
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Computers and Electrical Engineering 102 (2022) 108230

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

# SQEIR: An epidemic virus spread analysis and prediction model

Yichun Wu [a], Yaqi Sun [a,b,*], Mugang Lin [a,b]

[a] College of Computer Science and Technology, Hengyang Normal University, Hengyang, 421002, China
[b] Hunan Provincial Key Laboratory of Intelligent Information Processing and Application, Hengyang, 421002, China

ARTICLE INFO

*Keywords:*
Susceptible quarantined exposed infective
Infectious disease model
Mathematical model
COVID-19

ABSTRACT

In 2019, a new strain of coronavirus pneumonia spread quickly worldwide. Viral propagation may be simulated using the Susceptible Infectious Removed (SIR) model. However, the SIR model fails to consider that separation of patients in the COVID-19 incubation stage entails difficulty and that these patients have high transmission potential. The model also ignores the positive effect of quarantine measures on the spread of the epidemic. To address the two flaws in the SIR model, this study proposes a new infectious disease model referred to as the Susceptible Quarantined Exposed Infective Removed (SQEIR) model. The proposed model uses the weighted least squares for the optimal estimation of important parameters in the infectious disease model. Based on these parameters, new differential equations were developed to describe the spread of the epidemic. The experimental results show that this model exhibits an accuracy 6.7% higher than that of traditional infectious disease models.

## 1. Introduction

In the history of mankind, infectious diseases have consistently been the main culprit of human life, affecting the lives of individuals, the peril of nations, and the direction of history. The infectious disease model started in 1760 when Daniel Bernoulli [1] studied vaccination against smallpox in one of his papers. The advancement of the real deterministic mathematical model [2] of infectious diseases began in the early part of the 20th century. Hamer and Ross [3], among others, conducted considerable research to establish mathematical models of infectious diseases. In 1927, Kermack and McKendrick [4] proposed the Susceptible Infected Recovered (SIR) warehouse model while studying the black death epidemic in London. Based on these models, the threshold theory [5] in the dynamics of infectious diseases was proposed. The SIR model is the most common and basic among infectious disease models, with a foundational contribution to the study of infectious disease dynamics.

On January 26, 2020, the National Health Commission announced that COVID-19 during the incubation period is contagious [6–7]. The population division of the SIR model does not apply to the COVID-19 virus. This study proposes a novel infectious disease model, the Susceptible Quarantined Exposed Infective Removed (SQEIR) model, which adds newly exposed persons and considers the quarantine period. Both infected and exposed persons pose a threat to susceptible persons because of their infectivity, which is more consistent with the characteristics of the "new crown pneumonia" epidemic. The new infectious disease model proposed in this study builds an E(Exposed) warehouse, reestablishes a balance equation for the population in each warehouse over time, and re-predicts the epidemic trend.

The remainder of this paper is as follows. The second part briefly introduces the work related to the establishment of the SIR model.

* Corresponding author.
*E-mail address:* happysyq@hynu.edu.cn (Y. Sun).

https://doi.org/10.1016/j.compeleceng.2022.108230
Received 16 March 2022; Received in revised form 29 June 2022; Accepted 10 July 2022
Available online 10 August 2022

The third part introduces the SQEIR model, which re-predicts the epidemic situation. The fourth part presents experiments to explore the fit between the model and the actual situation of the epidemic. Finally, in the fifth part, a conclusion is drawn.

Compared with other infectious diseases models, the SQEIR model has the following major contributions:

(1) To consider the effect of the incubation period of the virus and the patient quarantine period on the prevention and control of the epidemic, we proposed a novel infectious disease model, the SQEIR model, to predict the spread of the epidemic.
(2) The infection rate and recovery rate are two essential factors in the infectious illness model. To estimate the parameters, we applied the weighted least squares to optimize the predicted number of daily infections.
(3) The distribution of infected, recovered, susceptible, and incubated patients under quarantine measures was described using a novel differential equation, which enhanced the accuracy of model predictions.

We will introduce the related work of this paper in the next Section 2, and describe the proposed method, the improved model and the comparison between the models in Section 3. Section 4 shows the effect of important parameters in the model and the accuracy of the improved model, and finally is the conclusion of this work.

## 2. Related Work

### 2.1. SIR infectious disease model

People have a long history of understanding and studying infectious diseases. As early as the Qin dynasty, leprosy patients were isolated to break the transmission path of infectious diseases and thus control and eliminate epidemics. Epidemiological theories, also known as mathematical models [8] of epidemiology, were developed during the "struggle" between human beings and infectious diseases. According to relevant historical records, the use of mathematics to study infectious diseases began as early as 1760 during a large-scale outbreak of smallpox [9]. The French mathematician Daniel Bernoulli [10] established the world's earliest mathematical epidemiological chamber model. After 150 years, in the early 20th century, the use of deterministic models to study infectious diseases began in earnest.

In the dynamics of infectious diseases, the infectious disease warehouse model [11] is one of the most commonly used models. The initial trial model, referred to as Kermack–Mckendrick theory [12], is relatively complex and considers the infectious recovery ability of different age groups. Its differential equation is shown in Eq. (1).

$$\begin{cases} \dfrac{ds}{dt} = -\lambda S \\ \dfrac{\partial i}{\partial t} + \dfrac{\partial i}{\partial a} = \delta(a)\lambda S - \gamma(a) i \\ \dfrac{dR}{dt} = \displaystyle\int_0^{\infty} \gamma(a) i(a,t) da \end{cases} \tag{1}$$

When age—that is, $\beta(a)$, is a constant, which is in the Dirac function [13] $\lambda = \int_0^{\infty} \beta(a) i(a,t) da$, $I(t) = \int_0^{\infty} i(a,t) da$. The SIR infectious disease model [14] can be obtained, with its differential equation shown in Eq. (2).

$$\begin{cases} \dfrac{dS}{dt} = -\beta SI \\ \dfrac{dI}{dt} = \beta SI - \gamma I \\ \dfrac{dR}{dt} = \gamma I \end{cases} \tag{2}$$

In 1927, W.O. Kermack and A.G. McKendrick proposed a complete mathematical model [15] to investigate the spread of infectious diseases by studying the Black Death that ravaged London in the 17th century and the plague that was prevalent in Mumbai in the early 20th century, and established an infectious disease warehouse model—the SIR infectious disease model [16]. Subsequently, numerous scientists studying epidemic diseases presented different models, in addition to the SIR model. Similar to the SIS model, the SIR model is typically used to describe virus transmission [17]. The SIR model is more suitable for describing bacterial transmission. However, the SIS infectious disease model cannot be used to predict infectious diseases, hence the broader application of the SIR model as a current virus model. The SIR model is the most classic and basic model among infectious disease models, with a foundational contribution to the study of infectious disease dynamics. The SIR model is still widely used and continuously developed.

The SIR model is used to divide the population in the natural state into several warehouses, and the individuals at different stages of infectious diseases represent the corresponding warehouses. The SIR model categorizes the total population into the following [18]: The number of susceptibles, $S(t)$, representing the number of people not infected at time t but may be infected by this type of disease; the number of infectives, $I(t)$, indicating the number of people at time t who have been infected and have become patients and are infectious; the number of recovered people, $R(t)$, referring to the number of people who have been removed from the infected at time t. Its transmission mechanism is presented in Fig.1.

As shown in Fig. 1, a set of differential equations can be established to represent the marginal changes in the three populations over time, with $\beta$ as the infection coefficient [19] and $\gamma$ as the cure coefficient [20]. The SIR model can be used to predict various infectious diseases but is not sufficiently detailed to classify the population and does not explicitly consider quarantine as a factor. In reality, isolation of suspected patients is an effective means to control the spread of the epidemic. The model has no feedback mechanism. In the prediction process, the accuracy is reduced if the data within a long period are predicted only based on the existing data. Moreover, the solution to the differential equations is difficult to obtain and sensitive to initial values, exerts a substantial effect on the robustness of the model.

The SIR infectious disease model is deficient in population division because of the transmissibility of COVID-19 during the incubation period. In the following study, we redivided the population distribution and considered the effects of isolation, vaccination, and other measures on the spread of the epidemic. We then established a new virus model based on the findings.

### 2.2. Parameter estimation

Not all facts on the infectious process of an epidemic can be acquired. Data such as the number of new infections and the number of recovered persons daily can usually be observed. However, in the SIR infectious disease model, data such as the number of susceptible people, transmission intensity, and removal strength are unknown. Behind the two parameters is the basis of the equation, which is important. Their parameters need to be estimated.

Parameter estimation refers to the process of estimating model parameters based on the input and output data of the known model [21]. Methods of parameter estimation vary, and the most commonly used techniques can be classified into moment estimation, the least squares method, and the maximum likelihood method. Considered ideal, the least squares method is increasingly used by researchers to estimate parameters. The least squares method is an approach to parameter estimation, which estimates the parameter value by minimizing the sum of squares of the difference between the sample data and the estimated value. It performs parameter estimation or system identification of regression models from the perspective of error fitting. The least squares method has been widely used in various fields, such as parameter estimation [22], system identification, and prediction [23].

The least squares method is usually credited to Gauss but was first published by Legendre [24]. The least square is the best fit—that is, the minimization of the sum of the squares of residuals (the gap between the observed value and the fitting value provided by the model) [25]. In the virus model, the variables in the research problem exhibit a nonlinear relationship. Parameter estimation using the nonlinear least squares method [26]. As indicated in [27], if a linear regression model $\overline{Y}_i(\widehat{a}_{(0)}) = Z_i(\widehat{a}_{(0)})\widehat{a} + \varepsilon_i$ is constructed, the least squares method can be used to estimate the parameters, thus obtaining an approximate estimate $\widehat{a}_{(1)}$ of the parameter values of the original nonlinear model. If the obtained value meets the requirements, the calculation ends; otherwise, the new given value $\widehat{a}_{(1)}$ of as $\widehat{a}$ is followed by Taylor expansion, and the same aforementioned operation is performed to determine the second iteration value $\widehat{a}_{(2)}$ and so on, until convergence.

In the SIR model, the daily number of infected persons is the most important data; thus, the unknown parameters are estimated optimally by minimizing the sum of squares of residuals of its observed value and predicted value. In the general least squares method, each data point in the time series is equal in importance; however, data in different time series often vary in significance in practice. Specifically, virus models simulating the spread of an epidemic are time-sensitive, and recent data are often more important than long-term data. Thus, in this study, we introduce the weighted least squares.

## 3. Mainstream infectious disease models and optimization

### 3.1. SQEIR model formulation and preliminaries

The SIR model divides the population into infected ($I$), susceptible ($S$), and recovered ($R$). This population division is used to study the spread of COVID-19 with a large error, considering that COVID-19 has an incubation period. To redivide the population, we build a new model, SQEIR.

The study determined that the new virus model SQEIR requires three premises, as follows:

(1) Assume that Coronavirus Disease 2019(COVID-19) occurs in a closed environment. We consider the population as a constant $N$, regardless of factors such as birth, death, and movement of the population. We denote the total population as $N$ and obtain Eq. (3).

$$S_t + R_t + I_t + E_t = N \tag{3}$$

(2) In the unit time at time $t$, $\beta$ is expressed as infection intensity. The number of susceptible persons that can be infected by an infected person is proportional to the total number of susceptible persons S at this time. The number of newly infected persons at time $t$ is determined as $\beta S_t I_t$.

(3) In the unit time at time t, the number of recovered persons is proportional to the number of patients. The strength of removal is $\gamma$, and the number of recovered people from all infected persons per unit time at time t is determined as $\gamma I_t$.


**Fig. 1.** SIR model infection mechanism.

In the study of infectious diseases, some epidemiological indicators are typically usually to measure its transmissibility. Among the important indicators are the basic reproduction number ($R$) and the effective reproduction number ($R_t$).

The basic reproduction number is calculated as $R = \frac{S_0}{\mu} = S_0\frac{\beta}{\gamma}$. The basic reproduction number R consists of three parameters: susceptible persons, infection intensity, and removal intensity. When $R > 1$, the number of newly infected persons per unit time exceeds the number of newly removed persons per unit time, hence the enlarged scale of the infectious disease. By contrast, when $R \leq 1$, infectious diseases tend to die out. Thus, the basic reproduction number is an important indicator for assessing the trend of infectious diseases. $R = 1$is the threshold to distinguish whether it erupts or not. This is the threshold theory for studying infectious diseases. Therefore, during the spread of infectious diseases, the number of susceptible people $S_0$, the infection intensity $\beta$, and the removal intensity $\gamma$ are the most critical parameters. From these parameters, the size of the basic reproduction number $R$ is derived, and the development trend of the infectious disease is evaluated.

When the initial number of infected people is negligible relative to the number of susceptible people, the parameter $\mu$ can be approximately expressed as $\mu = \frac{S_0 - S_\infty}{\ln S_0 - \ln S_\infty}$. The basic reproduction number is given by $R = S_0\frac{\ln S_0 - \ln S_\infty}{S_0 - S_\infty}$. The effective reproduction number $R_t$ represents the number of people that a single infected person can infect in unit time and can also be expressed as $R_t = \frac{S_t}{\mu} = S_t\frac{\beta}{\gamma}$.

The data available during the spread of infectious diseases are incomplete. The daily number of newly infected and recovered persons can be obtained from the official website of the National Health Commission of the People's Republic of China. However, other data in the SQEIR virus model, such as the number of susceptible persons $S_t$, infection intensity $\beta$, and removal intensity $\gamma$ remain unknown. The latter two parameters are the basic parameters used to establish the differential equation of the model.

According to the aforementioned threshold theorem, the effective reproduction number consists of $S_t$, $\beta$, and $\gamma$ the threshold that determines the occurrence of an outbreak. Therefore, a realistic estimation of the parameters $S$, $\beta$, and $\gamma$ is crucial. The weighted least squares used to estimate the required parameters is thus presented.

In the infectious disease model, the daily number of infections is the most important data for us. Thus, the unknown parameters are optimally estimated by minimizing the residual square sum of its observations and predictions.

In the SQEIR model, for the unknown parameters $\beta$ and $S_0$, assume $\theta = (\beta, S_0)$; if the parameter $\gamma$ is given, let the observation interval be $[1, M]$. The actual observed number of infected people is $I = \{I_i | 1 \leq i \leq M\}$, the simulated value of the number of infected persons is $Y(\theta) = \{y_i(\theta), 1 \leq i \leq M\}$, and the residual can be expressed as $V = Y(\theta) - I$. The residual sum of squares is thus given in Eq. (4).

$$SSE(\theta) = \| V \|^2 = \sum_{i=1}^{M} [y_i(\theta) - I_i]^2 \tag{4}$$

The goal is to find a set of $\beta$ and $S_0$ to minimize the value of $SSE(\theta)$. The partial derivative of $\beta$ and $S_0$ in Eq. (4) can be determined using Eqs. (5) and (6).

$$\frac{\partial SSE(\theta)}{\partial S_0} = 2\sum_{i=1}^{M}[y_i(\theta) - I_i]\frac{\partial y_i(\theta)}{\partial S_0} \tag{5}$$

$$\frac{\partial SSE(\theta)}{\partial \beta} = 2\sum_{i=1}^{M}[y_i(\theta) - I_i]\frac{\partial y_i(\theta)}{\partial \beta} \tag{6}$$

In the general least squares method, the data in the time series are of equal importance; however, in practice, the data in different time series vary in significance. Recent data are more important than long-term data for various reasons. Therefore, a more reasonable

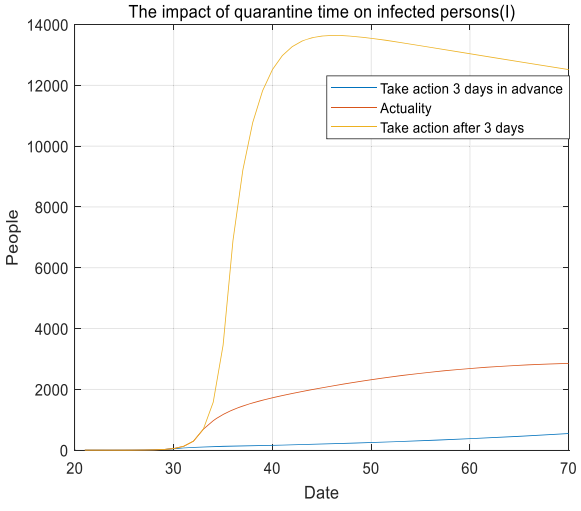


**Fig. 2.** Effect of quarantine measures on infected persons.

method is to use a weighted method—that is, assigning a larger weight to the recent data and a smaller weight to the long-term data. We thus introduce weighted least squares in the SQEIR infectious disease model. This approach uses exponential weights $w^{N-i}$, $0 < w < 1$. The residual sum of squares $SSE(\theta)$ can thus be used to obtain Eq. (7).

$$SSE(\theta) = \| V \|^2 = \sum_{i=1}^{N} w^{N-i}[y_i(\theta) - I_i]^2 + \sum_{i=N+1}^{M} [y_i(\theta) - I_i]^2 \tag{7}$$

### 3.2. Optimization of the SQEIR model

According to the National Health Commission of the People's Republic of China, the new coronavirus is infectious during the incubation period. Therefore, we reclassify the population to establish the SQEIR model. Meanwhile, the government has implemented quarantine measures to control the source of infection and prevent the spread of infectious diseases. We use Huanggang City, Hubei as an example to evaluate study the effect of quarantine measures on infected persons (I). We then compare the data on quarantine measures taken three days in advance and three days later with the actual situation, as shown in Fig. 2.

As shown in Fig. 2, if the infected people are not promptly isolated, the spread of the virus is accelerated. If quarantine measures are implemented in advance, the infection rate is greatly reduced. We can conclude that the quarantine measures adopted by the government influence the spread of the epidemic. Merely quarantining for three days in advance largely affects the number of infected patients; thus, quarantining as soon as possible positively influences epidemic prevention and control.

In the proposed infectious disease model, referred to as the SQEIR model, we consider the effect of quarantine measures on the spread of the epidemic. We introduce three parameters: quarantine ratio ($q$), quarantine release rate ($\lambda$), and the rate of conversion of exposed persons into susceptible persons ($\omega$). We also assume that the infection rate of exposed persons is the same as that of susceptible persons. On the basis of the aforementioned parameters and conditions, we establish a new differential equation to describe the SQEIR model, as shown in Eq. (8).

$$\begin{cases} S_i = S_{i-1} - \dfrac{r\beta I_{i-1}S_{i-1}}{N} - \dfrac{r_1\beta_2 E_{i-1}S_{i-1}}{N} \\ \quad - \dfrac{qr(1-\beta)I_{i-1}S_{i-1}}{N} - \dfrac{qr_1(1-\beta_2)E_{i-1}S_{i-1}}{N} + \varpi EQ_{i-1} \\ E_i = E_{i-1} + \dfrac{r\beta I_{i-1}S_{i-1}}{N} + \dfrac{r_1\beta_2 E_{i-1}S_{i-1}}{N} \\ \quad - \alpha E_{i-1} - qEQ_{i-1} \\ I_i = I_{i-1} + \alpha E_{i-1} - \gamma I_{i-1} + \alpha EQ_{i-1} \\ R_i = R_{i-1} + \gamma I_{i-1} \\ SQ_i = SQ_{i-1} + \dfrac{qr(1-\beta)I_{i-1}S_{i-1}}{N} \\ \quad + \dfrac{qr_1(1-\beta_2)E_{i-1}S_{i-1}}{N} - \lambda SQ_{i-1} \\ EQ_i = EQ_{i-1} + qr_1\beta_2 E_{i-1} - \varpi EQ_{i-1} \end{cases} \tag{8}$$

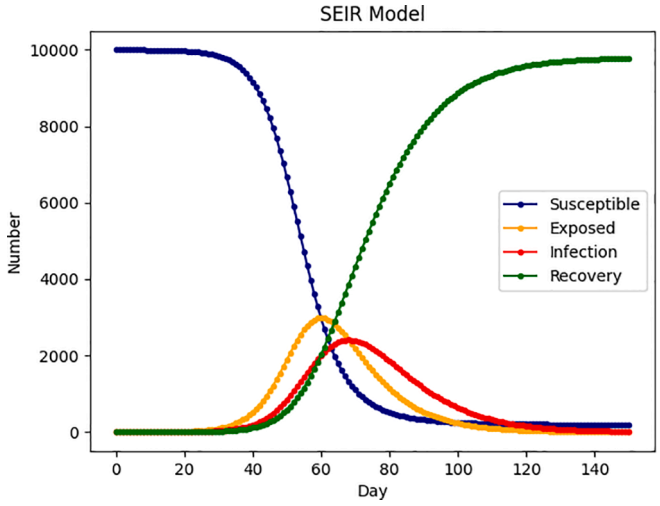


**Fig. 3.** Fitting diagram for the SEIR model.

where $i$ is the number of days COVID-19 has spread. $N = S_i + I_i + R_i + E_i$, where $N$ indicates the total number of people in the area; $S_i$, $E_i$, $I_i$, and $R_i$ represent the number of susceptible persons, the number of exposed persons, the number of infected persons, and the number of recovered persons on the $i$th day, respectively; and $\alpha$ denotes the probability of exposed persons becoming infected. $\beta$, $\beta_2$ is the probability of infection.

With quarantine measures considered, the average number of infected person contacts per day is assumed to be $r = 10$, and the number of exposed person contacts is $r_1 = 8$ in the actual scenario. During the period, the number of susceptible people is equal to the total number minus the number of exposed people, and the change rate of the infected population is $\frac{dI_i}{ddi} = \frac{\beta I_i S_i}{N} - \gamma I_i \approx (\beta - \gamma) I_i$, estimate $\beta = \beta_2 \approx 0.45$ after fitting based on the data released by the National Health and Family Planning Commission during the outbreak period. From $\alpha = \frac{I_i}{E_i}$, $\alpha \approx 0.10$ can be derived using the threshold theorem.

$Y(i)$ is the number of people infected in the last $i$ days; the actual number comes from the National Bureau of Statistics. The incubation period is represented by $T_L$, and the infection period is denoted by $T_i$. The calculation of the generation time $T_g$ can be approximated by the sequence interval between $T_L$ and $T_i$, calculated as $T_g \in (7.0, 8.4)$. $P$ is the ratio of the incubation period to the generation time. We can calculate $P = 0.7$ for COVID-19 from the formula $P = \frac{T_L}{T_g}$. We then derive the basic reproduction number $R_0$ from the formula $R_0 = 1 + \lambda T_g + P(1 - P) \times (\lambda T_g)^2$, where $\lambda = \frac{\ln Y(i)}{i}$ is the early infection rate. Given the same parameters, we compared the SEIR model and the SQEIR model to assess the distribution of different populations during the COVID-19 outbreak, as shown in Figs. 3 and 4.

On the basis of the comparison of Figs. 3 and 4, the SQEIR model we proposed influences the simulation of the spread of the new crown epidemic. The figure evidently shows that the number of infected persons (I) is markedly decreased because of isolation; similarly, the number of exposed persons is reduced to a certain extent. According to the actual scenario, the peak arrival time of the number of infected persons and the number of exposed persons should be basically the same. This observation shows that the SQEIR model is more accurate than the SEIR model.

### 3.3. SQEIR model comparison

The SQEIR model we proposed divides the population into four categories: susceptible persons ($S$), exposed persons ($E$), infected persons ($I$), and recovered persons ($R$). Given that the new coronavirus has an incubation period, we also consider the infectivity of the virus during the incubation period in the SQEIR model. Meanwhile, on the basis of the aforementioned research on quarantine regulations, the measures taken by the government positively affect the spread of the epidemic. Therefore, we define new parameters to evaluate the effect of prevention and control measures on the simulated spread of the epidemic in the SQEIR infectious disease model.

The epidemic data mainly come from the National Health Commission of the People's Republic of China. However, there are still many incomplete places before the SQEIR virus model is applied, which requires preprocessing of the original data. In this study, data preprocessing mainly focuses on data combination; matching and fusion with a third-party database; feature engineering; and filtering out useless feature columns.

The epidemic data repository mainly includes three tables containing time-series data on cumulative confirmed cases, deaths, and cured cases. However, the severity of an epidemic in a region cannot be determined only from the number of confirmed cases but from its corresponding population as well. This information is missing in the original data table. Thus, in this study, the population of the region was acquired using the National Bureau of Statistics census. The region feature columns in the COVID-19 data were matched with those in the population data. However, the two tables labeled the same province differently; consequently, the fields could not be

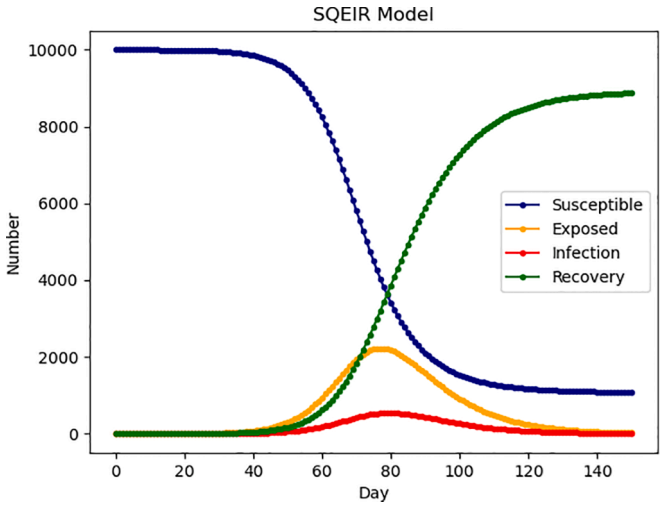


**Fig. 4.** Fitting diagram for the SQEIR model.

matched, resulting in data loss. These inadequacies were addressed by adding editing fields between the data set and the merged data component to modify the field values, adjusting the data values for the region feature columns of the two tables to achieve consistency, and cleaning again the data. Data standardization was thus realized.

We applied the processed datasets to both the SEIR infectious disease model and the proposed SQEIR model. The two models predicted the number of infected people and were compared with the actual scenario, as shown in Fig. 5.

On the basis of Fig. 5, we can conclude that compared with the SEIR model, the SQEIR infectious disease model predicts that the number of infected people is closer to the actual situation, indicating that the proposed SQEIR model can more accurately simulate epidemic spread, compared with the SEIR model.

## 4. Experiments and Discussion

### 4.1. Important parameters in the SQEIR model

(1) Influence of $\beta$.

In the actual spread of infectious diseases, two issues have to be considered: (i) the infection rate ($\beta$) during the actual spread of infectious diseases and (ii) the influence of the medium of infectious diseases on the spread. The infection rate ($\beta$) is an important parameter to characterize the speed of disease spread, and its size directly affects the spread of the epidemic, as shown in Fig. 6.

As shown in Fig. 6, the experimental results for different $\beta$ values vary, and the distribution of the population changes as well. When the value of $\beta$ rises, the infected population also increases, However, the SQEIR model adopts quarantine measures, preventing an extremely large increase in the number of infected people, and prompting an increase in the number of recoveries. These scenarios are consistent with the actual situation and reflects the superiority of China's epidemic prevention policy.

(2) Influence of $\gamma$.

Research has shown that diseases can be eliminated by increasing the cure rate ($\gamma$) control threshold. The cure rate ($\gamma$) mainly depends on the treatment efficiency and the severity of the disease. The number of deaths is also included, as shown in Fig. 7.

On the basis of Fig. 7, we conclude that the size of different cure rates ($\gamma$) directly affects the extent of the spread of the epidemic. For the SQEIR model, which considers numerous factors, even if the cure rate is markedly increased, transmission occurs at a small scale.

### 4.2. Accuracy of the SQEIR model

This study collects real-time outbreak dynamics data from the official website of the National Health Commission of the People's Republic of China. Our database includes the regional population, the corresponding cumulative number of confirmed cases, the cumulative number of deaths, and the cumulative number of cured cases in that region. We selected Huanggang City as the sample space for model training and collected data on the first COVID-19 case from December 31. Subsequently, to verify the validity and accuracy of the SQEIR model, we used the spread of the epidemic in Harbin for model validation. We applied the same data to simulate the spread of the epidemic in Harbin by using the SIR, SEIR, and SQEIR infectious disease models. The experiments evaluated the changes in recovered and infected people over time and then compared the two populations with real data, as shown in Fig. 8.

The experimental results indicate that the model fitting for recovered cases matches the real data in time, and its peak can be accurately inferred. For the prediction of infected individuals, the SQEIR model shows the least deviation from the real data, and the

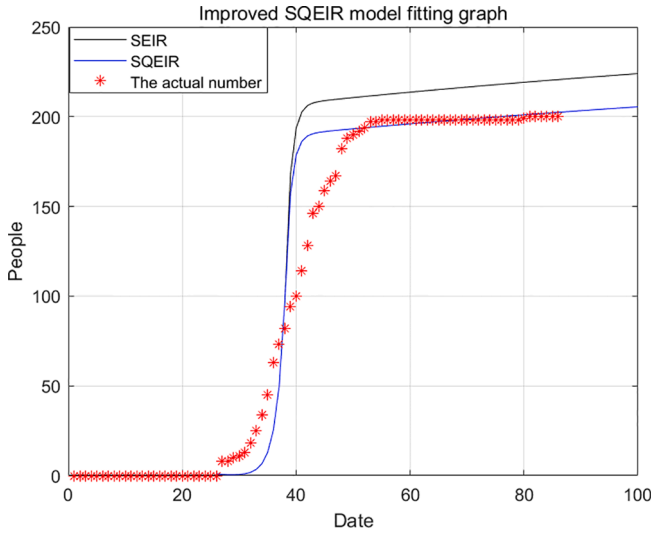


**Fig. 5.** Comparison of data fitting for infected persons between the two models.

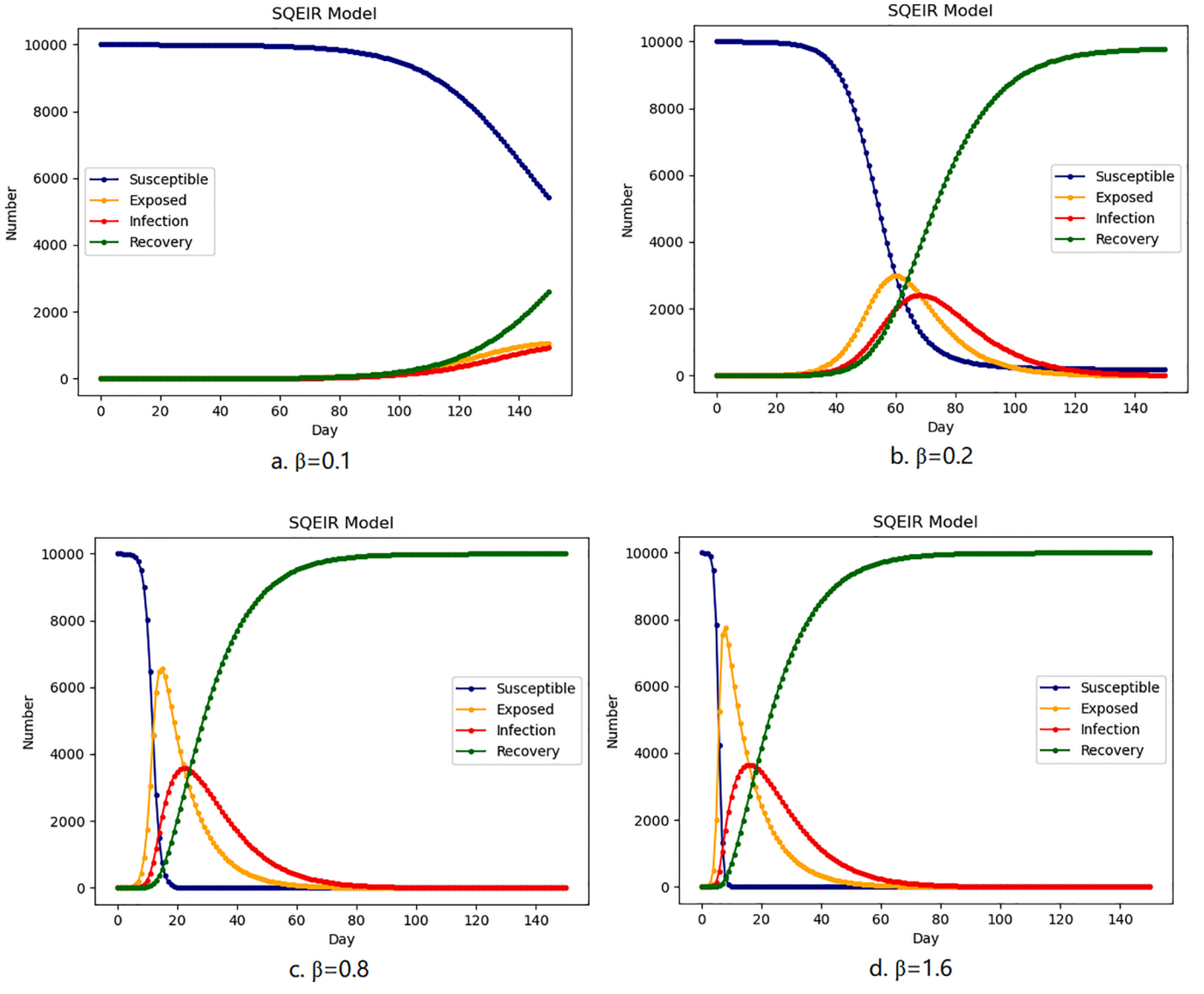


**Fig. 6.** Effect of the value of β on the SQEIR model.

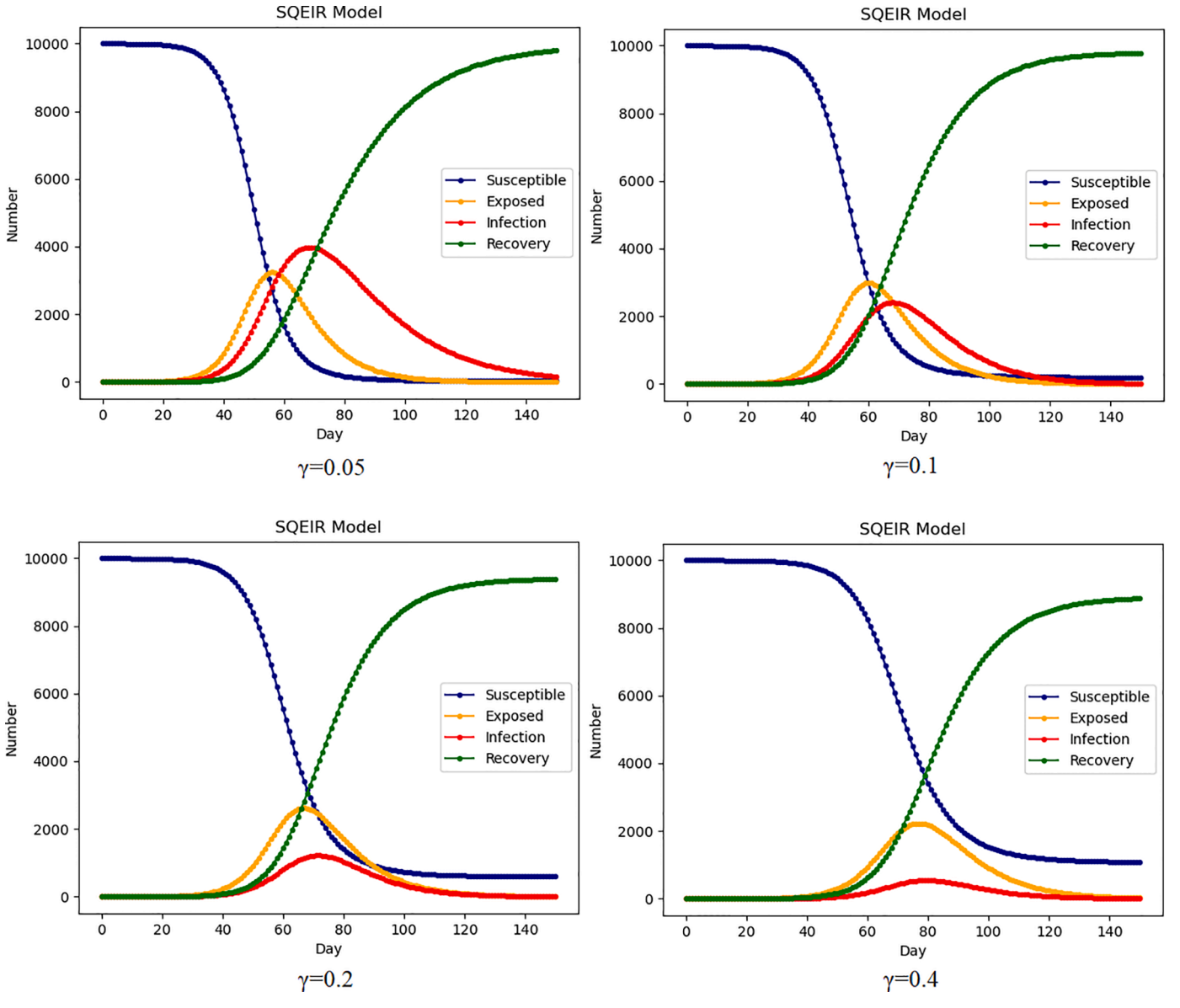


**Fig. 7.** Effect of γ on the SQEIR model.

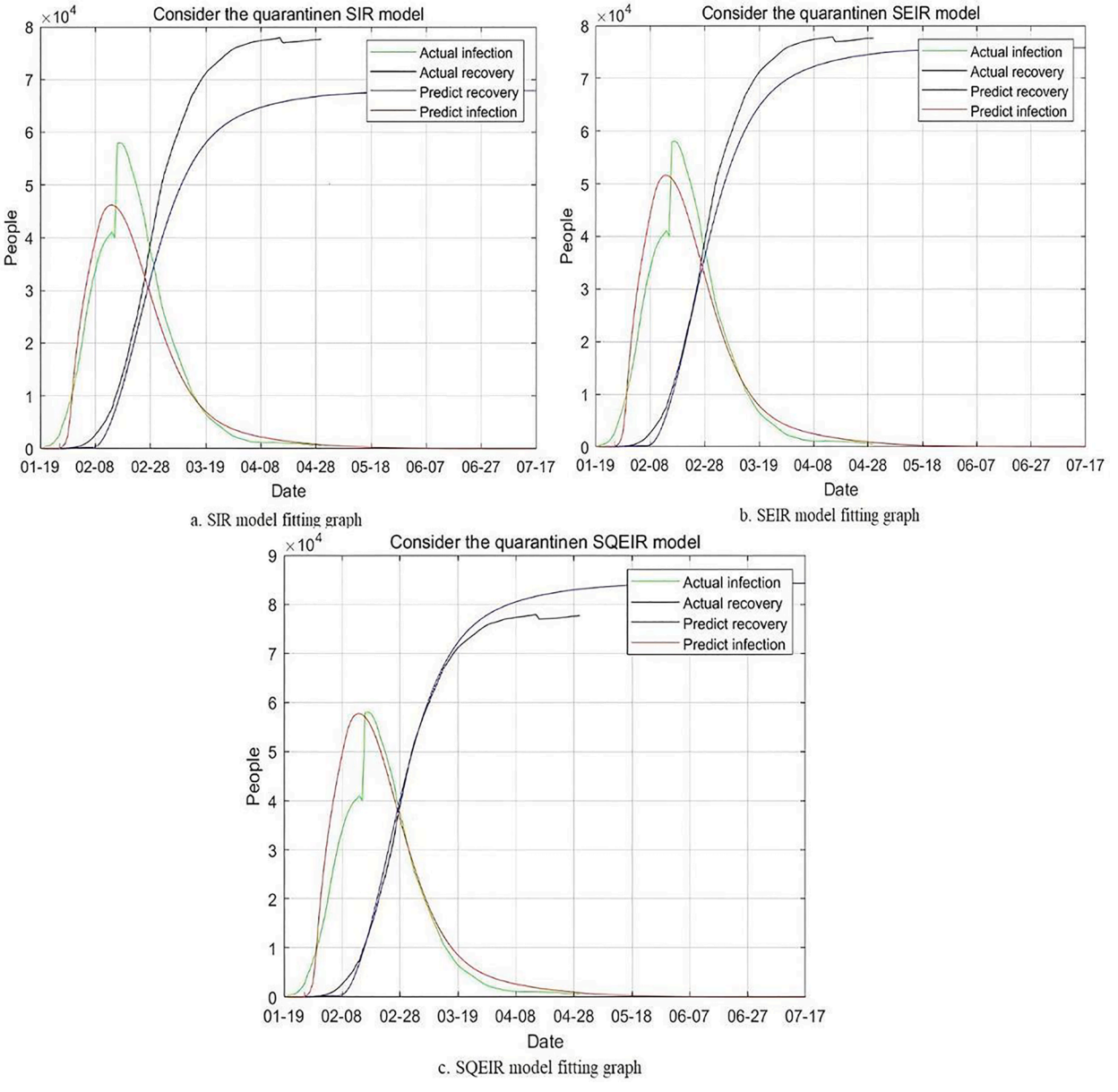


**Fig. 8.** Comparison chart for SIR, SEIR, and SQEIR model fitting.

**Table 1**
Comparison of SIR, SEIR, and SQEIR models

| Model | RMSE | MAE | R2 |
|---|---|---|---|
| SIR | 35.29 | 1.23 | 0.32 |
| SEIR | 14.12 | 0.49 | 0.61 |
| **SQEIR** | **3.53** | **0.12** | **0.84** |

trend of infected individuals is consistent with the real-life scenario. However, despite the numerous factors considered in the SQEIR model, the prediction results vary because of real-time changes in the epidemic. The accuracy of the SQEIR model is higher than all other classical models and 6.7% higher than most classical infectious disease model.

To better visualize the accuracy of the SQEIR infectious disease model, we compared the RMSE, MAE, and R2 values calculated under the SIR, SEIR, and SQEIR infectious disease models. The results are listed in Table 1.

Table 1 shows that the SQEIR model has lower RMSE and MAE values than the other two models and a higher R2 value than the other models. Variations in data indicate that the SQEIR infectious disease model is closer to the real epidemic scenario.

## 5. Conclusion

Infectious disease models have drawn heightened interest because of the COVID-19 outbreak. Mathematical modeling plays an important role in the study of infectious disease dynamics. It reveals the main characteristics of infectious diseases through hypotheses, parameters, variables, and the links between them.

Mathematical modeling using dynamic methods allows the study of whether an infectious disease will spread in an area and continue to be “endemic” in that area, or whether it will eventually be eliminated. Mathematical models are commonly used to discover the transmission mechanism of infectious diseases and predict the epidemic trend of infectious diseases.

In summary, the number of officially confirmed cases and the number of suspected cases from January 19, 2020 to July 17, 2020 are presented in this study. With the infectivity of the novel coronavirus in the incubation period considered, this study proposes a novel infectious disease model referred to as the SQEIR model. The experimental results show that after the introduction of lurkers, the distribution of different populations more closely fits the actual scenario, and the simulation of epidemic spread is more accurate. The results also indicate that prevention and control measures such as isolation and vaccination can affect the accuracy of the model. Compared with other general virus models, the SQEIR model, which considers these factors, has a 6.7% higher accuracy and a smaller error with the true value. However, while our model is as close to the ground truth as possible, some gaps still need to be filled. On the one hand, the virus mutates, and the infection intensity varies because of changes in the genetic material causing the virus; on the other hand, overseas import alters the local epidemic situation. This method and the new SQEIR virus model can contribute to medical initiatives and government epidemic prevention.

## Funding Statement

This work was supported by the National Natural Science Foundation of China (61772179), Hunan Provincial Natural Science Foundation of China (2022JJ50016, 2020JJ4152), the Science and Technology Innovation Program of Hunan Province (2016TP1020), Scientific Research Fund of Hunan Provincial Education Department (21B0649), Double First-Class University Project of Hunan Province (Xiangjiaotong [2018]469), Postgraduate Scientific Research Innovation Project of Hunan Province (CX20190998), Degree and Postgraduate Education Reform Project of Hunan Province (2019JGYB266, 2020JGZD072), Industry University Research Innovation Foundation of Ministry of Education Science and Technology Development Center (2020QT09), Hengyang Technology Innovation Guidance Projects (2020h052805,Hengcaijiaozhi [2020]-67), Postgraduate Teaching Platform Project of Hunan Province (Xiangjiaotong [2019]370-321).

## CRediT authorship contribution statement

**Yichun Wu:** Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Yaqi Sun:** Methodology, Writing – review & editing. **Mugang Lin:** Writing – review & editing.

## Declaration of Competing Interest

We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability

No data was used for the research described in the article.

## References


[1] Bloom D E, Cadarette D. Infectious disease threats in the twenty-first century: strengthening the global response[J]. Front Immunol 2019;10:549.
[2] Lu H, Zhang M, Xu X, Li Y, Shen HT. Deep fuzzy hashing network for efficient image retrieval[J]. IEEE Trans Fuzzy Syst 2020;29(1):166–76.
[3] Ma C, Li X, Li Y, Tian X, Wang Y, Kim H, Serikawa S. Visual information processing for deep-sea visual monitoring system[J]. Cognitive Robotics 2021;1:3–11.
[4] Lu H, Zhang Y, Li Y, Jiang C, Abbas H. User-oriented virtual mobile network resource management for vehicle communications[J]. IEEE Trans Intell Transp Syst 2020;22(6):3521–32.
[5] Lu H, Tang Y, Sun YDRRS-BC. Decentralized routing registration system based on blockchain[J]. IEEE/CAA J Automatica Sinica 2021;8(12):1868–76.
[6] Kang S, Wu H, Yang X, Li Y, Yao J, Chen B, Lu H. Discrete-time predictive sliding mode control for a constrained parallel micropositioning piezostage[J]. In: IEEE Transactions on Systems, Man, and Cybernetics: Systems; 2021.
[7] Wang P, Wang D, Zhang X, Li X, Peng T, Lu H, Tian X. Numerical and experimental study on the maneuverability of an active propeller control based wave glider [J]. Appl Ocean Res 2020;104:102369.
[8] Lu H, Yang R, Deng Z, Zhang Y, Gao G, Lan R. Chinese image captioning via fuzzy attention-based DenseNet-BiLSTM[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 17; 2021. p. 1–18.
[9] Zheng Q, Zhu J, Tang H, Liu X, Li Z, Lu H. Generalized label enhancement with sample correlations[J]. IEEE Trans Knowl Data Eng 2021.
[10] Nakayama Y, Lu H, Li Y, Kamiya T. WideSegNeXt: semantic image segmentation using wide residual network and NeXt dilated unit[J]. IEEE Sensors J 2020;21 (10):11427–34.
[11] Sun J, Li Y. Multi-feature fusion network for road scene semantic segmentation[J]. Comput Electrical Eng 2021;92:107155.
[12] NATIONAL HEALTH COMMITTEE OF THE PEOPLE'S REPUBLIC OF CHINA, http://www.nhc.gov.cn/, last accessed 2020/3/23.
[13] Tang L, Zhou Y, Wang L, Purkayastha S, Zhang L, He J, Wang F, Song Peter X-K. A review of multi-compartment infectious disease models[J]. Int Statistical Rev 2020;88(2):462–513.
[14] Yeling L, Jing W. SIR infectious disease model based on age structure and constant migration rate and its dynamics properties[C]. In: 2020 International Conference on Public Health and Data Science (ICPHDS). IEEE; 2020. p. 158–65.
[15] Peter OJ, Akinduko OB, Oguntolu FA, Ishola CY. Mathematical model for the control of infectious disease[J]. J Appl Sci Environ Manage 2018;22(4):447–51.
[16] Bin S, Sun G, Chen C C. Spread of infectious disease modeling and analysis of different factors on spread of infectious disease based on cellular automata[J]. Int J Environ Res Public Health 2019;16(23):4683.
[17] Scarpino S V, Petri G. On the predictability of infectious disease outbreaks[J]. Nat Commun 2019;10(1):1–8.
[18] Chae S, Kwon S, Lee D. Predicting infectious disease using deep learning and big data[J]. Int J Environ Res Public Health 2018;15(8):1596.
[19] Hunter E, Mac Namee B, Kelleher J D. A model for the spread of infectious diseases in a region[J]. Int J Environ Res Public Health 2020;17(9):3119.
[20] Wong Z S Y, Zhou J, Zhang Q. Artificial intelligence for infectious disease big data analytics[J]. Infection, disease & health 2019;24(1):44–8.
[21] Kypraios T, Minin V N. Introduction to the special section on inference for infectious disease dynamics[J]. Statistical Sci 2018;33(1):1–3.
[22] Minter A, Retkute R. Approximate Bayesian Computation for infectious disease modelling[J]. Epidemics 2019;29:100368.
[23] Chen X, Yu B. First two months of the 2019 Coronavirus Disease (COVID-19) epidemic in China: real-time surveillance and evaluation with a second derivative model[J]. Global health res pol 2020;5(1):1–9.
[24] Kartono A, Wahyudi S T, Setiawan A A, Sofian I. Predicting of the coronavirus disease 2019 (COVID-19) epidemic using estimation of parameters in the logistic growth model[J]. Infect Dis Rep 2021;13(2):465–85.
[25] Büyüktahtakın İ E, des-Bordes E, Kıbış EY. A new epidemics–logistics model: Insights into controlling the Ebola virus disease in West Africa[J]. Eur J Operational Res 2018;265(3):1046–63.
[26] Pastore y Piontti A, Perra N, Rossi L, Samay N, Vespignani A. Infectious disease spreading: From data to models[M]. Charting the next pandemic. Cham: Springer; 2019. p. 3–10.
[27] Heng K, Althaus CL. The approximately universal shapes of epidemic curves in the Susceptible-Exposed-Infectious-Recovered (SEIR) model[J]. Sci Rep 2020;10 (1):1–6.



Yichun Wu graduated from Hengyang Normal University with a bachelor's degree in 2021. She is currently studying at Hengyang Normal University, School of Computer Science and Technology, majoring in Electronics and Communication Engineering for a master's degree. Her current research interests include image style transfer and video image style transfer.

Yaqi Sun received her M.S. degree in Control Engineering from Guilin University of Technology, China in 2013. She is currently an Assistant Professor at the College of Computer Science and Technology, Hengyang Normal University. Her current research interests include machine learning and image processing.

Mugang Lin received his Ph.D. degree from Central South University, Changsha, China in 2019. He is currently an Associate Professor at the College of Computer Science and Technology, Hengyang Normal University, Hengyang, China and at the Hunan Provincial Key Laboratory of Intelligent Information Processing and Application. His current research interests include computer algorithms and parameterized algorithms.