportion of patients and latent persons under this parameter, the initial conditions of the proportion of patients and latent persons have a direct impact on the time of occurrence, peak, and end of the epidemic, but it has little effect on the shape and characteristics of the epidemic curve. The epidemic curve under different initial value conditions almost shifts along the time index. This shows that if human intervention such as treatment, prevention, and control is not carried out, the spread of the epidemic has little relationship with the proportion of initially sick and latent people.

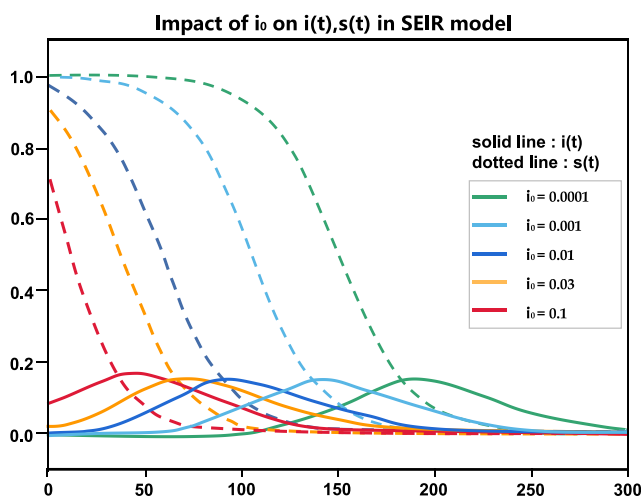


FIGURE 14 Impact of i_0 on $i(t)$, $s(t)$ in the SEIR model. SEIR, Susceptible Exposed Infected Removed. [Color figure can be viewed at wileyonlinelibrary.com]

7 | CONCLUSION AND FUTURE WORKS

In the face of a sudden epidemic, big data can play a key role and can quickly provide a basis for decision-making.^{31,49-51} The visual presentation mode, based on the picture form, displays the data in a deeper level. This article visualizes the epidemic data in China and the world, so that readers can understand the current epidemic situation more clearly and intuitively. For China and the world, the epidemic is still a topic that needs to be focused on. Therefore, on the basis of visual analysis, a large amount of actual data is used to establish a logistic growth model based on existing mathematical knowledge to realize the fitting of factual data and predictive models. At the same time, mathematical models of infectious diseases are used to predict future epidemics, so as to better prevent and respond to epidemics. By establishing the SEIR infectious disease model, the changes in the data can be well reflected, and the current situation can also be analyzed based on the existing diagnosis data. Through these analyses, we summarize experience and lessons, and adopt a series of effective measures to control the development of infectious diseases in the future, thereby reducing the amount of infections.

We use a combination of visual analysis and modeling and prediction to help readers understand the current situation of the epidemic from multiple perspectives and predict the progress of the epidemic. This study still has certain limitations, it can only make a rough forecast of the future development of the epidemic, and cannot predict the development of the epidemic with 100% accuracy. Due to time reasons, we will combine epidemic data with machine learning in the follow-up, improve the accuracy of the prediction model by training the model, and further analyze the epidemic data.

ACKNOWLEDGMENTS

This study is supported by the Foundation of National Natural Science Foundation of China (Grant Nos. 62072273, 72111530206, 61962009, 61873117, 61832012, 61771231, and 61771289), The Major Basic Research Project of Natural Science Foundation of Shandong Province of China (Grant No. ZR2019ZD10), Natural Science Foundation of Shandong Province (Grant No. ZR2019MF062),

Shandong University Science and Technology Program Project (Grant No. J18A326), Guangxi Key Laboratory of Cryptography and Information Security (Grant No. GCIS202112), The Major Basic Research Project of Natural Science Foundation of Shandong Province of China (Grant No. ZR2018ZC0438), Major Scientific and Technological Special Project of Guizhou Province (Grant No. 20183001), Foundation of Guizhou Provincial Key Laboratory of Public Big Data (Grant No. 2019BD-KFJJ009), and Talent project of Guizhou Big Data Academy, Guizhou Provincial Key Laboratory of Public Big Data ([2018]01).

DATA AVAILABILITY STATEMENT

No. Research data are not shared.

ORCID

Yiting Zhang  <http://orcid.org/0000-0002-8729-830X>

REFERENCES

1. Khan G. A novel coronavirus capable of lethal human infections: an emerging picture. *Virol J.* 2013;10:1-6.
2. Zhu N, Zhang D, Wang W, Li X, Tan W. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727-733.
3. Yu X, Wang H, Zheng X, Wang Y. Effective algorithms for vertical mining probabilistic frequent patterns in uncertain mobile environments. *Int J Ad-Hoc Ubiquitous Comput.* 2016;23(3-4):137-151.
4. Zheng X, Liu H. A scalable coevolutionary multi-objective particle swarm optimizer. *Int J Intell Syst.* 2010;3(5):590-600.
5. Ganapathy S, Kulothungan K, Muthurajkumar S, Vijayalakshmi M, Yogesh P, Kannan A. Intelligent feature selection and classification techniques for intrusion detection in networks: a survey. *EURASIP J Wireless Commun Networking.* 2013;2013(1):1-16.
6. Wu C, Li W. Enhancing intrusion detection with feature selection and neural network. *Int J Intell Syst.* 2021;36(7):3087-3105.
7. Shaban WM, Rabie AH, Saleh AI, Abo-Elsoud M. A new COVID-19 patients detection strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. *Knowl-Based Syst.* 2020;205:106270.
8. Ganapathy S, Vijayakumar P, Yogesh P, Kannan A. An intelligent CRF based feature selection for effective intrusion detection. *Int Arab J Inf Technol.* 2016;13(1):44-50.
9. Kanimozhi U, Ganapathy S, Manjula D, Kannan A. An intelligent risk prediction system for breast cancer using fuzzy temporal rules. *Natl Acad Sci Lett.* 2019;42(3):227-232.
10. Narin A. Accurate detection of COVID-19 using deep features based on X-Ray images and feature selection methods. *Comput Biol Med.* 2021;137(March):104771.
11. Wang S, Zhang Y, Cheng X, Zhang X, Zhang YD. PSSPNN: PatchShuffle stochastic pooling neural network for an explainable diagnosis of COVID-19 with multiple-way data augmentation. *Comput Math Methods Med.* 2021;2021:6633755.
12. Habibzadeh P, Stoneman EK. The novel coronavirus: A bird's eye view. *Int J Occup Environ Med.* 2020;11(2):65-71.
13. LeungCK, Chen Y, Hoi CS, Shang S, Wen Y, Cuzzocrea A. *Big Data Visualization and Visual Analytics of COVID-19 Data.* Vol 134. IEEE; 2020:415-420.
14. Rehouma R, Buchert M, Chen YP. Machine learning for medical imaging-based COVID-19 detection and diagnosis. *Int J Intell Syst.* 2021;36(9):5085-5115.
15. Li J, Ye H, Li T, et al. Efficient and secure outsourcing of differentially private data publishing with multiple evaluators. *IEEE Trans Dependable Secure Comput.* 2022;19(1):67-76.
16. Yan H, Chen M, Hu L, Jia C. Secure video retrieval using image query on an untrusted cloud. *Appl Soft Comput.* 2020;97:106782.
17. Almeshal AM, Almazrouee AI, Alenizi MR, Alhajeri SN. Forecasting the spread of COVID-19 in Kuwait using compartmental and logistic regression models. *Appl Sci.* 2020;10(10):3402.

18. Chen DG, Chen X, Chen JK. Reconstructing and forecasting the COVID-19 epidemic in the United States using a 5-parameter logistic growth model. *Global Health Res Policy*. 2020;5(1):1-7.
19. Read JM, Bridgen JR, Cummings DA, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *Philos. Trans. R. Soc. Lond., B, Biol. Sci*. 2020;376(1829):20200265. doi:10.1101/2020.01.23.20018549
20. Pandey G, Chaudhary P, Gupta R, Pal S. Machine learning models for government to predict COVID-19 outbreak. *Digital Government: Res. Pract.* 2020;1(4):1-6. doi:10.48550/arXiv.2004.00958
21. Dur-E-Ahmad M, Imran M. Dynamics model of coronavirus COVID-19 for the outbreak in most affected countries of the world. *Int J Interactive Multimedia Artif Intell*. 2020;6(2):7-11.
22. López L, Rodo X. A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics. *Results Phys*. 2021;21:103746.
23. Annas S, Pratama MI, Rifandi M, Sanusi W, Side S. Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. *Chaos Solitons Fractals*. 2020;139:110072. doi:10.1016/j.chaos.2020.110072.pdf
24. Radulescu A, Williams C, Cavanagh K. Management strategies in a SEIR-type model of COVID 19 community spread. *Sci Rep*. 2020;10(1):1-16.
25. Kumar G, Singh UP, Jain S. Hybrid evolutionary intelligent system and hybrid time series econometric model for stock price forecasting. *Int J Intell Syst*. 2021;36(9):4902-4935.
26. Chen B, Shi M, Ni X, Ruan L, Jiang H, Yao H. Visual data analysis and simulation prediction for COVID-19. *Int J*. 2020;6(1):95-114. doi:10.48550/arXiv.2002.07096
27. Li T, Li J, Chen X, Liu Z, Lou W, Hou YT. NPMML: a framework for non-interactive privacy-preserving multi-party machine learning. *IEEE Trans Dependable Secure Comput*. 2020;18(6):2969-2982.
28. Kuang X, Zhang M, Li H, Zhao G, Cao H, Wu Z. *DeepWAF: Detecting Web Attacks Based on CNN and LSTM Models*. Springer; 2019:121-136.
29. Cai J, Wang Y, Liu Y, Luo JZ, Wei W, Xu X. Enhancing network capacity by weakening community structure in scale-free network. *Future Gener Comput Syst*. 2018;87:765-771.
30. Ge C, Susilo W, Baek J, Liu Z, Fang L. Revocable attribute-based encryption with data integrity in clouds. *IEEE Trans Dependable Secure Comput*. Published online March 17, 2021. doi:10.1109/TDSC.2021.3065999
31. Ge C, Susilo W, Baek J, Liu Z, Fang L. A verifiable and fair attribute-based proxy re-encryption scheme for data sharing in clouds. *IEEE Trans Dependable Secure Comput*. Published online April 29, 2021. doi:10.1109/TDSC.2021.3076580
32. Ajinaja M, Olarinde M, Olaniyan L, Adeola C. Data analytics and visualization of coronavirus COVID-19 epidemic in Nigeria based on recovered and death cases. *Soc. Sci. Electron. Publ*. 2020. <https://ssrn.com/abstract=3632420>
33. Chen C, Huang T. Camdar-adv: generating adversarial patches on 3D object. *Int J Intell Syst*. 2021;36(3):1441-1453.
34. Biswas P, Saluja KS, Arjun S, Murthy L, Prabhakar G, Sharma VK. COVID-19 data visualization through automatic phase detection. *Digital Gov: Res Pract*. 2020;1(4):1-8.
35. Ndiaye BM, Balde MA, Seck D. Visualization and machine learning for forecasting of COVID-19 in Senegal. *arXiv preprint arXiv: 2008.03135*. 2020. <https://arxiv.org/pdf/2008.03135.pdf>
36. Teh JKL, Bradley DA, Chook JB, Lai KH, Ang WT, Teo KL. Multivariate visualization of the global COVID-19 pandemic: a comparison of 161 countries. *PLoS ONE*. 2021;16:e0252273. <https://doi.org/10.1371/journal.pone.0252273.pdf>
37. Costa PJ. *Applied Mathematics for the Analysis of Biomedical Data: Models, Methods, and MATLAB*. John Wiley & Sons; 2017. doi:10.1002/9781119269540
38. Jiang N, Jie W, Li J, Liu X, Jin D. GATrust: a multi-aspect graph attention network model for trust assessment in OSNs. *IEEE Trans Knowl Data Eng*. Published online May 10, 2022. doi:10.1109/TKDE.2022.3174044
39. Gao C, Li J, Xia S, Choo KKR, Lou W, Dong C. Mas-encryption and its applications in privacy-preserving classifiers. *IEEE Trans Knowl Data Eng*. 2020;34(05):2306-2323. doi:10.1109/TKDE.2020.3009221
40. Ai S, Hong S, Zheng X, Wang Y, Liu X. CSRT rumor spreading model based on complex network. *Int J Intell Syst*. 2021;36(5):1903-1913.

41. Yan H, Hu L, Xiang X, Liu Z, Yuan X. PPCL: privacy-preserving collaborative learning for mitigating indirect information leakage. *Inf Sci.* 2021;548:423-437.
42. Li J, Hu X, Xiong P, Zhou W. The dynamic privacy-preserving mechanisms for online dynamic social networks. *IEEE Trans Knowl Data Eng.* 2020;34(06):2962-2974. doi:10.1109/TKDE.2020.3015835
43. Li J, Huang Y, Wei Y, Lv S, Liu Z, Dong C. Searchable symmetric encryption with forward search privacy. *IEEE Trans Dependable Secure Comput.* 2019;18(1):460-474.
44. Tianqing Z, Zhou W, Ye D, Cheng Z, Li J. Resource allocation in IoT edge computing via concurrent federated reinforcement learning. *IEEE Internet Things J.* 2021;9(2):1414-1426.
45. Mo K, Tang W, Li J, Yuan X. Attacking deep reinforcement learning with decoupled adversarial policy. *IEEE Trans Dependable Secure Comput.* Published online January 18, 2022. doi:10.1109/TDSC.2022.3143566
46. Arcede JP, Caga-Anan RL, Mentuda CQ, Mammari Y. Accounting for symptomatic and asymptomatic in a SEIR-type model of COVID-19. *Math Modell Nat Phenom.* 2020;15:34.
47. Meng W, Li W, Wang Y, Au MH. Detecting insider attacks in medical cyber-physical networks based on behavioral profiling. *Future Gener Comput Syst.* 2020;108:1258-1266.
48. Hu L, Yan H, Li L, Pan Z, Liu X, Zhang Z. MHAT: an efficient model-heterogeneous aggregation training scheme for federated learning. *Inf Sci.* 2021;560:493-503.
49. Li T, Wang Z, Yang G, Cui Y, Chen Y, Yu X. Semi-selfish mining based on hidden Markov decision process. *Int J Intell Syst.* 2021;36(7):3596-3612. doi:10.1002/int.22428
50. Yuan F, Chen S, Liang K, Xu L. *Research on the Coordination Mechanism of Traditional Chinese Medicine Medical Record Data Standardization and Characteristic Protection Under Big Data Environment.* Vol 1, 1st ed. Shandong People's Publishing House; 2021.
51. Li T, Wang Z, Chen Y, Li C, Jia Y, Yang Y. Is semi-selfish mining available without being detected? *Int J intell syst.* 2021. doi:10.1002/int.22656.

How to cite this article: Wang Y, Zhang Y, Zhang X, Liang H, Li G, Wang X. An intelligent forecast for COVID-19 based on single and multiple features. *Int J Intell Syst.* 2022;37:9339-9356. doi:10.1002/int.22995