Soft Computing (2023) 27:5437–5501
https://doi.org/10.1007/s00500-022-07778-2

DATA ANALYTICS AND MACHINE LEARNING

Predicting COVID-19 using lioness optimization algorithm and graph convolution network

Dong Li[1] · Xiaofei Ren[1] · Yunze Su[1]

Accepted: 21 December 2022 / Published online: 9 January 2023

Abstract

In this paper, a graph convolution network prediction model based on the lioness optimization algorithm (LsOA-GCN) is proposed to predict the cumulative number of confirmed COVID-19 cases in 17 regions of Hubei Province from March 23 to March 29, 2020, according to the transmission characteristics of COVID-19. On the one hand, Spearman correlation analysis with delay days and LsOA are used to capture the dynamic changes of feature information to obtain the temporal features. On the other hand, the graph convolutional network is used to capture the topological structure of the city network, so as to obtain spatial information and finally realize the prediction task. Then, we evaluate this model through performance evaluation indicators and statistical test methods and compare the results of LsOA-GCN with 10 representative prediction methods in the current epidemic prediction study. The experimental results show that the LsOA-GCN prediction model is significantly better than other prediction methods in all indicators and can successfully capture spatio-temporal information from feature data, thereby achieving accurate prediction of epidemic trends in different regions of Hubei Province.

Keywords COVID-19 · Epidemic prediction · Lioness optimization algorithm · Graph convolutional network

1 Introduction

Corona Virus Disease 2019 (COVID-19), which emerged in late 2019 and broke out in early 2020, is one of the most severe public health emergencies since the founding of the People's Republic of China. As of 18:00 on November 30, 2021, China has reported a total of 127, 859 confirmed cases and a total of 5697 deaths. Currently, there are still 2913 confirmed cases; a total of 262, 505, 360 confirmed cases and a total of 5, 227, 122 deaths have been reported globally, and there are presently 20, 202, 074 confirmed cases. Figures 1 and 2 show the distribution of cumulative confirmed cases in China and globally, respectively.[1]

Due to its long duration and wide spread, infectious diseases not only seriously threaten the life safety of people, but also affect people's mental health for a long time. Improper handling can easily cause social chaos. Especially in recent years, sudden epidemic infectious diseases have continued to appear in the world. For example, SARS in 2002 quickly spread to 32 countries and regions in a short time. As of August 16, 2003, the global death toll reached 919, with a case fatality rate of nearly 11%. In 2009, the H1N1 broke out in the United States on a large scale and spread to 214 countries and regions, causing nearly 200, 000 deaths. In addition, there are Ebola virus, HIV, dengue, and so on.

At present, the number of confirmed cases of COVID-19 is still increasing. How to monitor and predict sudden epidemics in a timely and effective manner, and eliminate the crisis in the bud, plays an essential role in the prevention and control of the epidemic. Given these reasons,

✉ Dong Li
ddli1009@xupt.edu.cn

Xiaofei Ren
171506694@qq.com

Yunze Su
yjysyz@live.com

1 College of Economics and Management, Xi'an University of Posts and Telecommunications, Xi'an 710061, Shaanxi, People's Republic of China

[1] Source of the picture: https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_aladin_banner.

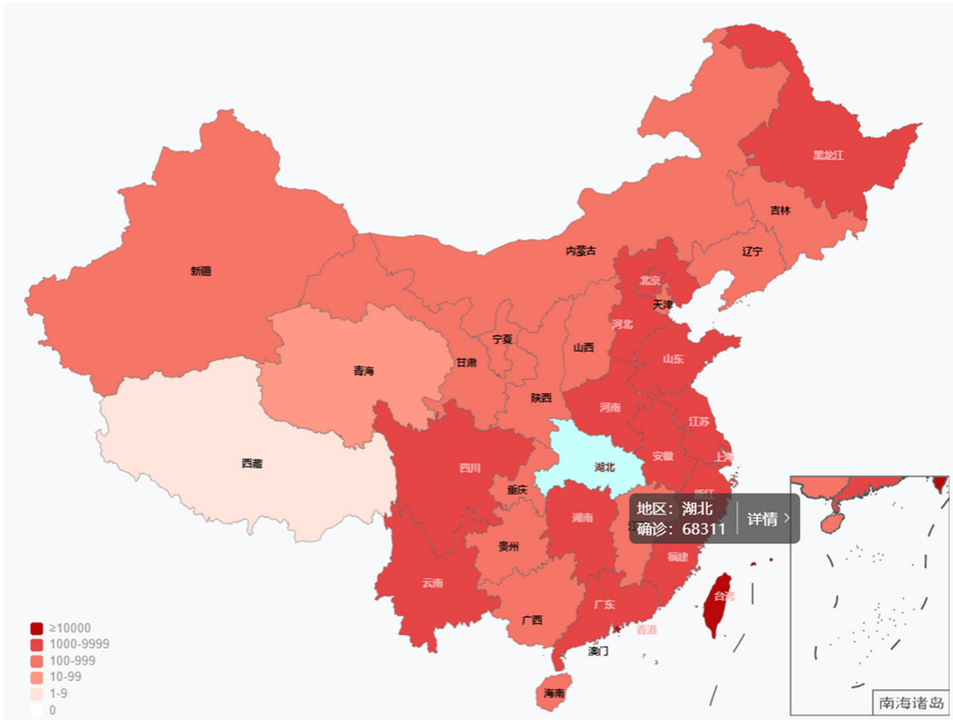


Fig. 1 The cumulative number of confirmed cases in China, including cured and deaths

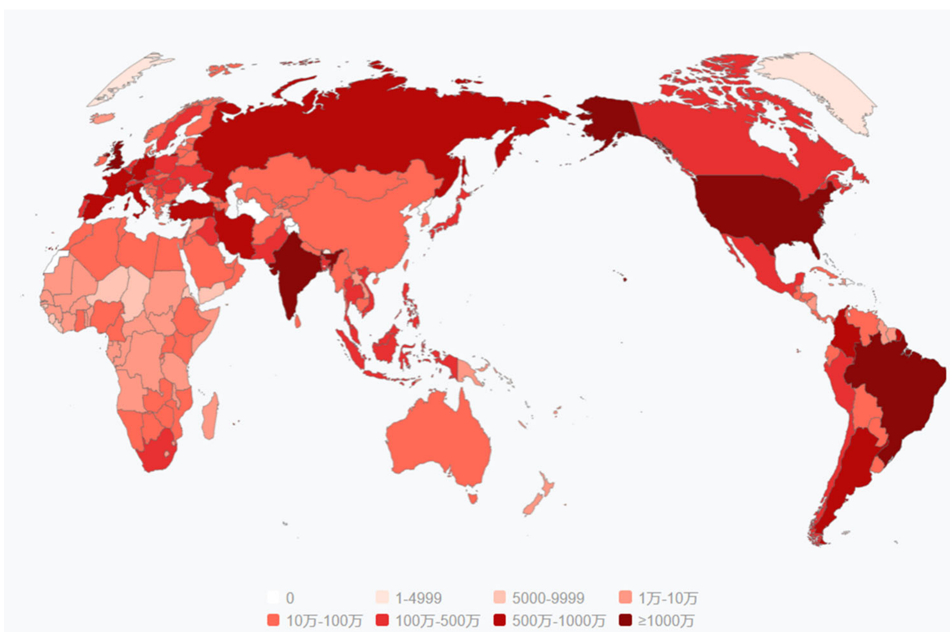


Fig. 2 The cumulative number of confirmed cases globally, including cured and deaths

domestic and foreign scholars have carried out a series of studies on the prediction of infectious diseases. To sum up, these methods can be divided into the following categories:

(1) Time series prediction model. The traditional time series prediction model is an effective tool for disease prediction, which mainly realizes the analysis of the spreading law and development trend of the epidemic by fitting epidemic data. Zhang et al. (2016) decomposed the monthly time series data of 11 class C notifiable diseases in China from 2009 to 2014 into a seasonal index representing the seasonal law of infectious diseases and a linear regression model describing the long-term trend of

infectious diseases. And based on this, an autoregressive integrated moving average (ARIMA) model was established to predict the incidence of these infectious diseases. Wang et al. (2019) selected the monthly incidence data of brucellosis from 2004 to 2013 in Jinzhou city, Liaoning Province, China. They analyzed the epidemiological characteristics and prevalence situation by using the autoregressive integrated moving average (ARIMA) model and constructed a prediction model to predict the incidence of brucellosis from 2005 to 2014. Qi et al. (2020) collected the monthly incidence data of Hemorrhagic Fever with Renal Syndrome (HFRS) in Shandong Province from 2009 to 2018. And they used the seasonal autoregressive fractionally integrated moving average (SARFIMA) model to predict the incidence of HFRS. The model evaluation results showed that the SARFIMA model could better fit the dynamic changes of the monthly incidence of HFRS in Shandong Province, and the excellent degree of fitting and predictive ability. Singh et al. (2020) used an autoregressive integrated moving average (ARIMA) model with covariates to fit daily confirmed COVID-19 cases in Malaysia from 22 January to 31 March 2020. Then, they verified this prediction model with confirmed cases from 1 to 17 April 2020. Finally, the daily confirmed cases in Malaysia from 18 April to 1 May 2020 were predicted with satisfaction. Aslan et al. (2022) developed a SEIQR-type deterministic model, which mainly used a system of ordinary differential equations (ODEs) to analyze the dynamics of the epidemic in Hubei Province, and achieved accurate prediction of the local epidemic situation through social distancing rate and confirmed case rate.

(2) Infectious diseases model. In the field of epidemic transmission research, scholars have put forward many models of infectious diseases that can be used to predict the number of infected people, among which the representative ones are: SIR, SIRS, SEIR models, etc. The main idea of this type of model is to divide the population within the epidemic range of infectious diseases into susceptible, exposed, infectious, recovered, etc., and establish and solve differential equations based on the transformation relationship between different groups. Guo et al. (2016) analyzed the transmission data of the Ebola virus by introducing the Jacobian matrix to solve the differential equation of the SIR model and constructed a comprehensive compartment model to fit and predict the transmission trend of the Ebola epidemic. Kuperman et al. (2001) used the SIRS model to study the transmission rules of infectious diseases on the small world network and found that when network nodes change from a recovered state to a susceptible state, infectious diseases will exist in the network for a long time, but the infection rate is low. Ding et al. (2004) used the SIJR model to analyze SARS epidemic data in Hong Kong, Singapore, and Canada and obtained essential parameters such as the epidemic transmission rate, and then realized the prediction of SARS transmission by adjusting the parameters in this model. Merler et al. (2011) and Wu et al. (2020) both established the SEIR model and included "exposed" into the simulated population to simulate the development trend of H1N1 in Europe and the COVID-19 in major cities of China, respectively. In addition, there are still many scholars who extend the infectious disease model by adding or adjusting the internal structure (Wang et al. 2022a, b; Yi et al. 2021; Liao et al. 2020; Cao et al. 2020; Bi et al. 2021). Liu et al. (2020) proposed a susceptible-asymptomatic-infected-removed (SAIR) model based on the social network to describe the spread of COVID-19 and analyzed the real epidemic data in Wuhan and finally selected an appropriate method to control the spread of COVID-19. In the classical SIR model, the transmission speed is a constant. However, in reality, the transmission speed will change through various interventions such as personal protection, community lockdowns, and urban control. Therefore, Jia et al. (2020) added a transmission modifier to the SIR model, established a dynamic extended SIR model (eSIR) for infectious diseases, which can dynamically change the transmission speed and predict the epidemic trend of COVID-19 in Italy. In view of the fact that the SEIR model does not consider the influence caused by random factors such as population migration and flow between regions, Yang et al. (2020) introduced input/output flows on the basis of the SEIR model, making the modified SEIR (M-SEIR) model is more suitable for the actual population change when it is used to predict the urban epidemic. Wang et al. (2022a, b) introduced the complex network model and the intelligent agent on the basis of M-SEIR model, and proposed a city-level structured prediction and evaluation (CSPE) model for COVID-19 epidemic. The model has effectively deduced the epidemic situation in Hubei Province. For the COVID-19 pandemic, Manuel et al. (2020) also proposed a susceptible-exposed-symptomatic-infectious-asymptomatic infectious-recovered (SEIAR) model considering population and disease mortality based on the SEIR model, which aims to reduce the number of infectors by assessing the isolation situation.

(3) Machine learning prediction method. In recent years, many machine learning algorithms have also begun to be used for the prediction of infectious diseases. Wang et al. (2017) used the annual county-level human brucellosis (HB) cases and data of 37 environmental variables in Inner

Mongolia, China, to build an artificial neural network (ANN) model, which was used to determine essential predictors of human brucellosis and predict the annual county-level cases of human brucellosis. Chen et al. (2018) introduced different predictors for different diseases, countries, and prediction windows, established different LASSO prediction models, and analyzed the importance of different predictors under various conditions by evaluating the prediction performance of these models. Research results showed that different predictors have different effects in different situations, and short-term forecasts usually perform better than long-term forecasts. Hasan (2020) proposed a hybrid model combining ensemble empirical mode decomposition (EEMD) and artificial neural network (ANN). This model first used EEMD to decompose the epidemic time series data and then sent these decomposed IMF components into the neural network model for training. Eventually, the trained neural network was used to predict the trend of COVID-19. Fang et al. (2020) established a random forest model based on meteorological factors to predict the incidence of infectious diarrhea in Jiangsu Province, China. The prediction model better fitted the dynamic characteristics of the disease epidemic. Gumaei et al. (2021) used gradient boosting regression (GBR) to build a training model to predict the total number of daily confirmed COVID-19 cases around the world. Xie et al. (2021) proposed a nonlinear time-varying transmission rate model based on support vector regression in order to overcome the shortcomings of a single model for insufficient extraction of effective information and low prediction accuracy, and applied it to long-term prediction of COVID-19 epidemic in Hubei Province. Smith and Alvarez (2021) took COVID-19 patients in Wuhan, Hubei, China, as the research object and applied a series of machine learning models to analyze the factors affecting the death of COVID-19 patients. They finally obtained the conclusion that the Shapley value is helpful to determine the death factors of COVID-19.

(4) Deep learning prediction method. In recent years, deep learning methods have performed well in fields such as data mining, analysis, and prediction. They have provided new technologies and techniques for studying epidemic transmission. Chae et al. (2018) used search query data, social media big data, and weather data to sort out the factors that affect the occurrence of infectious diseases to build OLS, ARIMA, DNN, and LSTM prediction models and compare their prediction performance. The study found that the prediction model using deep learning is more suitable for predicting the trend of infectious diseases. Wang et al. (2020) used the LSTM network and rolling update mechanism to continuously input new prediction results into the next iteration model for training, and finally made a long-term prediction of epidemic trends in Russia, Peru, and Iran. In this article, the authors also introduced the Diffusion Index (DI) to evaluate the effectiveness of preventive measures such as social isolation and lockdown. Hao et al. (2020) used Elman neural network, LSTM and SVM to predict the cumulative confirmed cases, deaths, and cured cases in Wuhan, Hubei Province, and used SVM with fuzzy granulation to predict the confirmed new cases, new deaths and new cured cases. Finally, they analyzed the applicability of different methods. Rauf et al. (2021) proposed to use long short-term memory (LSTM) network, Recurrent Neural Network (RNN), and Gate Recurrent Unit (GRU), while considering time variables and data nonlinearity, to evaluate each model feature to predict the number of confirmed COVID-19 cases in the Asia–Pacific region in the next 10 days. Chen et al. (2021) developed a data-driven workflow to extract, process, and develop deep learning methods, and built multivariate long and short-term memory (LSTM) models with different architectures to predict the COVID-19 epidemic, thereby supplementing existing modeling methods.

In addition, when predicting the incidence trend of specific infectious diseases, not all features should be included in the prediction model. Sometimes, you can get better prediction results by deleting some unimportant features. Meta-heuristic algorithm is a common feature selection method, which has been widely concerned by scholars (Jia et al. 2019; Turabieh et al. 2021; Samieiyan et al. 2022). Depending on the source of inspiration, meta-heuristic algorithms can be divided into the four categories: (1) biological evolutionary algorithms that simulate the process and mechanism of biological evolution in nature. Among these algorithms, the most widely used is genetic algorithm (GA)(Holland 1992) with the basic operations of selection, crossover and mutation; in addition, there are differential evolution (DE), evolutionary strategy (ES), etc.; (2) algorithms inspired by physical phenomena and laws in nature, such as multiverse optimizer (MVO) inspired by three concepts in cosmology: white hole, black hole and wormhole, simulating the transfer process of matter in the universe from white hole to black hole through the wormhole (Mirjalili et al. 2015); gravitational search algorithm (GSA) based on Newton's law of universal gravitation (Rashedi et al. 2009); lightning search algorithm (LSA) based on the lightning mechanism (Shareef et al. 2015), etc.; (3) algorithms that simulate various human behaviors in society, such as teaching–learning-based optimization (TLBO) that simulates the two stages of "teaching" and "learning" in the learning process to improve the ability of each individual(Rao et al. 2011);

heap-based optimizer (HBO) that uses the heap data structure to map corporate rank hierarchy to simulate three human behaviors(Askari et al. 2020), etc.; (4) swarm intelligence algorithms that simulate the social behavior of animal groups in nature, such as particle swarm optimization (PSO) designed to simulate the predatory behavior of birds(Eberhart and Kennedy 1995); whale optimization algorithm (WOA) proposed by imitating the hunting behavior of humpback whales(Mirjalili and Lewis 2016); chimp optimization algorithm (ChOA) for modeling hunting behavior of chimps according to differences in intelligence and sexual motivation(Khishe and Mosavi 2020); In addition, there are recently proposed marine predators algorithm (MPA)(Faramarzi et al. 2020), k-means clustering optimizer (KO)(Minh et al. 2022), etc. The principles of these algorithms are different, and their abilities to solve different types of problems are also different. As the "no free lunch (NFL)" theorem states, no algorithm can solve all types of optimization problems with equal simplicity and efficiency (Wolpert and Macready 1997). Therefore, we think it is necessary to develop new optimization algorithms to better achieve the purpose of feature selection.

In conclusion, many scholars have made great efforts in infectious disease prediction including COVID-19 prediction, and the results have been fruitful. However, by analyzing the current situation of infectious disease prediction research, we found that the existing models still have the following problems:

(1) Some models are constructed based on the historical data of a single region, ignoring the spatial correlation of infectious disease transmission, which affects the improvement of the model prediction accuracy;
(2) Some prediction models only use directly related data, such as the number of confirmed cases and the number of isolated people, which makes the prediction results unable to timely reflect the influence of other factors (such as social environment and government response policies) in the process of infectious disease transmission, and finally makes the prediction accuracy difficult to be satisfied;
(3) Some models do not consider the time lag effect of eclipse period on transmission. Virus infection can have a certain eclipse period (the time from contact with the source of transmission to the onset of symptoms), and even asymptomatic patients can appear. Ignoring the time lag will also affect the prediction performance;
(4) Finally, when there are many feature data input into the model, there may be some irrelevant features. If these features are forcibly incorporated into the prediction model without selection, it will not only increase the computation, but also affect the prediction performance of the prediction model.

In order to solve the above problems, this paper proposed a graph convolution network prediction model based on the lioness optimization algorithm (LsOA-GCN) according to the transmission characteristics of the COVID-19. The main contributions are as follows:

(1) A novel population-based lioness optimization algorithm was proposed to filter the feature data with time delay (Baidu index), and based on this, a feature matrix for epidemic prediction was constructed;
(2) A COVID-19 prediction model, LsOA-GCN, which comprehensively uses the lioness optimization algorithm and graph convolutional network (GCN) (Kipf and Welling 2017), was proposed;
(3) Experiments were conducted on epidemic data from 17 regions in Hubei Province, China, and compared with 10 prediction methods, which verified the superiority of the LsOA-GCN prediction model.

The rest of this paper is composed as follows. The second part introduces the LSOA-GCN prediction model in detail, including problem definition, adjacency matrix construction, data source and model construction, etc. The third part compares and analyzes the performance of different prediction methods through model evaluation indicators and statistical test methods. The last part summarizes the proposed prediction model.

2 LsOA-GCN model

2.1 Problem definition

(1) City network graph:G. In this paper, $G = (V, E)$ is used to describe the topological structure of Hubei Province's city network. We regard each region in Hubei Province as a node in the network graph, where, $V = \{v_1, v_2, ..., v_n\}$ represents the set of city nodes in Hubei Province, n is the total number of nodes, and E represents the set of edges between cities.

(2) Adjacency matrix: $A^{n\times n}$. The spatial information between cities is mapped into the city network graph G, and the degree of correlation between city nodes is represented by edges. This paper intends to use the migration relationship graph to reflect the degree of mutual influence between various regions in Hubei Province, which is represented by $G = (V, A)$, where $A \in R^{n\times n}$ is the adjacency matrix.

(3) Feature matrix: $X^{n \times t \times d}$ We use the Baidu Search Index value of keywords related to the epidemic as the attribute features of the nodes in the network, which was represented by $X \in R^{n \times t \times d}$, where t is the number of days, and d is the dimension of the node attribute features. For example, $X_t \in R^{n \times i}$ represents the Baidu Search Index value of i keywords in n regions on the t-th day.

Therefore, the epidemic prediction problem solved in this paper can be expressed as: under the premise of a given city network topology graph G and the feature matrix X, we planned to predict the cumulative number of confirmed cases in each region of Hubei province in the future through graph convolution network. The formula is as follows:

$$[Y_{t+1}, \ldots, Y_{t+T}] = GCN(G; (X_{t-s}, \ldots, X_{t-1}, X_t)) \tag{1}$$

where s is the length of the historical time series, T is the length of the future time series that needs to be predicted, Y_{t+1} represents the cumulative number of confirmed cases in each region of Hubei Province on the $(t+1)$-th day, and X_t represents the feature vector of each region on day t.

2.2 Construction of adjacency matrix

Since the outbreak of the COVID-19 has the characteristic of "person-to-person transmission", this paper assumes that the large-scale population movement during the epidemic will have a certain impact on the spread of the epidemic. And COVID-19 has a certain eclipse period. During the eclipse period, some patients have only mild respiratory symptoms, such as fever, fatigue, dry cough, etc., but they are also infectious, and there are even asymptomatic patients. Based on this, this paper collected the population migration index values of 17 regions in Hubei province from January 10, 2020 to March 15, 2020 (a total of 66 days) and performed a time-delayed Spearman correlation analysis with the cumulative confirmed cases (from January 24, 2020 to March 29, 2020) in each region to construct an adjacency matrix.

$$A_{ij} = ave|(spearman(x_i, y_{ji}))| \quad i,j \in [1, 17] \tag{2}$$

where the Spearman correlation coefficient $\rho = \frac{\sum_{i,j}(x_i - \bar{x})(y_{ji} - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_{i,j} (y_{ji} - \bar{y})^2}}$, x_i represents the cumulative number of confirmed cases in region i, and y_{ji} represents the population migration index from region j to region i.

2.3 Data source

The epidemic data in this paper come from the daily epidemic notification provided by the National Health Commission (http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml) and the COVID-19 epidemic situation in Hubei Province provided by the Hubei Health Committee (https://wjw.hubei.gov.cn/) to collect the cumulative number of confirmed COVID-19 cases in 17 regions of Hubei Province from January 24, 2020, to March 29, 2020.

The migration data used to construct the adjacency matrix in this paper come from Baidu Maps Huiyan (http://qianxi.baidu.com/) developed by Baidu, and we collected and processed the migration scale index of 17 regions in Hubei Province from January 10, 2020 to March 15, 2020. The keyword data used to construct the feature matrix is also derived from the Baidu Search Index (https://index.baidu.com/v2/index.html#/) provided by Baidu. Set the location parameter as "cities in Hubei Province", the time parameter as "2020-01-10 $\sim$ 2020-03-15", and the source parameter as "PC+ Mobile".

2.4 Model construction

2.4.1 Overall overview

In this section, we give a comprehensive overview of how to use the LsOA-GCN model for epidemic prediction. The LsOA-GCN model is mainly composed of two parts. One is the construction of the feature matrix for capturing temporal information, the other is the graph convolution operation for capturing spatial information. First, the keyword data was sorted out through Spearman correlation analysis with a time lag, and the initial feature matrix was constructed to capture the time series information between the current number of confirmed cases and the Baidu Search Index at different periods in the past. Then LsOA was used to further filter the feature matrix to improve the accuracy of the model and avoid unnecessary feature interference. Finally, the feature matrix was used as input to obtain the spatial features of the epidemic-related data through the graph convolution network, to get the prediction results.

2.4.2 Capture temporal information

In an information society, with the continuous innovation and development of PC and mobile terminals, the Internet has become the primary and critical way for people to collect information from the outside world. Netizens can use various information search platforms to gather the information they want to know and obtain real-time data on the network. According to the 47th "Statistical Reports on Internet Development in China" released by the China Internet Network Information Center (CNNIC), as of December 2020, there have been 989 million Chinese netizens, an increase of 85.4 million compared with March 2020. The Internet has become an essential channel for the public to obtain information. Among the many search platforms, Baidu is the most popular information search platform, with nearly 90% of users in China, accounting for a significant market share. Among them, Baidu Index is a data sharing platform based on massive netizens' behavior data in Baidu. Through the Baidu Index, we can study keyword search trends, gain insights into the interests and needs of netizens, monitor public opinion trends, and locate audience features. In recent years, Baidu Search Index has been used to predict some epidemic diseases, such as dengue (Li et al. 2017), HIV/AIDS (Huang et al. 2020) and Brucellosis (Zhao et al. 2020a). Therefore, it is scientific and feasible to use Baidu Search Index as a feature to predict the COVID-19.

During the epidemic, there is a complicated time series correlation between the current number of confirmed cases and the Baidu Search Index at different periods in the past. Therefore, obtaining the temporal features is also a key issue in epidemic prediction. The first step to obtain time information is the selection of keywords. The selection methods include empirical word selection, technical word selection (Ginsberg et al. 2009), and range word selection (Alessa and Faezipour 2018). The keyword selection used to construct the feature matrix this time mainly adopted the range word selection method, supplemented by the empirical word selection method. Firstly, the selection range of keywords should be roughly determined according to the features of the epidemic, transmission routes, protective measures, etc. Then, relevant words can be recommended through the demand graph function in Baidu Index, while reducing the workload, as far as possible to ensure the completeness of the epidemic-related vocabulary. Based on these, we set up a keyword database from the four dimensions of prevention, symptoms, common words and other related aspects of COVID-19, and initially selected 40 specific keywords, namely "vaccine", "hand washing", "mask", "disinfection", "nucleic acid", "flu", "diarrhea", "dry cough", "pneumonia", "cough", "fever", "dyspnea", "high temperature", "stuffy nose", "asymptomatic infection", "close contact", "suspected case", "aerosol transmission", "infection", "novel coronavirus", "NCP", "SARS-CoV-2", "COVID-19", "novel coronary pneumonia", "novel coronavirus pneumonia", "virus pneumonia", "coronavirus", "novel", "epidemic", "Center for Disease Control and Prevention", "Red Cross Society", "Spring Festival", "Health Commission", "SARS", "Zhong Nanshan", "children", "medical", "government", "Wuhan" and "lockdown of the city".

The specific construction steps of the feature matrix are as follows:

(1) Establish a keyword database from the 4 dimensions mentioned above, and initially screen out 40 keywords;

(2) Time-delayed Spearman correlation analysis is performed between Baidu Index (search index, information index) of 40 keywords and the cumulative number of COVID-19 cases;

(3) Further screen the keyword database and select keyword data with a correlation coefficient greater than 0.5;

(4) Collect keyword data according to the delay days corresponding to the maximum correlation coefficient of each keyword in the thesaurus, and construct a feature matrix based on this;

(5) When dealing with practical problems, not all features must be included in the prediction model, and sometimes deleting some unimportant features may improve the performance of the model. Feature selection is a common data dimensionality reduction method, which has been widely concerned by scholars. They regard feature selection as a 0–1 problem and usually choose the meta-heuristic algorithm to solve (Guo et al. 2022; Jia et al. 2022; Sha et al. 2020). Inspired by this, we used the swarm intelligence optimization algorithm to further filter the feature matrix. Its main purpose is to delete the feature data that is not related to the problem in the original feature matrix, reduce the feature dimension and reduce the running time. This point will be explained in detail in the next part.

2.4.3 Filter feature matrix

In recent years, various intelligent optimization algorithms have emerged one after another, such as the classic Particle Swarm Optimization (Eberhart and Kennedy 1995), Whale Optimization Algorithm (Mirjalili and Lewis 2016), and the newly developed Marine Predator Algorithm (Faramarzi et al. 2020), etc. However, when dealing with practical problems, these algorithms will converge prematurely and easily fall into the risk of local optimum, making it difficult to solve all practical application problems (Faramarzi et al. 2020).

Therefore, we proposed a population-based meta-heuristic optimization algorithm, Lioness Optimization

Algorithm (LsOA), to further filter the feature matrix to avoid the interference of unnecessary features. The lion-inspired optimization algorithm appeared in 2012: The Lion's Algorithm (2012) (Rajakumar 2012), which mainly simulates the three behaviors of lion mating, territory takeover, and territory defense, to find optimal solution; The second is Lion pride optimizer (2022) (Wang et al. 2012), which is developed around the mating behavior of lions, and on this basis, three evolutionary strategies are proposed. Survival of the fittest is its main idea; The third chapter is the Lion Optimization Algorithm (2016) (Yazdani and Jolai 2016), which not only draws on the mating and competition behaviors of the lions, but also introduces hunting, migration, patrol and other mechanisms. It can be seen that although the evolution process of the above algorithms incorporate the natural behaviors of lions hunting and mating, they fail to consider some of the core optimization mechanisms in the lion hunting process, resulting in their shortcomings such as slow convergence speed and difficulty in finding global extreme value. Aiming at the shortcomings of the above algorithms, this paper proposed LsOA. Because the optimization process of the algorithm is closer to the hunting behavior of the lions, the lioness is the main performer of the lion hunting. Inspired by this, this algorithm mainly focuses on the hunting behavior of lionesses, and comprehensively considers the two mechanisms of team hunting and elite hunting. During the period, Levy flight and Brownian motion are also introduced. While retaining the strong global exploration ability of the original lion pride algorithms, the local mining ability of the algorithm is improved, which can solve various mathematical optimization problems more effectively, which is completely different from the previous algorithms.

In LsOA, the prey is the global optimal solution, and the current optimal solution is considered to be lioness A or the top lioness. The lions mentioned are all candidate solutions, and they are all lionesses by default. LsOA is mainly composed of the following six mechanisms.

(1)Track prey.

In the hunting process, the lions tend to spread out into a fan shape to surround the prey in the center, and the group of lions in the "center" of the fan shape will consciously adjust their positions according to the positions of other lions and prey. The mathematical model of this behavior is shown in Eqs. (3) and (4), where Eq. (3) defines the step length and direction of the current lion group to the target position, and Eq. (4) defines the position of the lion group after the update.

$$\vec{D} = \left| \vec{C} \cdot \overrightarrow{\text{Prey}} - \vec{X}(t) \right| \tag{3}$$

$$\vec{X}(t+1) = \overrightarrow{\text{Prey}} - \vec{A} \cdot \vec{D} \tag{4}$$

where t is the current iteration number, $\vec{X}(t)$ is the position vector of the lioness, $\vec{D}$ is the distance between the current lioness and the prey, represents the product, $\overrightarrow{\text{Prey}}$ is the position vector of the prey, $\vec{A}$ and $\vec{C}$ are coefficient vectors, and their calculation formulas are as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{5}$$

$$\vec{C} = 2\vec{r}_2 \tag{6}$$

where $\vec{a} = 2 - \frac{2 \cdot \text{Iter}}{\text{Max\_iter}} \in [0, 2]$, to simulate approaching the prey, as the iteration progresses, the value of $\vec{a}$ decreases linearly from 2 to 0, and $\vec{A} \in [-\vec{a}, \vec{a}]$. If $|\vec{A}| < 1$, it will attack the prey and focus on exploitation at this time; if $|\vec{A}| > 1$, it will force to stay away from the prey and focus on exploration at this time. $\vec{C} \in [0, 2]$ is a random value, which represents the weight value of the influence of the position of the current optimal solution on the updated position. $\vec{r}_1, \vec{r}_2$ are both random vectors between [0, 1].

(2) Capture prey.

When capturing prey, we believe that the location of the target prey should be near the "central circle" of the hunting team. And this center circle is composed of the four best candidate solutions in the population (lioness A, lioness B, lioness C, and lioness D) and the average of their four. The mathematical model is as follows:

$$\begin{aligned} \vec{D}_A &= \left| \vec{C}_1 \cdot \vec{X}_A - \vec{X} \right|, \vec{D}_B = \left| \vec{C}_2 \cdot \vec{X}_B - \vec{X} \right| \\ \vec{D}_C &= \left| \vec{C}_3 \cdot \vec{X}_C - \vec{X} \right|, \vec{D}_D = \left| \vec{C}_4 \cdot \vec{X}_D - \vec{X} \right| \end{aligned} \tag{7}$$

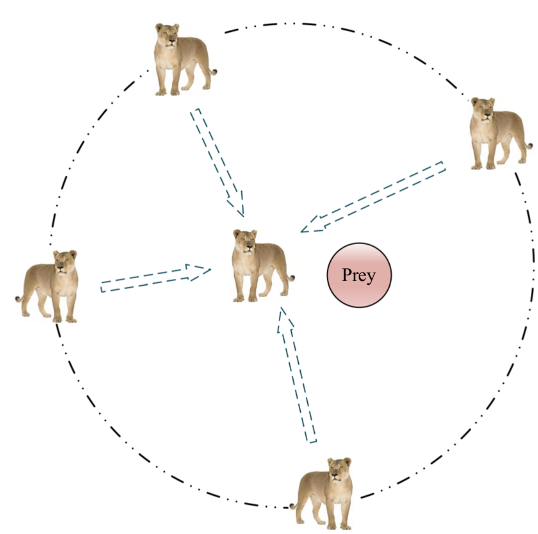


Fig. 3 The construction process of the "central circle"

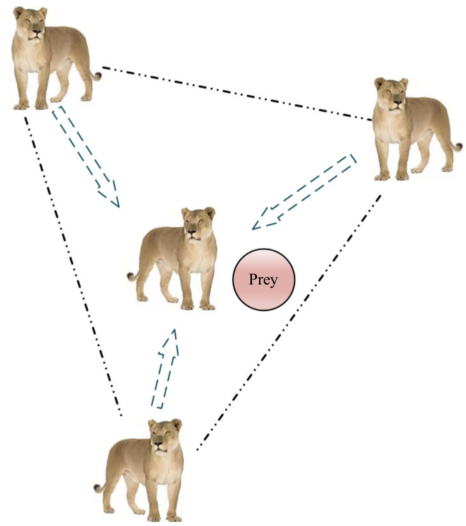


Fig. 4 The construction process of the elite matrix

$$\begin{aligned}\vec{X}_a = \vec{X}_A - \vec{A}_1 \cdot (\vec{D}_A), \vec{X}_b = \vec{X}_B - \vec{A}_2 \cdot (\vec{D}_B) \\ \vec{X}_c = \vec{X}_C - \vec{A}_3 \cdot (\vec{D}_C), \vec{X}_d = \vec{X}_D - \vec{A}_4 \cdot (\vec{D}_D)\end{aligned} \tag{8}$$

$$\vec{X}_{ave} = \frac{\vec{X}_a + \vec{X}_b + \vec{X}_c + \vec{X}_d}{4} \tag{9}$$

where the calculation of A_i, C_i(i=1, 2, 3, 4) are shown in Eq. (5) and Eq. (6), $\vec{X}_A, \vec{X}_B, \vec{X}_C, \vec{X}_D$ are the positions of lioness A, lioness B, lioness C and lioness D, respectively, $\vec{X}_{ave}$ is the average of $\vec{X}_a, \vec{X}_b, \vec{X}_c, \vec{X}_d$, and $\vec{X}$ is the position of the current prey.

(3) "Central Circle" and Elite Matrix.

"Central Circle":

According to Eqs. (8) and (9), the positions of the lions at the center of the hunting team can be determined. It is updated based on the positions of the 4 lionesses closest to the prey in the current population, which are $\vec{X}_a, \vec{X}_b, \vec{X}_c, \vec{X}_d$, respectively, and the last candidate is the average $\vec{X}_{ave}$ of these four candidates. These five positions constitute the hunting team's " Center circle":

$$\mathrm{E\_Team} = \left[\vec{X}_a; \vec{X}_b; \vec{X}_c; \vec{X}_d; \vec{X}_{ave}\right] \tag{10}$$

Figure 3 shows the construction process of the "central circle". Since the position of the target prey in the search space is not known a priori, this article assumes that the position of the target prey may be located near any of these five positions, and the probability of the five candidates being selected is the same, which is 0.2.

Elite Matrix:

The elite matrix is designed based on the elite hunting mechanism. At this time, the population evolves into some agents of the elite, through which agents are used to continuously test the direction and position of the elite, and finally achieve the goal of capturing the prey.

In designing the elite hunting mechanism, this paper did not use a single lioness to carry out optimization iterations, but took the elite hunting mechanism as inspiration, combined with the swarm intelligence mechanism to design: that is, according to the results of the population iteration, the position of the elite is determined on the basis of the lioness of a specific size, and then the optimization iteration is carried out under the guidance of the group search mechanism.

Figure 4 shows the construction process of the elite matrix. In order to avoid the risk of falling into a local optimum caused by relying only on a single elite position, here we draw on the principle of triangular stability, and replace the common method for determining the position of the elite from the common last-round best to a joint decision by the top three agent positions (lioness A, lioness B, and lioness C) in the last-round fitness value. The calculation formula is as follows:

$$\mathrm{Top\_lioness\_pos} = \frac{\vec{X}_A + \vec{X}_B + \vec{X}_C}{3} \tag{11}$$

where $\vec{X}_A, \vec{X}_B, \vec{X}_C$ are the position vectors of lioness A, lioness B and lioness C, respectively. After the position of the top lioness is determined, copy it n times to build the elite matrix:

$$Elite\_\mathrm{lioness} = \begin{bmatrix} X^I_{1,1} & X^I_{1,2} & \cdots & X^I_{1,d} \\ X^I_{2,1} & X^I_{2,2} & \cdots & X^I_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ X^I_{n,1} & X^I_{n,2} & \cdots & X^I_{n,d} \end{bmatrix}_{n\times d} \tag{12}$$

where $\overrightarrow{X^I}$ is the position vector of the top lioness in the lion group.

(4) The march strategy.

With the increasing number of optimization iterations, the risk of the lions falling into local optimization is increasing. In order to enable the lions to quickly explore a new area, this paper designed the march strategy. The implication is that when $r2 \le M$, unifies all dimensions of the predator into a unique value (march). Otherwise, each dimension will randomly select a value in the *E_Team* to be updated (random). The formulas are as follows:

$$\Pr ey(i,j) = E\_Team(randi(size(E\_Team,1)),\\ :), \text{IF}(r2 > M) \tag{13}$$

$$\Pr ey(i,:) = E\_Team(randi(size(E\_Team,1)),\\ :), \text{IF}(r2 \leq M) \tag{14}$$

where *randi*() is used to generate pseudo-random integers; size() is used to return the size of the vector. M indicates the degree of influence of the march strategy on the process. The prey here is a matrix with the same dimension as the elite matrix.

(5) Phase-focused mechanism.

In order to avoid premature convergence of LsOA and in reference to the behavioral process of lion hunting, the whole iteration process was divided into three stages in this paper, namely the early iteration, the middle iteration, and the late iteration (Faramarzi et al. 2020). And different search mechanisms were used in each stage.

① the early iteration.

While $Iter < \frac{1}{3}\, Max\_iter$

$$\overrightarrow{\text{step}}_i = \overrightarrow{\text{RB}} \otimes \left(\overrightarrow{\text{Elite\_lioness}_i} - \overrightarrow{\text{RB}} \otimes \overrightarrow{\text{Prey}_i}\right) i = 1, \ldots n \\ \overrightarrow{\text{Prey}}_i = \overrightarrow{\text{Prey}}_i + P \cdot \vec{R} \otimes \overrightarrow{\text{step}}_i \tag{15}$$

where *Iter* is the current iteration number and *Max_iter* is the maximum iteration number. $\overrightarrow{\text{RB}}$ is a vector of random numbers representing Brownian motion, the symbol $\otimes$ means multiplication term by term. P= 0.5, R is a random vector in [0, 1].

② the middle iteration.

While $\frac{1}{3}\, Max\_iter < Iter < \frac{2}{3}\, Max\_iter.$

For prey:

$$\overrightarrow{\text{step}}_i = \overrightarrow{\text{RL}} \otimes \left(\overrightarrow{\text{Elite\_lioness}_i} - \overrightarrow{\text{RL}} \otimes \overrightarrow{\text{Prey}_i}\right) i = 1, \ldots, n/2 \\ \overrightarrow{\text{Prey}}_i = \overrightarrow{\text{Pre }y}_i + P \cdot \vec{R} \otimes \overrightarrow{\text{step}}_i \tag{16}$$

where $\overrightarrow{\text{RL}}$ is the random number vector representing Levy's flight and $\overrightarrow{\text{RL}} \otimes \overrightarrow{\text{Prey}}_i$ simulates the behavior of prey. For lions, this study assumed:

$$\overrightarrow{\text{step}}_i = \overrightarrow{RB} \otimes \left(\overrightarrow{RB} \otimes \overrightarrow{\text{Elite\_lioness}_i} - \overrightarrow{\text{Prey}_i}\right) i = n/2, .., n \\ \overrightarrow{\text{Prey}}_i = \overrightarrow{\text{Elite\_lioness}_i} + P.CF \otimes \overrightarrow{\text{step}}_i \tag{17}$$

where $\text{CF} = \left(1 - \frac{\text{Iter}}{\text{Max\_Iter}}\right)^{\left(2\frac{\text{Iter}}{\text{Max\_Iter}}\right)}$, it is regarded as an adaptive parameter that controls the step length of the lion's movement. $\overrightarrow{RB} \otimes \overrightarrow{\text{Elite\_lioness}_i}$ simulates the behavior of a predator (lioness).

③ the late iteration.

While $Iter > \frac{2}{3}\, Max\_iter$

$$\overrightarrow{\text{step}}_i = \overrightarrow{RL} \otimes \left(\overrightarrow{RL} \otimes \overrightarrow{\text{Elite\_lioness}_i} - \overrightarrow{\text{Prey}_i}\right) i = 1, \ldots, n \\ \overrightarrow{\text{Prey}}_i = \overrightarrow{\text{Elite\_lioness}_i} + P.CF \otimes \overrightarrow{\text{step}}_i \tag{18}$$

where $\overrightarrow{RL} \otimes \overrightarrow{\text{Elite\_lioness}_i}$ simulates the behavior of a predator (lioness).

(6) The impact of FADs.

FADs (Filmalter et al. 2011) is called "fish aggregating devices", which is a kind of raft-like structure, usually made of some waste products. This kind of device is easy to attract and gather a large number of fish, making it difficult for them to escape. There are similar facilities (regions) on the African savannah, such as wetlands or swamps. When lions hunt here, they often get caught in them and cannot get out.

Given this point, this paper introduced FADs into the algorithm. FADs is considered to be local optima, and the effect is that lions are trapped in these areas. The FADs effect is expressed mathematically as:

$$\overrightarrow{\text{Prey}}_i = \begin{cases} \overrightarrow{\text{Prey}}_i + CF\left[\vec{X}_{\min} + \vec{R} \otimes \left(\vec{X}_{\max} - \vec{X}_{\min}\right)\right] \otimes Ma & \text{if } r \leq FADs \\ \overrightarrow{\text{Prey}}_i + [FADs(1-r) + r]\left(\overrightarrow{\text{Prey}}_{r1} - \overrightarrow{\text{Prey}}_{r2}\right) & \text{if } r > FADs \end{cases} \tag{19}$$

where $FADs = 0.2$, which means the probability that FADs will affect the optimization process.r is a uniform random number in [0, 1]. $r1$, $r2$ represents the random index of the prey matrix. $\vec{X}_{\max}$, $\vec{X}_{\min}$ are the maximum and minimum values of each dimension, respectively. Ma is a binary vector containing only 0 and 1. First, generate a random vector in [0, 1], if it is less than 0.2, it will be regarded as 0; otherwise, it will be regarded as 1. The pseudo-code of LsOA is as follows:

Initialize search agents(Prey) population $i = 1, \ldots, n$
Assign free parameters: $FADs = 0.2$; $P = 0.5$; $Q = 0.5$; $M = 0.9$
While $Iter < Max\_iter$
    Calculate the fitness of each search agent
    $\vec{X}_A$ = the best search agent
    $\vec{X}_B$ = the second best search agent
    $\vec{X}_C$ = the third best search agent
    $\vec{X}_D$ = the fourth best search agent
    Update CF, $\vec{a}$, $\overrightarrow{RL}$ and $\overrightarrow{RB}$
        If ($q \le Q$)   % Team hunting
            For each search agent
                Update $\vec{A}$ and $\vec{C}$ by the Eq. (5) and Eq. (6)
                Use $\vec{A}$ and $\vec{C}$ to calculate $\vec{D}$
                Calculate $\vec{X}_a, \vec{X}_b, \vec{X}_c$ and $\vec{X}_d$ by the Eq. (8)
                $\vec{X}_{ave} = \dfrac{\vec{X}_a + \vec{X}_b + \vec{X}_c + \vec{X}_d}{4}$
                Construct the "center circle": $E\_Team = \{\vec{X}_a, \vec{X}_b, \vec{X}_c, \vec{X}_d, \vec{X}_{ave}\}$
                If ($r_2 < M$)
                    Update the position of the current search agent by the Eq.(13)
                else if ($r_2 \ge M$)
                    Update the position of the current search agent by the Eq.(14)
                end if
            End for
        Else if ($q \le Q$)   % Elite hunting
            Top_lioness_pos= $\dfrac{\vec{X}_A + \vec{X}_B + \vec{X}_C}{3}$
            Construct the Elite matrix and accomplish memory saving
            For each search agent
                If $Iter < Max\_iter/3$
                    Update the position of the current search agent by the Eq.(15)
                else if $Max\_iter/3 < Iter < 2 \cdot Max\_iter/3$
                    For the first half of the populations (i=1,…,n/2)
                        Update the position of the current search agent by the Eq.(16)
                    End for
                    For the other half of the populations (i=n/2,…,n)
                        Update the position of the current search agent by the Eq.(17)
                    End for
                else if $Iter > 2 \cdot Max\_iter/3$
                    Update the position of the current search agent by the Eq.(18)
                end if
            End for
        End if
    Update $Top\_lioness\_pos$ if there is a better solution
    Applying FADs effect and update the position of the current search agent by the Eq.(19)
    $Iter = Iter + 1$
End while
Return $Top\_lioness\_pos$

Note: The value of each dimension is within the range of [0, 1]. If the value is greater than 0.5, it is regarded as 1; otherwise, it is regarded as 0. If the value in a certain dimension is 1, it means that the corresponding feature vector in this dimension is selected, otherwise, it will not be selected. The calculation formula of the fitness value is $fitness = GCN(G; X)$, G is the city network graph, X is the feature matrix, and the output result is the value of Root Mean Squared Error (RMSE).

2.4.4 Capture spatial information

At the beginning of the outbreak, it was the Spring Festival travel season. Most people are taking chances, and their awareness of epidemic prevention is weak. The large-scale population movement has caused the epidemic to spread quickly across the country in a short time. Since the outbreak of the COVID-19 has the feature of "person-to-person transmission", when forecasting the number of confirmed COVID-19 cases in a certain region, it is necessary to consider not only the number of confirmed cases in the region in the past period of time, but also the inflow of confirmed cases from the surrounding regions to the region. How to characterize the interrelationships between complex spaces has become another vital issue in epidemic prediction.

The conventional convolutional neural network (convolutional neural network, CNN) can get local spatial information, but it can only apply to Euclidean space, such as images, regular grids, and so on. The principle is to get

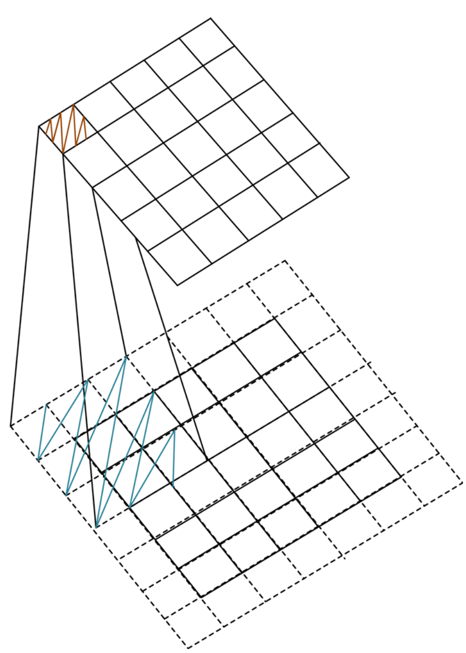


Fig. 5 A schematic diagram of common CNN two-dimensional convolution

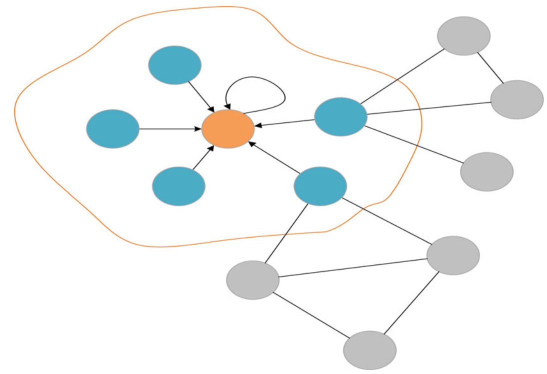


Fig. 6 Capture spatial features by obtaining topological relationships between nodes

the spatial relationship features of the image through the convolution kernel. The typical CNN two-dimensional convolution is shown in Fig. 5. However, the city network is not a two-dimensional grid, but in the form of a graph, which means that CNN cannot reflect the complex topology of the city network and cannot precisely obtain spatial information. In order to be able to extend convolution to topological graphs of non-European data structures, GCN is born. It can extract features from many irregular data structures in real life, typical of which is graph structure, or topological structure.

Give a set of graph data, which contains N nodes, each node has its own features, we used the nodes with D-dimensional features to construct a feature matrix $X \in R^{N \times D}$, and then used the mutual relationship between the nodes to construct the adjacency matrix $A \in R^{N \times N}$. Finally, input the feature matrix X and the adjacency matrix A into the network model. GCN is also a neural network layer, and its layer-to-layer propagation mode is as follows:

$$X^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X^{(l)}W^{(l)}\right) \tag{20}$$

where $X^{(l)}$ represents the feature matrix of the l-th layer, $W^{(l)}$ represents the learnable parameters of the l-th layer. $\tilde{A} = A + I_N$, A is the adjacency matrix, which is used to reflect the network topology in the graph data. I_N is the identity matrix, which is used to extract the features of its own node. $\tilde{D} = \sum \tilde{A}_{ij}$ is the degree matrix, and the degree matrix is used to measure the importance of neighbor nodes. σ is a nonlinear activation function.

It can be seen from the above formula that the GCN model is equivalent to a feature extractor, which acts on the nodes of the graph, and gets together its first-order neighbors according to different weights, thereby capturing the spatial features between nodes. The specific operation is shown in Fig. 6. Suppose the orange node is the central node, and the blue nodes are the first-order neighbor node of the central node. According to the network topology, the central node aggregates the features of itself and its first-order neighbor nodes, and finally obtains spatial information from the feature data.

The LsOA-GCN model used in this paper constructed a GCN model by superimposing two convolutional layers. A two-layer GCN model can be expressed as:

$$X^{(l+2)} = Relu\left(\hat{A}Relu\left(\hat{A}X^{(l)}W^{(l)}\right)W^{(l+1)}\right) \quad (21)$$

where $X^{(l)}$ represents the feature matrix of the l-th layer; Relu() is the activation function; $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$, $\tilde{A} = A + I_N$, A is the adjacency matrix, and $\tilde{D} = \sum \tilde{A}_{ij}$ is the degree matrix; $W^{(l)}W^{(l+1)}$ represent the learnable parameters of layer l and layer $l+1$, respectively.

2.4.5 Loss function

The loss function is the most basic and critical element of the prediction model in the training process. Its purpose is to minimize the error between the actual value and the predicted value of cumulative confirmed cases. The loss function we defined in the LsOA-GCN model was the mean square loss function. The specific formula is as follows:

$$loss = \frac{\sum_{t=1}^{T}(Y_t - \hat{Y}_t)^2}{T} \quad (22)$$

where Y_t and $\hat{Y}_t$, respectively, represent the actual and predicted the cumulative number of confirmed cases on day t.

2.4.6 LsOA-GCN model

In order to capture the temporal and spatial information of the feature data simultaneously, we used a graph convolutional network prediction model based on the lioness optimization algorithm (LsOA-GCN). The overall process of LsOA-GCN is shown in Fig. 7. The right part is the flow chart of this prediction model, and the left part is the feature screening process.

In short, the LSOA-GCN model can handle complex temporal dependence and spatial correlation. On the one hand, Spearman correlation analysis with delay days and LsOA are used to capture the dynamic changes of feature information to get the temporal features. On the other hand, we obtain the structure of the city network by graph convolutional network, to get spatial information and finally realize the prediction task.

3 Experiments

3.1 Data set

To verify the performance of the LsOA-GCN model, we predicted the number of confirmed cases of the epidemic for 66 days in 17 regions of Hubei Province. Due to proper epidemic control measures, the number of newly confirmed cases in Hubei Province was almost zero in the later period. Therefore, the prediction target of the confirmed epidemic data set in Hubei Province was set as the cumulative number of daily confirmed cases.

In the experiment, we selected 17 regions in Hubei Province as the primary research areas, namely Wuhan, Ezhou, Enshi, Huanggang, Huangshi, Jingmen, Jingzhou, Qianjiang, Shennongjia, Shiyan, Suizhou, Tianmen, Xiantao, Xianning, Xiangyang, Xiaogan and Yichang. The experimental data was mainly composed of three parts: The first was the migration data from January 10, 2020 to March 15, 2020. The 17*17 adjacency matrix was constructed mainly based on the correlation between the migration scale index and the cumulative number of confirmed cases; The second was the keyword data from January 10, 2020 to March 15, 2020 (the date range varies according to the number of delayed days), which was used to construct a 17*66*44 feature matrix; The third was the epidemic data from January 24, 2020, to March 29, 2020. The daily cumulative number of confirmed cases in each region of Hubei Province was used as the final target value to form a 17*66 target value matrix. During the period, all experimental data used were normalized, and according to the ratio of 9:1, the data set was divided into the training set and the test set. The length of the historical time series was set to 1, and finally the cumulative number of confirmed cases in the next day was predicted. The specific division is shown in Fig. 8.

3.2 Evaluation indicator

At present, there are many indicators used to evaluate regression prediction models (Zhao et al. 2020b). RMSE, mean absolute error (MAE), mean absolute percentage error (MAPE), accuracy (Acc), coefficient of determination (R2), and explained variance score (Var) were selected as evaluation indicators in this paper to evaluate the difference between the real value Y_t and the predicted value $\hat{Y}_t$ of the cumulative number of confirmed cases. RMSE, MAE, and MAPE are mainly used to measure the difference between the predicted value and the real value. The smaller the three indicators are, the smaller the difference between the predicted value and the actual value is, and the better the performance of the prediction model is. Accuracy is

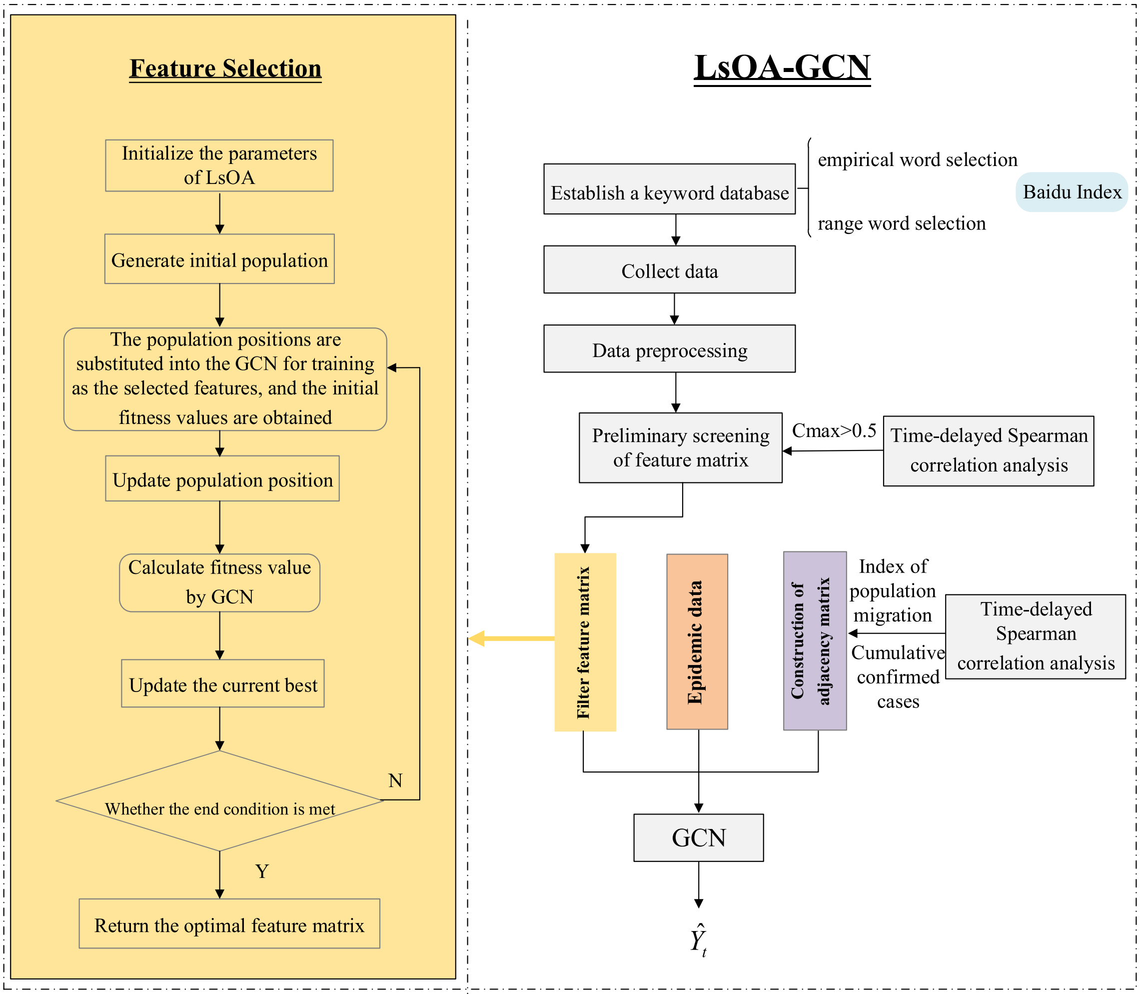
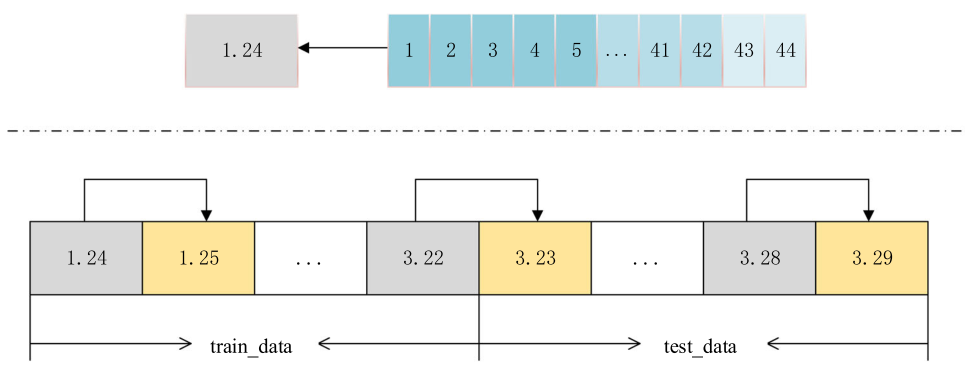


Fig. 7 LsOA-GCN model structure

Fig. 8 Division of data set



mainly used to test the prediction accuracy of the model. R2 and Var are mainly used to measure the fitting degree between the predicted value and the actual value. The larger the three indicators are, the better the prediction effect is. The specific calculation formulas of these six indicators are as follows:

(1) RMSE:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \hat{Y}_t\right)^2} \tag{23}$$

(2) MAE:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|Y_t - \hat{Y}_t\right| \tag{24}$$

(3) MAPE:

$$MAPE = \frac{100\%}{T} \times \sum_{t=1}^{T}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \tag{25}$$

(4) Acc:

$$Accuracy = 1 - \frac{\| Y - \hat{Y} \|_F}{\| Y \|_F} \tag{26}$$

(5) R2:

$$R^2 = 1 - \frac{\sum_{t=1}\left(Y_t - \hat{Y}_t\right)^2}{\sum_{t=1}\left(Y_t - \overline{Y}\right)^2} \tag{27}$$

(6) Var:

$$\mathrm{var} = 1 - \frac{Var\{Y - \hat{Y}\}}{Var\{Y\}} \tag{28}$$

where T is the predicted days; $\| \cdot \|_F$ is the Frobenius norm that refers to as F-norm, which is a matrix norm that can be used to compare the error between the real matrix and the estimated matrix; $Var\{\cdot\}$ is the variance calculation.

3.3 Correlation analysis between Baidu Index and the cumulative number of confirmed cases of COVID-19

Since Baidu Index includes search index and information index, 40 keywords constitute 80 feature vectors. The search index shows the degree of Internet users' attention to keyword search and the continuous changes. It is based on the data of Internet users' search volume on Baidu, with keywords as statistical objects, scientifically analyzes and calculates the weight of the search frequency of each keyword in Baidu. The information index shows the degree of attention and coverage of specific keywords and continuous changes in news information on the Internet. It is based on Baidu's intelligent distribution and recommended content data and is obtained by the weighted summation of the number of netizens' reading, commenting, forwarding, like, dislike and other behaviors. We first preprocessed the data, deleted the search index or information index of keywords with all zeros, and finally contained only 61 feature vectors.

Figure 9 shows the trend comparison between the five randomly selected feature vectors and the cumulative number of confirmed cases, from January 10, 2020, to March 15, 2020. The five feature vectors are coronavirus (S), dry cough (S), fever (S), diarrhea (X), and hand washing (S), where (S) represents the search index of the keyword and (X) describes the information index of the keyword. Since the ordinate used a logarithmic scale with a base of 10 and negative or zero values could not be drawn correctly on the logarithmic graph, keywords with the Baidu Index value of 0 would appear truncated and had no value in the graph.

As can be seen from Fig. 9, the change of Baidu Index values of individual keywords is earlier than the change of cumulative confirmed cases of COVID-19. This showed that, based on certain keywords, we could predict the epidemic trend in the future to a certain extent, and there was an apparent lag between the Baidu Index of keywords and the cumulative number of confirmed cases. Because of this, we assumed that the Baidu Index could be used as a predictive indicator in a period of time before the

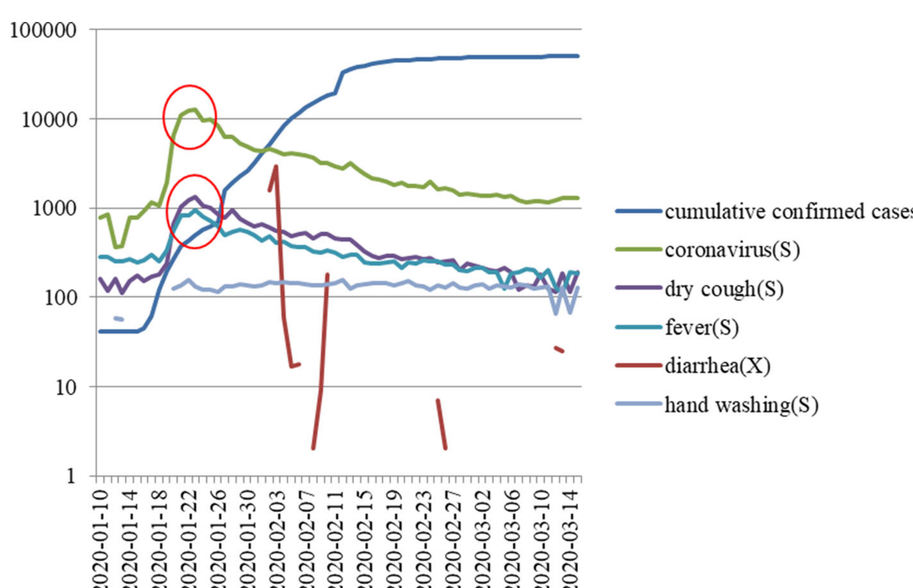


Fig. 9 Trend comparison between the five randomly selected feature vectors and cumulative number of confirmed cases

diagnosis, that is, the feature vector, and then the number of confirmed cases could be predicted.

Based on the above analysis, in this part, we mainly analyzed the correlation between the daily cumulative number of confirmed cases of the epidemic and the Baidu Index. Due to the eclipse period of COVID-19, there are even asymptomatic infections. And as the main feature vector of the prediction model, the Baidu Index needs to be at least a certain number of days in advance to have practical application value. Therefore, the Baidu Index values of keywords 0 to 14 days in advance (epidemic data was relatively lagging by 0 to 14 days) were successively used for correlation analysis with epidemic data.

Taking Wuhan as an example, we performed Spearman correlation analysis on the Baidu Search Index (search index, information index) of keywords with time delay (0-14d) and the cumulative number of confirmed cases in Wuhan. The results are shown in Table 1, and the correlation analysis results of other 16 regions are shown in Appendix Tables 13–28.

Table 1 shows the correlation coefficient value of each keyword with a delay of 0–14 days and the cumulative number of confirmed cases, where the correlation coefficient value corresponding to the optimal delay days has been shown in bold. For example, for novel coronary pneumonia (S), we found that the 14-day delay has the strongest explanatory power. For fever (S), the most explanatory delay time is 3 days. In this experiment, we believed that it is most appropriate to use the keyword data corresponding to the delay days with the strongest explanatory power, and used it for subsequent prediction tasks. In order to highlight the role of feature vectors and improve prediction performance, based on comprehensive consideration of 17 regions in Hubei Province, we only screened out the keyword data with the maximum correlation coefficient greater than 0.5, a total of 44, as shown in Table 2. Then, according to the delay days corresponding to the maximum correlation coefficient, the keyword data was moved, filled, and sorted out to obtain the final feature vectors.

After screening out 44 feature vectors, in order to further determine the importance of correlation analysis with delay days for epidemic prediction, we, respectively, predicted the sorted feature data and unsorted feature data on the GCN model, and the prediction results are shown in Tables 3 and 4, where Table 3is a comparison of the overall prediction results of Hubei Province, and Table 4 is a comparison of the prediction results of 17 regions in Hubei Province. It was worth noting that due to proper epidemic control measures, the number of newly confirmed cases in various regions of Hubei Province in the later period was almost zero, and the daily cumulative number of confirmed cases was almost unchanged. Therefore, according to the calculation formulas of evaluation indicators, when comparing the prediction results of various regions, we only used the four indicators of RMSE, MAE, MAPE, and Acc. It can be clearly seen from Tables 3 and 4 that the sorted feature data are better than the unsorted feature data in predicting effect, and it is better than the unsorted feature data in all indicators. This result proves that the use of correlation analysis with delay days to sort the feature data can improve the performance of the prediction model.

3.4 Performance analysis of lioness optimization algorithm on benchmark functions

After determining the optimal delay days for each feature, we then considered using lioness optimization algorithm to solve the feature selection problem, whose main task is to delete redundant and irrelevant features from the original feature data. In chapter 2.4.3, we introduced the lioness optimization algorithm. In this part, we mainly tested the performance of the proposed algorithm.

In this experiment, we tested the algorithm through 23 commonly used benchmark functions. The benchmark functions mainly include four categories: (1) unimodal functions (F1-F7), which have the unique optimal solution and are mainly used to test the exploitation performance of the algorithm. (2) Multimodal functions (F8-F13), with multiple local optimal solutions, are mainly used to test the exploration ability of the algorithm. (3) Multimodal functions with fixed dimension (F14-F23), which has low and fixed dimensions. The optimization algorithms involved in the comparison included classical particle swarm optimization (PSO), grey wolf optimizer(GWO) (Mirjalili et al. 2014), whale optimization algorithm(WOA), teaching–learning-based optimization(TLBO), and newly developed heap-based optimizer(HBO), marine predators algorithm(MPA) and K-means clustering algorithm(KO). The parameter settings of various algorithms and more details can be found in the corresponding references. To provide reliable comparison results, for all test functions, the algorithms independently ran 30 times in the same environment, and the number of function evaluations was 15, 000. Therefore, under the same other conditions, the differences between algorithms are caused by different performance. Table 5 shows the test results of different algorithms on common benchmark functions.

It can be seen from Table 5 that for unimodal functions (F1-F7), the average values obtained by LsOA on F5 and F6 were far superior to other algorithms. On F7, LsOA also ranked first with a slight advantage. On F1-F4, compared with KO, LsOA performed poorly, but it was also superior to other algorithms by absolute advantages, showing strong competitiveness. For multimodal functions (F8-F23), LsOA reached global optimization on F9, F11, and F15-

Table 1 Correlation analysis between keywords and the cumulative number of confirmed cases–Wuhan

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.1247	0.1787	0.2346	0.2929	0.3559	0.4239	0.4956	0.5729	0.6464	0.7330	0.7933	**0.7954**	0.7911	0.7824	0.7743
Red Cross Society(S)	0.0283	0.0172	0.0596	0.1123	0.1690	0.2237	0.2814	0.3437	0.4134	0.4897	0.5693	0.6542	0.7477	0.8416	**0.8720**
Spring Festival(S)	0.8880	0.8911	0.8946	0.8986	**0.9009**	0.8994	0.8926	0.8860	0.8812	0.8788	0.8749	0.8701	0.8647	0.8581	0.8502
Health Commission(S)	0.0678	0.1150	0.1691	0.2274	0.2875	0.3560	0.4242	0.4935	0.5700	0.6539	0.7429	0.7821	**0.7844**	0.7733	0.7651
Vaccine(S)	0.1460	0.1085	0.0679	0.0239	0.0231	0.0749	0.1362	0.1963	0.2563	0.3217	0.3915	0.4672	0.4932	0.5363	**0.5418**
SARS(S)	0.4405	0.5084	0.5448	0.6189	0.6970	0.7842	0.8693	0.9261	0.9524	0.9817	0.9895	**0.9902**	0.9899	0.9894	0.9887
Hand washing(S)	0.2298	0.2060	0.1690	0.1284	0.0911	0.0426	0.0071	0.0598	0.1158	0.1774	0.2407	0.2931	**0.3013**	0.2609	0.2885
Children(S)	**0.6136**	0.6043	0.6105	0.6064	0.6061	0.6011	0.6014	0.5873	0.5707	0.5475	0.5355	0.5267	0.5175	0.6036	0.5960
Mask(S)	0.1552	0.2114	0.2701	0.3329	0.4003	0.4692	0.5413	0.6218	0.7057	0.7974	**0.8964**	0.8913	0.8851	0.8786	0.8714
Coronavirus(S)	0.2614	0.3194	0.3820	0.4483	0.5207	0.5978	0.6791	0.7663	0.8588	0.9574	0.9889	**0.9898**	0.9893	0.9888	0.9881
Novel(S)	0.1858	0.2432	0.3022	0.3663	0.4346	0.5055	0.5812	0.6635	0.7504	0.8434	0.9445	0.9667	**0.9670**	0.9655	0.9635
Disinfection(S)	0.0703	0.1208	0.1701	0.2234	0.2868	0.3465	0.4113	0.4833	0.5615	0.6373	0.7211	**0.7317**	0.7175	0.7026	0.6851
Lockdown of the city(S)	0.0240	0.0235	0.0742	0.1232	0.1733	0.2286	0.2912	0.3597	0.4327	0.5085	0.5891	0.6770	0.7717	**0.8739**	0.8665
Wuhan(S)	0.2114	0.2670	0.3326	0.3995	0.4659	0.5390	0.6134	0.6947	0.7869	0.8841	0.9433	**0.9455**	0.9434	0.9418	0.9384
Zhong Nanshan(S)	0.1781	0.2334	0.2904	0.3521	0.4205	0.4932	0.5689	0.6489	0.7341	0.8259	**0.9263**	0.9233	0.9192	0.9145	0.9095
epidemic(S)	**0.9191**	0.9160	0.9090	0.9039	0.8984	0.8924	0.8883	0.8845	0.8779	0.8714	0.8651	0.8571	0.8487	0.8399	0.8305
Flu(S)	0.4343	0.4723	0.5102	0.5521	0.5953	0.6490	0.7170	0.7802	0.8435	0.9111	0.9678	**0.9683**	0.9678	0.9665	0.9646
Diarrhea(S)	0.0656	0.1116	0.1652	0.2184	0.2790	0.3432	0.4081	0.4791	0.5538	0.6371	0.7232	0.8191	0.9184	0.9448	**0.9464**
Dry cough(S)	0.3505	0.3969	0.4594	0.5134	0.5879	0.6501	0.7144	0.7849	0.8594	0.9367	0.9860	**0.9893**	0.9888	0.9883	0.9876
Ncp(s)	**0.5801**	0.5725	0.5644	0.5561	0.5357	0.5233	0.5042	0.4817	0.4578	0.4324	0.4036	0.3730	0.3455	0.3103	0.2721
novel Coronary pneumonia(S)	**0.9480**	0.9439	0.9383	0.9373	0.9371	0.9346	0.9326	0.9303	0.9247	0.9206	0.9170	0.9114	0.9062	0.9009	0.8951
COVID-19(S)	**0.7171**	0.7119	0.7045	0.6986	0.6863	0.6765	0.6684	0.6559	0.6447	0.6330	0.6178	0.6023	0.5860	0.5676	0.5474
Novel coronavirus pneumonia(S)	0.0032	0.0408	0.0900	0.1460	0.2025	0.2663	0.3286	0.3944	0.4657	0.5457	0.6286	**0.7207**	0.7192	0.7165	0.7111
Asymptomatic infection(S)	**0.5199**	0.5022	0.4800	0.4601	0.4378	0.4142	0.3834	0.3512	0.3215	0.2901	0.2560	0.2189	0.1770	0.1314	0.0820
Aerosol transmission(S)	**0.4709**	0.4491	0.4250	0.4014	0.3768	0.3486	0.3181	0.2867	0.2514	0.2147	0.1752	0.1320	0.0849	0.0345	0.0202
Pneumonia(S)	0.1877	0.2412	0.3021	0.3647	0.4326	0.5065	0.5831	0.6624	0.7497	0.8428	0.9425	**0.9525**	0.9524	0.9501	0.9472
Cough(S)	0.4075	0.4556	0.4821	0.5211	0.5778	0.6139	0.6796	0.7160	0.7874	0.8513	0.8985	**0.9050**	0.9027	0.8981	0.8926
Fever(S)	0.7782	0.7958	0.8099	0.8295	0.8477	0.8644	0.8989	0.9202	0.9354	0.9597	0.9717	**0.9718**	0.9705	0.9689	0.9670
dyspnea(S)	0.2216	0.2765	0.3356	0.4018	0.4715	0.5451	0.6238	0.7083	0.7809	0.8764	0.9698	**0.9724**	0.9715	0.9704	0.9686
high temperature(S)	0.3187	0.3761	0.4367	0.5013	0.5578	0.6306	0.7048	0.7879	0.8781	0.9480	**0.9695**	0.9691	0.9674	0.9656	0.9635
novel coronavirus(S)	0.1182	0.1676	0.2229	0.2823	0.3448	0.4140	0.4871	0.5632	0.6446	0.7322	0.8260	0.9278	0.9549	**0.9650**	0.9636
SARS-CoV-2(S)	**0.7241**	0.7163	0.7067	0.7000	0.6957	0.6893	0.6792	0.6683	0.6540	0.6413	0.6280	0.6137	0.5978	0.5805	0.5616
Spring Festival(X)	0.7765	0.7760	0.7778	0.7833	**0.7858**	0.7786	0.7703	0.7597	0.7528	0.7427	0.7364	0.7262	0.7105	0.6941	0.6762
Red Cross Society(X)	0.1147	0.0777	0.0379	0.0030	0.0468	0.0968	0.1471	0.2017	0.2591	0.3080	0.3584	0.4287	0.5046	**0.5567**	0.5470

Table 1 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
government(X)	0.1745	0.1704	0.1351	0.0981	0.0884	0.0974	0.1565	0.2103	0.2334	0.2439	0.2567	0.2494	0.2648	**0.3005**	0.2989
Health Commission(X)	0.1732	0.2229	0.2187	0.2570	0.3143	0.3699	0.3919	0.4613	0.5331	0.6098	**0.6642**	0.6540	0.6587	0.6479	0.6289
medical(X)	0.1602	0.1231	0.0859	0.0501	0.0135	0.0306	0.0792	0.1337	0.1894	0.2473	0.3086	0.3637	0.4229	**0.4341**	0.4220
vaccine(X)	0.1378	0.1077	0.0765	0.0441	0.0028	0.0442	0.0953	0.1484	0.2012	0.2579	0.3200	0.3437	0.3684	**0.3790**	0.3529
SARS(X)	0.3243	0.3812	0.4357	0.4917	0.5574	0.5790	0.5926	0.6194	0.6805	0.7358	0.7976	**0.8013**	0.7932	0.7820	0.7694
children(X)	0.3235	0.3048	0.2971	0.3468	0.4064	0.4691	0.4541	0.4354	0.4128	0.4341	**0.5031**	0.4853	0.4593	0.4306	0.4025
mask(X)	0.1042	0.0666	0.0301	0.0150	0.0609	0.1053	0.1575	0.2168	0.2742	0.3381	0.4061	0.4742	**0.4861**	0.4657	0.4360
infection(X)	0.3778	0.4038	0.4311	0.4598	0.4899	0.5216	0.5550	0.5901	0.6270	0.6659	0.7071	0.7506	**0.7966**	0.7835	0.7690
coronavirus(X)	**0.5493**	0.5341	0.5183	0.5029	0.4854	0.4656	0.4433	0.4168	0.3938	0.3696	0.3436	0.3145	0.2855	0.2558	0.2229
disinfection(X)	0.6499	0.6513	0.6526	0.6508	0.6505	0.6964	0.6981	0.6997	0.7018	0.7022	0.7034	0.7043	0.7052	0.7520	**0.7540**
nucleic acid(X)	**0.4535**	0.4280	0.4065	0.3823	0.3535	0.3253	0.2918	0.2558	0.2199	0.1806	0.1393	0.0957	0.0481	0.0035	0.0592
lockdown of the city(X)	**0.6956**	0.6932	0.6902	0.6872	0.6858	0.6853	0.6844	0.6812	0.6778	0.6742	0.6709	0.6673	0.6632	0.6587	0.6537
Wuhan(X)	0.0945	0.0912	0.1337	0.1827	0.2412	0.3006	0.3612	0.4063	0.4712	0.5192	0.5644	0.5469	0.5413	**0.5848**	0.5643
Zhong Nanshan(X)	0.0065	0.0335	0.0766	0.1237	0.1759	0.2338	0.2898	0.3537	0.4185	0.4877	0.5629	**0.6452**	0.6338	0.6157	0.5959
epidemic(X)	0.2695	0.2373	0.2054	0.1743	0.1386	0.0993	0.0572	0.0069	0.0407	0.0898	0.1436	0.2026	0.2658	**0.3348**	0.3241
flu(X)	**0.4739**	0.4532	0.4316	0.4130	0.3924	0.3645	0.3330	0.3015	0.2694	0.2353	0.1992	0.1599	0.1174	0.0768	0.1008
diarrhea(X)	0.0572	0.0468	0.0358	0.0243	0.0023	0.0183	0.0300	0.0498	0.0654	0.0818	0.1045	0.1283	0.1509	0.1764	**0.2035**
stuffy nose(X)	0.0938	0.1273	0.1524	0.1789	0.1974	0.2250	0.2478	0.2884	0.3185	0.3471	0.4070	0.4629	0.5096	0.5600	**0.6164**
dry cough(X)	0.1158	0.0810	0.0439	0.0048	0.0403	0.0873	0.1414	0.1983	0.2547	0.3149	0.3813	**0.4389**	0.4070	0.3990	0.3897
NCP(X)	**0.2867**	0.2724	0.2662	0.2596	0.2475	0.2283	0.2122	0.1936	0.1769	0.1592	0.1403	0.1178	0.0954	0.0714	0.0456
close contact(X)	**0.7132**	0.7060	0.6949	0.6858	0.6782	0.6702	0.6597	0.6444	0.6300	0.6163	0.6016	0.5857	0.5682	0.5483	0.5266
suspected case(X)	**0.4649**	0.4406	0.4211	0.3985	0.3715	0.3440	0.3136	0.2746	0.2405	0.2043	0.1641	0.1214	0.0748	0.0237	0.0315
pneumonia(X)	**0.4079**	0.3867	0.3628	0.3407	0.3154	0.2871	0.2562	0.2203	0.1859	0.1511	0.1134	0.0717	0.0273	0.0144	0.0290
cough(X)	0.4278	0.4759	0.5282	0.5770	0.5706	0.5538	0.5509	0.5581	0.5833	0.5958	0.6682	0.7430	0.7566	0.7674	**0.7698**
dyspnea(X)	0.1956	0.1643	0.1255	0.0834	0.0386	0.0009	0.0116	0.0539	0.0519	0.0503	0.1023	0.1543	0.2101	0.2443	**0.2555**
virus pneumonia(X)	**0.7062**	0.7032	0.6900	0.6832	0.6807	0.6787	0.6726	0.6653	0.6805	0.6733	0.6620	0.6436	0.6273	0.6095	0.5865
novel coronavirus(X)	**0.6044**	0.5898	0.5741	0.5595	0.5424	0.5229	0.5009	0.4750	0.4518	0.4277	0.4017	0.3726	0.3411	0.3063	0.2688

Optimal values are shown in bold in the table

Table 2 The final keyword list of COVID-19

Category	Keyword
prevention	mask(S), mask(X), disinfection(X), nucleic acid(X)
symptoms	flu(S), diarrhea(S), dry cough(S), asymptomatic infection(S), pneumonia(S), cough(S), fever(S), dyspnea(S), high temperature(S), infection(X), stuffy nose(X), close contact(X), suspected case(X), pneumonia(X), cough(X)
common words	coronavirus(S), novel(S), epidemic(S), NCP(S), novel coronary pneumonia(S), COVID-19(S), novel coronavirus(S), coronavirus(X), virus pneumonia(X), novel coronavirus(X)
other related aspects of COVID-19	Red Cross Society(S), Spring Festival(S), Health Commission(S), SARS(S), lockdown of the city(S), Wuhan(S), Zhong Nanshan(S), Spring Festival(X), Red Cross Society(X), Health Commission(X), SARS(X), children(X), lockdown of the city(X), Wuhan(X), Zhong Nanshan(X)

The Baidu Index of the keyword here included search index and information index, where (S) describes the search index of the keyword, and (X) represents the information index of the keyword

F23. On F8, LsOA also had a weak advantage over other algorithms and achieved good results. For F10 and F14, LsOA tied for first with some algorithms and obtained the same results. On F12 and F13, LsOA showed an absolute advantage.

We comprehensively considered the test results of these benchmark functions. The last row in Table 5 is the final ranking result. The overall performance of LsOA on these 23 commonly used benchmark functions was excellent, with an average ranking of 1.17, ranking first in the comprehensive ranking, which was in turn superior to KO, MPA, TLBO, HBO, GWO, WOA, and PSO. The above experiments prove the superior performance of LsOA, which also shows that we can use this algorithm to screen the prediction features of COVID-19.

3.5 Setting training parameters and comparison methods

3.5.1 Training parameters

The parameters used in the GCN training model included the length of history\future time series, feature dimension, learning rate, training epoch, batch size, number of hidden layer nodes, optimizer, loss function, and so on. The processor of the experimental hardware environment was Intel (R) Pentium (R) CPU N3700 @ 1.60 GHz 1.60 GHz, a server with 4 GB of memory, and the software environment consisted of Python 3.6.12 and TensorFlow 1.14.0.

Among the training parameters, the number of hidden layer nodes is a very important parameter in the GCN model. Different numbers of hidden layer nodes will greatly affect the prediction accuracy. Too many nodes will cause over-fitting, and too few nodes will cause the model to be inaccurate. Therefore, in this part, we conducted a sensitivity analysis on the number of hidden layer nodes and conducted experiments with different numbers of hidden layer nodes, to analyze the performance of the model and finally determine the optimal number of hidden layer nodes by comparing the prediction results.

In this experiment, we determined the number of hidden layer nodes from {8, 16, 32, 64, 80, 100, 128, 144, 160}, conducted experiments on the premise that the number of training is 3000, and analyzed the change of prediction accuracy. Finally, the results of sensitivity analysis for different numbers of hidden layer nodes are shown in Fig. 10, where the abscissa represents the number of hidden layer nodes and the ordinate represents the corresponding value of each indicator.

Figure 10 shows the changes of various index values under different numbers of hidden layer nodes when the number of training times is 3000, where Fig. 10a and c, respectively, shows the changes of RMSE, MAE, and MAPE as the number of hidden layer nodes increases. It can be seen from these figures that when the number of hidden layer nodes is 16, the corresponding RMSE and MAE are the smallest, and MAPE is the smallest when the number of hidden layer nodes is 32, indicating that the difference between the predicted value and the actual value in the corresponding state is also the smallest. Figure 10b shows the changes of Accuracy and Var as the number of hidden layer nodes increases. It can be concluded from Fig. 10b that when the number of hidden layer nodes is 16, these three indicators all reach the maximum value, indicating that the model fitting effect is the best at this time. In summary, we could conclude that when the training times are set to 3000, the model has the best effect when the number of hidden layer nodes is set to 16.

We can also see from these figures that as the number of hidden layer nodes increases, the prediction accuracy first decreases, then increases, and then decreases. This is mainly because when the hidden layer node is larger than a certain level, the model complexity and calculation difficulty will greatly increase, thereby reducing the prediction

Table 3 Comparison of prediction results in Hubei Province

Data set	Indicator	Data (sorted)	Data (unsorted)
Hubei Province	RMSE	322.277	1030.61
	MAE	74.0729	308.891
	MAPE	0.02144	0.08120
	Acc	0.97360	0.91558
	R2	0.99922	0.99202
	Var	0.99925	0.99274

accuracy. Therefore, according to the results of sensitivity analysis, we also set the number of hidden layer nodes as 16 in the subsequent experiment of LsOA-GCN. Finally, the parameters used for the LsOA-GCN training model we obtained are shown in Table 6.

3.5.2 Comparison methods

This part mainly compared several representative time series prediction methods in current epidemic prediction research, including multiple linear regression, support vector regression, backpropagation neural network, kernel ridge regression, Gaussian process regression, stochastic gradient descent regression, GAT, LSTM, and T-GCN models. The nine comparison methods are described as follows:

(1) Multiple Linear Regression (Yuchi et al. 2019). When performing regression analysis, we call regression with two or more independent variables under linear correlation conditions as multiple linear regression (MLR). This method uses the optimal combination of multiple independent variables to predict or estimate the dependent variable together. Because the relationship between multiple independent variables is integrated, it is usually a better method than linear regression, which is more effective and more realistic than using only one independent variable for prediction or estimation. In the comparison experiment, the parameters used in this method were all default values.

(2) Support Vector Regression (Brereton and Lloyd 2010). Support vector machine (SVM) is a class of generalized linear classifiers that perform binary classification of data in a supervised learning manner. Support vector regression (SVR) is an application model of support vector machines in regression problems. The support vector machine regression model has many variants based on different loss functions. The core idea is to find a separating hyperplane (hypersurface) that minimizes the expected risk. In the comparison experiment, the kernel type used in this method was 'poly', the penalty parameter C = 1.1, the kernel coefficient gamma of 'poly' was set as 'auto', the degree of the polynomial kernel function ('poly') was 3, epsilon = 0.1, the independent term in the kernel function coef0 = 1.0, and the rest of the parameters used default values.

Table 4 Comparison of prediction results of 17 regions in Hubei Province

Data		Data (sorted))				Data (unsorted)			
Number	Region	RMSE	MAE	MAPE	Acc	RMSE	MAE	MAPE	Acc
0	Wuhan	1318.24	881.730	0.01763	0.97364	4219.93	3823.86	0.07647	0.91561
1	Ezhou	38.6784	26.1887	0.01879	0.97225	118.111	106.713	0.07655	0.91527
2	Enshi	6.72733	4.83126	0.01917	0.97330	21.2020	19.3225	0.07668	0.91586
3	Huanggang	83.8352	60.6461	0.02086	0.97116	248.487	226.466	0.07790	0.91452
4	Huangshi	27.9862	18.5732	0.01830	0.97243	86.9989	78.6410	0.07748	0.91429
5	Jingmen	24.3426	16.3154	0.01758	0.97377	79.0874	72.2794	0.07789	0.91478
6	Jingzhou	46.1738	31.7411	0.02009	0.97078	137.985	125.721	0.07957	0.91267
7	Qianjiang	5.51831	3.60669	0.01822	0.97213	16.7239	15.2088	0.07681	0.91554
8	Shennongjia	0.27478	0.20933	0.01903	0.97502	1.04942	0.97174	0.08834	0.90460
9	Shiyan	19.7104	13.8861	0.02066	0.97067	58.0828	53.0211	0.07890	0.91357
10	Suizhou	37.5744	25.8064	0.01974	0.97125	111.841	101.834	0.07791	0.91443
11	Tianmen	16.5381	12.1464	0.02449	0.96666	45.3164	41.8090	0.08429	0.90864
12	Xiantao	16.1987	10.9648	0.01907	0.97183	48.2514	43.9107	0.07637	0.91608
13	Xianning	47.6550	43.2417	0.05172	0.94300	103.582	98.5077	0.11783	0.87610
14	Xiangyang	32.7713	22.5675	0.01921	0.97211	101.452	92.3979	0.07864	0.91366
15	Xiaogan	96.6878	67.4415	0.01917	0.97252	302.343	275.598	0.07834	0.91406
16	Yichang	27.8529	19.3436	0.02078	0.97008	81.9497	74.8775	0.08043	0.91198

Table 5 Test results of different algorithms on common benchmark functions

Function		LsOA	PSO	GWO	WOA	TLBO	HBO	MPA	KO
F1	Ave	3.47E-167	2.49E + 00	1.47E-27	2.40E-74	1.42E-38	6.09E-07	1.16E-09	**1.81E-281**
	Rank	2	8	5	3	4	7	6	1
	Std	0.00E + 00	9.09E-01	2.82E-27	9.81E-74	1.11E-38	7.66E-07	9.16E-10	0.00E + 00
F2	Ave	4.19E-84	4.40E + 00	9.96E-17	2.35E-49	8.92E-20	8.15E-06	6.23E-06	**4.25E-145**
	Rank	2	8	5	3	4	7	6	1
	Std	2.30E-83	1.20E + 00	7.23E-17	1.18E-48	4.22E-20	1.06E-05	3.33E-06	1.58E-144
F3	Ave	5.23E-112	1.98E + 02	1.21E-05	3.95E + 04	5.99E-07	2.51E + 04	1.21E + 00	**2.64E-167**
	Rank	2	6	4	8	3	7	5	1
	Std	1.87E-111	6.53E + 01	2.20E-05	1.23E + 04	5.80E-07	7.02E + 03	1.31E + 00	0.00E + 00
F4	Ave	1.46E-84	1.96E + 00	1.05E-06	4.55E + 01	4.38E-16	1.28E + 01	3.55E-04	**8.45E-137**
	Rank	2	6	4	8	3	7	5	1
	Std	8.00E-84	1.93E-01	1.37E-06	2.88E + 01	2.63E-16	4.96E + 00	1.15E-04	2.45E-136
F5	Ave	**5.59E-17**	1.18E + 03	2.72E + 01	2.79E + 01	2.57E + 01	1.08E + 02	2.68E + 01	2.73E + 01
	Rank	1	8	4	6	2	7	3	5
	Std	3.06E-16	7.58E + 02	7.77E-01	5.31E-01	4.78E-01	5.51E + 01	5.09E-01	4.34E-01
F6	Ave	**1.28E-17**	2.20E + 00	7.89E-01	3.82E-01	3.01E-03	6.92E-07	1.02E-01	2.59E-01
	Rank	1	8	7	6	3	2	4	5
	Std	2.79E-17	9.58E-01	4.22E-01	2.34E-01	5.29E-03	1.25E-06	9.23E-02	1.50E-01
F7	Ave	**1.40E-04**	1.32E + 01	1.75E-03	3.87E-03	2.41E-03	3.52E-02	2.85E-03	4.97E-04
	Rank	1	8	3	6	4	7	5	2
	Std	1.23E-04	1.11E + 01	8.78E-04	3.91E-03	8.77E-04	1.12E-02	1.36E-03	4.57E-04
F8	Ave	**-1.26E + 04**	-6.25E + 03	-6.25E + 03	-1.02E + 04	-6.22E + 03	-1.16E + 04	-8.36E + 03	-6.36E + 03
	Rank	1	7	6	3	8	2	4	5
	Std	1.10E-09	1.53E + 03	7.54E + 02	1.91E + 03	1.23E + 03	3.44E + 02	4.28E + 02	1.37E + 03
F9	Ave	**0.00E + 00**	1.58E + 02	2.78E + 00	**0.00E + 00**	1.90E + 01	1.51E + 01	5.44E-04	**0.00E + 00**
	Rank	1	6	3	1	5	4	2	1
	Std	0.00E + 00	3.02E + 01	3.88E + 00	0.00E + 00	9.94E + 00	5.52E + 00	2.97E-03	0.00E + 00
F10	Ave	**8.88E-16**	2.71E + 00	1.07E-13	3.73E-15	5.03E-15	3.12E-02	6.47E-06	**8.88E-16**
	Rank	1	7	4	2	3	6	5	1
	Std	0.00E + 00	3.84E-01	1.90E-14	2.36E-15	1.35E-15	1.70E-01	2.65E-06	1.00E-31
F11	Ave	**0.00E + 00**	1.17E-01	5.84E-03	1.89E-02	**0.00E + 00**	1.73E-03	7.11E-09	**0.00E + 00**
	Rank	1	6	4	5	1	3	2	1
	Std	0.00E + 00	5.00E-02	9.38E-03	4.90E-02	0.00E + 00	4.25E-03	9.95E-09	0.00E + 00
F12	Ave	**9.93E-14**	5.27E-02	4.14E-02	2.19E-02	1.26E-04	3.46E-03	4.47E-03	2.56E-02
	Rank	1	8	7	5	2	3	4	6
	Std	4.34E-13	6.47E-02	1.91E-02	2.26E-02	4.13E-04	1.89E-02	3.22E-03	1.09E-02
F13	Ave	**6.05E-13**	5.14E-01	6.70E-01	4.85E-01	1.38E-01	2.56E-03	1.30E-01	2.20E-01
	Rank	1	7	8	6	4	2	3	5
	Std	2.90E-12	1.94E-01	2.47E-01	2.38E-01	1.18E-01	8.50E-03	7.24E-02	7.97E-02
F14	Ave	**0.99800**	3.19702	3.97097	3.74390	**0.99800**	**0.99800**	**0.99800**	1.19346
	Rank	1	3	5	4	1	1	1	2

Table 5 (continued)

Function		LsOA	PSO	GWO	WOA	TLBO	HBO	MPA	KO
	Std	2.52E-16	2.81E + 00	3.91E + 00	3.74E + 00	4.12E-17	0	4.19E-16	0.54228
F15	Ave	**0.00031**	0.00092	0.00713	0.00070	0.00118	0.00079	**0.00031**	0.00056
	Rank	1	5	7	3	6	4	1	2
	Std	2.43E-13	1.29E-04	9.52E-03	3.62E-04	3.64E-03	1.93E-04	2.30E-06	1.91E-04
F16	Ave	**-1.03163**	**-1.03163**	**-1.03163**	**-1.03163**	**-1.03163**	**-1.03163**	**-1.03163**	-1.03162
	Rank	1	1	1	1	1	1	1	2
	Std	6.10E-12	4.88E-16	1.81E-05	3.03E-09	6.58E-16	6.39E-16	3.32E-14	7.93E-06
F17	Ave	**0.39789**	**0.39789**	0.39790	0.39790	**0.39789**	**0.39789**	**0.39789**	**0.39789**
	Rank	1	1	2	2	1	1	1	1
	Std	2.10E-12	0.00E + 00	5.68E-05	2.26E-05	0.00E + 00	0.00E + 00	2.09E-12	4.59E-06
F18	Ave	**3.00000**	**3.00000**	3.00005	3.00010	**3.00000**	**3.00000**	**3.00000**	3.00018
	Rank	1	1	2	3	1	1	1	4
	Std	5.83E-14	5.34E-15	4.62E-05	2.38E-04	1.39E-15	1.38E-15	2.54E-14	2.09E-04
F19	Ave	**-3.86278**	**-3.86278**	-3.86149	-3.85491	**-3.86278**	**-3.86278**	**-3.86278**	-3.86210
	Rank	1	1	3	4	1	1	1	2
	Std	5.48E-12	1.89E-15	2.66E-03	1.11E-02	2.71E-15	2.71E-15	1.74E-12	8.32E-04
F20	Ave	**-3.32200**	-3.27047	-3.25757	-3.18350	-3.30784	**-3.32200**	**-3.32200**	-3.22099
	Rank	1	3	4	6	2	1	1	5
	Std	1.22E-09	5.99E-02	7.96E-02	1.79E-01	3.67E-02	5.95E-15	3.50E-10	6.66E-02
F21	Ave	**-10.1532**	-7.30681	-9.47737	-7.58607	-10.1493	-9.31835	**-10.1532**	-10.1422
	Rank	1	7	4	6	2	5	1	3
	Std	9.96E-08	3.21E + 00	1.75E + 00	3.01E + 00	1.66E-02	2.21E + 00	2.59E-09	1.89E-02
F22	Ave	**-10.4029**	-8.21827	-9.97077	-7.77028	-10.4028	-10.1733	**-10.4029**	-10.3930
	Rank	1	6	5	7	2	4	1	3
	Std	2.44E-08	3.23E + 00	1.67E + 00	2.86E + 00	9.84E-04	1.26E + 00	2.18E-09	9.95E-03
F23	Ave	**-10.5364**	-9.03874	-10.3560	-7.57344	-10.3551	-10.1484	**-10.5364**	-10.3476
	Rank	1	6	2	7	3	5	1	4
	Std	4.86E-08	2.78E + 00	9.79E-01	3.33E + 00	9.92E-01	1.48E + 00	2.87E-09	9.86E-01
Sum rank		27	127	99	105	66	88	64	63
Mean rank		1.17	5.52	4.30	4.57	2.87	3.83	2.78	2.74
Final rank		1	8	6	7	4	5	3	2

Optimal values are shown in bold in the table

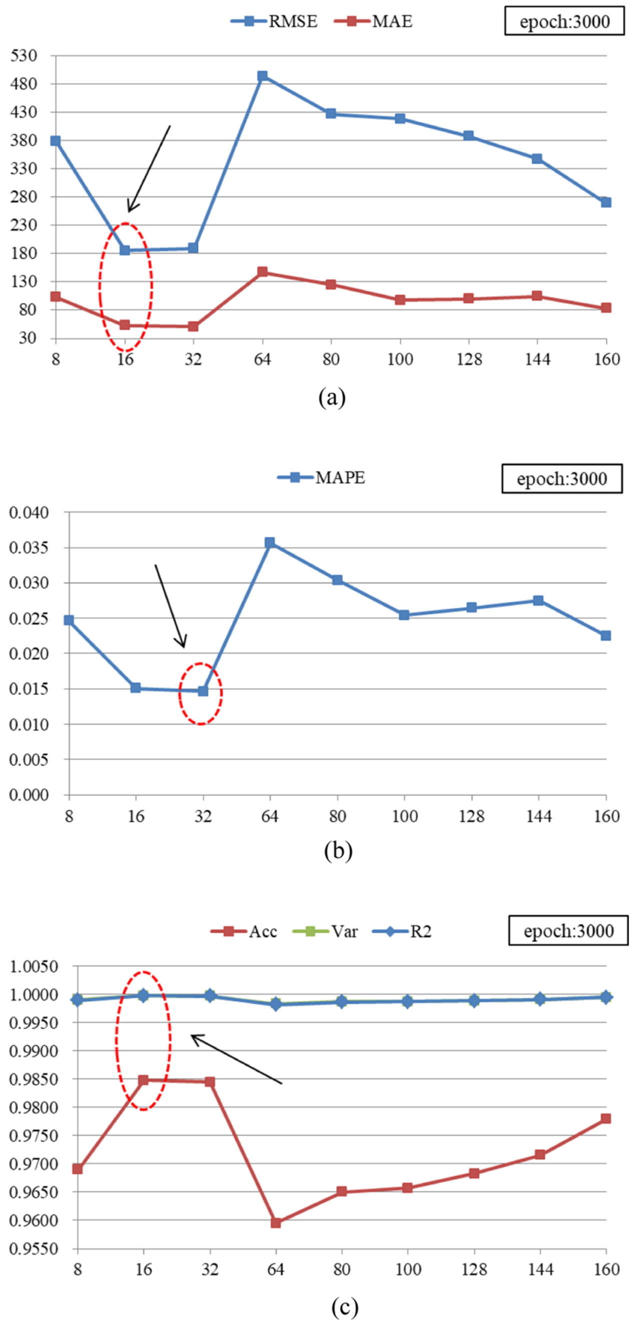


Fig. 10 Sensitivity analysis of hidden layer nodes

(3) BP neural network (Wang et al. 2015). BP neural network (backpropagation neural network, BPNN) is a multilayer feedforward network trained according to error backpropagation. Its basic idea is the gradient descent method, which uses gradient search technology to minimize the mean square error between the real output value and the expected output value of the network. It is the most widely used neural network. The BPNN used in the comparison experiment was a three-layer network structure, and the node number of input layer, hidden layer, and output layer was 16, 32, and 1, respectively. The activation function was relu, loss was the mean square loss function, the optimizer used Adam algorithm, the training epoch was set to 3000, and the batch size was 7.

(4) Kernel Ridge Regression (Wu and Zhao 2020). The kernel ridge regression (KRR) algorithm introduces the kernel method based on the ridge regression algorithm. The independent variable space is mapped to the high-dimensional feature space through the kernel function, and then, the ridge regression method is used to analyze and process the data in the high-dimensional feature space. In other words, kernel ridge regression is a kind of regression algorithm, and its essence is to use ridge regression to make predictions. Ridge regression is linear regression, and it fits a straight line. When facing a nonlinear situation, you can map the independent variable space to the high-dimensional feature space by adding a kernel function, and then perform linear regression in the high-dimensional feature space. For the KRR model used in the comparison experiment, alpha = 1.0, the kernel used a linear kernel, the regularization strength coefficient was set to 0.9, and the rest of the parameters were default values.

(5) Gaussian Process Regression (Arthur et al. 2020). Gaussian process (GP) is a commonly used supervised learning method, which aims to solve regression problems and probabilistic classification problems. Gaussian process regression (GPR) implements the Gaussian process model in the case of regression. It is a nonparametric model that uses Gaussian process priors to perform regression analysis on data. In the comparison experiment, the kernel object used in this method was kernel = DotProduct() + WhiteKernel(), the generator used to initialize the center was random_state = 0, and the rest of the parameters adopted default values.

(6) Stochastic Gradient Descent Regression (Ighalo et al. 2020). Stochastic gradient descent (SGD) is a simple but highly efficient method, which is mainly used for discriminant learning of linear classifiers under convex loss functions, such as (linear) support vector machines and logistic regression. Stochastic gradient descent regression supports different loss functions and penalties to fit linear regression models. The SGD model used in the comparison experiment, the maximum number of iterations was set as 1000, the stopping criterion tol was set as '0.001', and the rest of the parameters adopted default values.

(7) GAT (Velikovi et al. 2017). Graph attention network (GAT), which aggregates neighbor nodes through the attention mechanism and realizes the adaptive allocation of different neighbor weights. This is different from GCN. The weights of different neighbors in GCN are

Table 6 Parameters of training model

Training parameters	Mode or value
History time series/d	1
Future time series/d	1
Training set /d	59
Test set /d	7
Feature dimension	44
Learning rate	0.001
Training epoch	3000
Batch size	7
Hidden layer nodes	16
Network model	GCN
Optimizer	Adam
Number of graph convolutional layers	2
Loss function	ReLu
Feature optimization algorithm	LsOA
Population size	30
Max iteration	100

fixed, and they all come from the normalized Laplacian matrix. GAT greatly improves the expressive ability of the graph neural network model. The GAT model used in the comparison experiment was a double-headed three-layer network structure, the number of nodes in the three layers was 45, 16, and 1, respectively. The activation function was LeakyReLU, loss was the mean square loss function, the optimizer used the Adam algorithm, the default learning rate was 0.001, the training epoch was set to 3000, and the batch size was 7.

(8) LSTM (Hochreiter and Schmidhuber 1997). Long short-term memory (LSTM) is a time-cyclic neural network, which is specially designed to solve the long-term dependency problem of general recurrent neural network (RNN). The core concepts of LSTM lie in the cell state and the "gate" structure. The cell state is equivalent to the path of information transmission, so that information can be passed down in the serial connection, which can be regarded as the "memory" of the network. The addition and removal of information is achieved through the "door" structure, which learns which information should be saved or forgotten during the training process. The LSTM model used in the comparison experiment was a three-layer network structure. The number of nodes in the three layers was 45, 200, and 1, respectively. The optimizer used Adam algorithm, the initial learning rate was 0.005, the gradient threshold was set to 1, and after 125 cycles, the learning rate was reduced by multiplying by 0.1, the training epoch was 3000, and the batch size was 7.

(9) T-GCN (Zhao et al. 2020b). Temporal graph convolutional network (T-GCN) is proposed for traffic flow prediction and is mainly composed of two parts. One part is to use the graph convolutional network (GCN) to learn the connection relationship of roads in the traffic network graph, to model the spatial correlation of traffic temporal and spatial sequence data. The other part is to use gate recurrent unit (GRU) to obtain the time dependence of data. The T-GCN model used in the comparison experiment was composed of GCN and GRU. The optimizer used Adam algorithm, the initial learning rate was 0.001, the number of hidden layer nodes was 8, the activation function of the graph convolutional layer was sigmoid, the activation function of the GRU layer was tanh, the loss function was $loss = ||Y_t - \hat{Y}_t|| + \lambda L_{reg}$, the training epoch was 3000, and the batch size was 7.

3.6 Performance evaluation

In this section, in order to verify the effectiveness of the prediction model, we used the cumulative number of confirmed cases of COVID-19 from January 24, 2020, to March 29, 2020 (66 days in total), in 17 regions of Hubei Province and the Baidu Index of adjusted keywords as the data source, used the data in the data set before March 23, 2020, as the training set, and used the LsOA-GCN prediction model to predict the cumulative number of confirmed cases in the 7 days from March 23, 2020 to March 29, 2020. During training, the input data X_i is the keyword data on the i-th day, where $X_i = [x^1, x^2, \ldots, x^j]$, $1 \leq i \leq 66$, $1 \leq j \leq 44$, and the output value Y_{i+1} is the cumulative number of confirmed cases on the $(i+1)$-th day. Finally, the prediction results were compared with the results of 10 representative prediction methods in current epidemic prediction research, including multiple linear regression, support vector regression, backpropagation neural network, kernel ridge regression, Gaussian process regression, stochastic gradient descent regression, GAT, LSTM, T-GCN and original GCN model. In order to visually test the prediction accuracy of each model, we compared the predicted value of each model with the actual value through evaluation indicators and obtained the RMSE, MAE, MAPE, Acc, R2, and Var of each model. The final results are shown in Tables 7 and 8, where Table 7 shows the comprehensive prediction results of Hubei Province and Table 8 shows the respective prediction results of 17 regions in Hubei Province.

It can be seen from the comprehensive prediction results in Table 7 that compared with the other 10 methods, the RMSE, MAE, and MAPE values of the cumulative number of confirmed cases predicted by the LsOA-GCN model are

Table 7 Experimental results of LsOA-GCN and other methods–Hubei Province

Data	Metric	Methods										
		LsOA-GCN	GCN	LSTM	GAT	T-GCN	BPNN	GPR	KRR	MLR	SGD	SVR
Hubei Province	RMSE	**14.6572**	322.277	1047.15	2729.48	392.285	1459.15	3001.42	986.865	1250.69	1486.09	780.661
	MAE	**5.40047**	74.0729	259.864	740.637	308.649	327.823	690.042	312.087	328.534	494.022	222.975
	MAPE	**0.00350**	0.02144	0.03701	0.22437	0.55861	0.04046	0.06150	0.13179	0.05606	0.15712	0.04660
	Acc	**0.99880**	0.97360	0.91422	0.77642	0.96787	0.88047	0.75414	0.91916	0.89755	0.87827	0.93605
	R2	**1.00000**	0.99922	0.99176	0.94404	0.99884	0.98401	0.93233	0.99268	0.98825	0.98341	0.99542
	Var	**1.00000**	0.99925	0.99219	0.94621	0.99887	0.98473	0.93541	0.99336	0.98863	0.98484	0.99553

Optimal values are shown in bold in the table

the smallest, which are 14.65572, 5.4005, and 0.0035, respectively. It shows that the cumulative number of confirmed cases predicted by this method is closer to the actual value, and the error is smaller. For the three types of indicators, Acc, R2, and Var, the values are all close to 1, indicating that the prediction accuracy of the prediction model is high and it shows an absolute advantage. From the regional prediction results in Table 8, we could also see that when predicting the cumulative number of confirmed cases in Xianning, LsOA-GCN ranked third with only a slight difference in the two indicators of MAE and MAPE. In addition, for the three types of indicators, RMSE, MAE and MAPE, LsOA-GCN has an absolute advantage over other forecasting methods. For Acc, in these 17 regions, the value of Acc is within the range of 0.9914 to 0.9990. The closer Acc is to 1, the stronger the fitting degree between the actual value and the predicted value is, the higher the prediction accuracy is, and the more accurate the prediction method is. In addition, we also found that the same prediction model was used for 17 regions in Hubei Province, but the prediction results would be different to some extent. For example, when using LsOA-GCN for prediction, the RMSE value for Wuhan is 50.6662, Ezhou is 1.4078, and Enshi is 0.2504, which shows that the same prediction method has different prediction effects for different regions.

The convergence curve of Fig. 11 also proves the influence of using LsOA for feature screening on the prediction results, where the ordinate is the value of RMSE, and the abscissa is the number of iterations. It can be seen from Fig. 11 that at the beginning of the iteration, the RMSE value drops sharply. As the iteration progresses, the RMSE value gradually decreases and finally stabilizes. The final feature matrix was a 17*66*3 matrix, and its keyword features were dyspnea (S), novel coronavirus pneumonia (S) and stuffy nose (X). Among them, the two symptoms of dyspnea (S) and stuffy nose (X) are not unique to COVID-19, but they are easy to recognize and attract attention. For the general population, this is an indicator that may be infected with COVID-19. In addition, novel coronavirus pneumonia (S) is a commonly used term in the epidemic, indicating that as the epidemic continues to spread, people have further information needs for the basic knowledge of COVID-19. Based on the above analysis, we could believe that these three keywords can reflect the prevalence of COVID-19 to a certain extent and have certain predictive value.

In general, for these six types of evaluation indicators, the LsOA-GCN prediction model has shown superior performance in Hubei Province or in the 17 regions of Hubei and is the best prediction method.

Table 8 Experimental results of LsOA-GCN and other methods–17 regions in Hubei

Data	Metric	LsOA-GCN	GCN	LSTM	GAT	T-GCN	BPNN	GPR	KRR	MLR	SGD	SVR
Wuhan	RMSE	**50.6662**	1318.24	4310.42	11,134.0	787.196	6013.73	12,371.2	3992.33	5146.77	6062.66	3204.49
	MAE	**41.5418**	881.730	3832.77	8734.63	655.635	5071.99	10,897.2	3036.36	4731.97	5738.31	2964.49
	MAPE	**0.00083**	0.01763	0.07665	0.17467	0.01311	0.10143	0.21792	0.06072	0.09463	0.11475	0.05928
	Acc	**0.99899**	0.97364	0.91380	0.77735	0.98426	0.87974	0.75260	0.92016	0.89708	0.87876	0.93592
Ezhou	RMSE	**1.40783**	38.6784	43.1421	480.575	500.479	58.7762	128.618	151.441	137.801	221.969	57.7294
	MAE	**1.17987**	26.1887	34.2657	343.917	492.940	48.0810	105.159	122.883	115.187	211.224	51.6682
	MAPE	**0.00085**	0.01879	0.02458	0.24671	0.35362	0.03449	0.07544	0.08815	0.08263	0.15152	0.03706
	Acc	**0.99899**	0.97225	0.96905	0.65525	0.64098	0.95784	0.90773	0.89136	0.90115	0.84077	0.95859
Enshi	RMSE	**0.25039**	6.72733	9.18368	39.6688	258.790	7.19169	10.6373	55.7457	10.8859	47.8082	14.0554
	MAE	**0.20125**	4.83126	8.06020	33.0289	204.116	5.77993	8.63636	52.0468	8.79418	46.4684	13.3372
	MAPE	**0.00080**	0.01917	0.03198	0.13107	0.80999	0.02294	0.03427	0.20654	0.03490	0.18440	0.05293
	Acc	**0.99901**	0.97330	0.96356	0.84258	-0.02694	0.97146	0.95779	0.77879	0.95680	0.81029	0.94422
Huanggang	RMSE	**3.22323**	83.8352	114.250	786.510	287.178	90.6541	188.220	488.267	198.876	474.418	209.138
	MAE	**2.51078**	60.6461	96.6943	607.871	271.633	85.5400	160.247	483.764	168.108	461.406	191.306
	MAPE	**0.00086**	0.02086	0.03326	0.20911	0.09344	0.02943	0.05512	0.16641	0.05783	0.15872	0.06581
	Acc	**0.99889**	0.97116	0.96070	0.72944	0.90121	0.96882	0.93525	0.83204	0.93159	0.83680	0.92806
Huangshi	RMSE	**1.01808**	27.9862	23.0499	299.370	160.273	30.0910	69.7950	126.014	67.6836	147.611	60.1148
	MAE	**0.85338**	18.5732	19.3720	231.617	131.782	27.0843	41.0881	109.056	39.0791	139.819	53.9503
	MAPE	**0.00084**	0.01830	0.01909	0.22819	0.12983	0.02668	0.04048	0.10744	0.03850	0.13775	0.05315
	Acc	**0.99900**	0.97243	0.97729	0.70505	0.84210	0.97035	0.93124	0.87585	0.93332	0.85457	0.94077
Jingmen	RMSE	**0.98352**	24.3426	24.3273	322.172	513.461	7.52676	64.2156	124.861	65.6214	123.718	62.6998
	MAE	**0.79078**	16.3154	21.5688	252.195	511.540	6.41007	55.2451	109.424	56.6386	112.163	60.6048
	MAPE	**0.00085**	0.01758	0.02324	0.27176	0.55123	0.00691	0.05953	0.11791	0.06103	0.12086	0.06531
	Acc	**0.99894**	0.97377	0.97379	0.65283	0.44670	0.99189	0.93080	0.86545	0.92929	0.86668	0.93244
Jingzhou	RMSE	**1.68224**	46.1738	58.4515	297.973	486.231	45.3420	45.5418	160.072	47.9307	229.225	104.973
	MAE	**1.36306**	31.7411	48.5009	250.867	463.086	32.2055	39.8533	147.699	41.4051	218.332	91.0327
	MAPE	**0.00086**	0.02009	0.03070	0.15878	0.29309	0.02038	0.02522	0.09348	0.02621	0.13818	0.05762
	Acc	**0.99894**	0.97078	0.96301	0.81141	0.69226	0.97130	0.97118	0.89869	0.96966	0.85492	0.93356
Qianjiang	RMSE	**0.20008**	5.51831	9.17594	88.2524	500.587	17.6562	13.2697	39.1590	12.8428	45.1208	8.19654
	MAE	**0.16711**	3.60669	6.96984	62.9022	480.283	16.0132	10.6344	34.1029	10.5190	42.8084	7.39665
	MAPE	**0.00084**	0.01822	0.03520	0.31769	2.42567	0.08087	0.93298	0.17224	0.05313	0.21620	0.03736
	Acc	**0.99899**	0.97213	0.95366	0.55428	-1.52822	0.91083	0.05371	0.80223	0.93514	0.77212	0.95860
Shennongjia	RMSE	**0.01086**	0.27478	0.73771	1.60061	32.6534	0.75230	1.36115	2.38661	1.39890	2.05334	0.46899
	MAE	**0.00902**	0.20933	0.66089	1.25474	27.5576	0.57620	0.96659	2.06045	0.97459	1.95007	0.34058
	MAPE	**0.00082**	0.01903	0.06008	0.11407	2.50524	0.05238	0.08787	0.18731	0.08860	0.17728	0.03096
	Acc	**0.99901**	0.97502	0.93294	0.85449	-1.96849	0.93161	0.87626	0.78304	0.87283	0.81333	0.95736

Table 8 (continued)

Data	Metric	LsOA-GCN	GCN	LSTM	GAT	T-GCN	BPNN	GPR	KRR	MLR	SGD	SVR
Shiyan	RMSE	**0.71915**	19.7104	26.1772	157.081	274.934	37.1574	50.4028	104.800	51.0943	109.373	20.1123
	MAE	**0.57114**	13.8861	19.2230	131.183	214.508	29.4689	44.9231	98.6214	45.5895	105.590	17.4851
	MAPE	**0.00085**	0.02066	0.02861	0.19521	0.31921	0.04385	0.06685	0.14676	0.06784	0.15713	0.02602
	Acc	**0.99893**	0.97067	0.96105	0.76625	0.59087	0.94471	0.92500	0.84405	0.92397	0.83724	0.97007
Suizhou	RMSE	**1.39058**	37.5744	51.8354	645.079	468.449	28.2146	98.4990	263.211	96.7216	257.949	49.9525
	MAE	**1.12534**	25.8064	43.1949	388.660	398.648	23.9810	88.3543	246.807	85.9936	249.472	45.7563
	MAPE	**0.00086**	0.01974	0.03305	0.29737	0.30501	0.01835	0.06760	0.18883	0.06579	0.19087	0.03501
	Acc	**0.99894**	0.97125	0.96034	0.50644	0.64158	0.97841	0.92464	0.79861	0.92600	0.80264	0.96178
Tianmen	RMSE	**4.26481**	16.5381	17.3045	148.241	269.851	20.8620	39.0279	49.2429	51.5244	80.0063	21.3729
	MAE	**4.24731**	12.1464	14.1336	117.714	222.573	19.5843	28.7870	45.1768	37.9261	76.1957	15.5669
	MAPE	**0.00856**	0.02449	0.02850	0.23733	0.44873	0.03948	0.05804	0.09108	0.07646	0.15362	0.03138
	Acc	**0.99140**	0.96666	0.96511	0.70113	0.45595	0.95794	0.92131	0.90072	0.89612	0.83870	0.95691
Xiantao	RMSE	**0.60402**	16.1987	32.0901	231.922	186.895	37.8016	41.9303	89.9615	42.6721	118.372	21.3962
	MAE	**0.48564**	10.9648	25.1861	174.238	167.300	35.1417	33.5490	88.2869	34.1701	116.414	20.2326
	MAPE	**0.00084**	0.01907	0.04380	0.30302	0.29096	0.06112	0.05835	0.15354	0.05943	0.20246	0.03519
	Acc	**0.99895**	0.97183	0.94419	0.59666	0.67497	0.93426	0.92708	0.84355	0.92579	0.79414	0.96279
Xianning	RMSE	**32.1149**	47.6550	62.4732	535.267	152.921	51.3795	32.2676	113.252	35.3993	167.479	66.8367
	MAE	32.1073	43.2417	58.9477	360.307	122.046	46.4659	**24.5145**	104.926	27.0204	157.611	60.9347
	MAPE	0.03841	0.05172	0.07051	0.43099	0.14599	0.05558	**0.02932**	0.12551	0.03232	0.18853	0.07289
	Acc	**0.96159**	0.94300	0.92527	0.35973	0.81708	0.93854	0.96140	0.86453	0.95766	0.79967	0.92005
Xiangyang	RMSE	**1.23615**	32.7713	25.3289	262.735	427.916	35.6762	64.3560	131.025	65.7992	141.774	98.3506
	MAE	**0.98278**	22.5675	21.0696	195.030	419.414	30.1233	58.5119	120.568	59.9261	131.521	97.2247
	MAPE	**0.00084**	0.01921	0.01793	0.16598	0.35695	0.02564	0.04980	0.10261	0.05100	0.11193	0.08274
	Acc	**0.99895**	0.97211	0.97844	0.77640	0.63582	0.96964	0.94523	0.88849	0.94400	0.87934	0.91630
Xiaogan	RMSE	**3.59745**	96.6878	175.942	790.340	69.5158	47.2554	105.199	397.664	90.0374	503.097	82.9717
	MAE	**2.88957**	67.4415	136.026	538.574	55.8627	42.1431	97.4514	391.336	86.0844	463.377	72.2803
	MAPE	**0.00082**	0.01917	0.03867	0.15309	0.01588	0.01198	0.02770	0.11124	0.02447	0.13172	0.02055
	Acc	**0.99898**	0.97252	0.94999	0.77534	0.98024	0.98657	0.97010	0.88696	0.97441	0.85699	0.97642
Yichang	RMSE	**0.97179**	27.8529	36.2150	214.007	452.742	70.5527	48.6621	125.226	50.7516	140.394	29.1905
	MAE	**0.78178**	19.3436	31.0428	166.834	408.107	52.4005	35.5812	112.358	35.6877	125.705	26.9711
	MAPE	**0.00084**	0.02078	0.03334	0.17920	0.43835	0.05628	0.03822	0.12069	0.03833	0.13502	0.02897
	Acc	**0.99896**	0.97008	0.96110	0.77013	0.51370	0.92422	0.94773	0.86549	0.94549	0.84920	0.96865

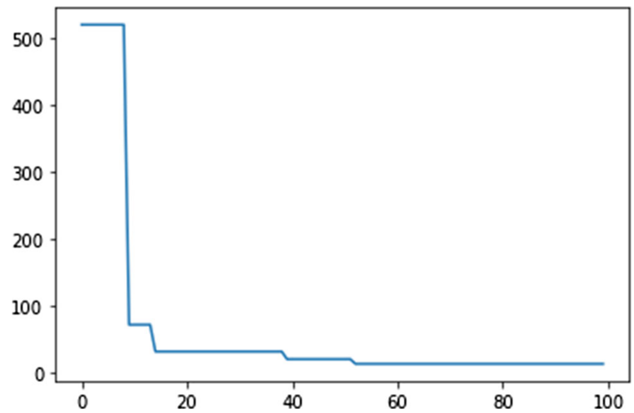


Fig. 11 Convergence curve

3.7 Statistical test

Before statistical analysis, Friedman was used to calculate the average ranking of various prediction methods on different indicators. The results are shown in Table 9. According to Friedman mean rank, LsOA-GCN performed best when considering 17 regions in Hubei Province, ranking first under the four indicators of RMSE, MAE, MAPE, and Acc. Among them, for RMSE, the average ranking of LsOA-GCN was 1.00, which was superior to GCN, LSTM, BPNN, SVR, GPR, MLR, KRR, SGD, T-GCN, GAT, respectively, by 182%, 324%, 329%, and 376%, 482%, 512%, 729%, 800%, 853%, 912%; for MAE, the average ranking of LsOA-GCN was 1.12, which was superior to GCN, LSTM, BPNN, SVR, GPR, MLR, KRR, SGD, T-GCN, GAT, respectively, by 121%, 284%, 295%, 363%, 405%, 426%, 637%, 721%, 753%, 800%; for MAPE, the average ranking of LsOA-GCN was 1.12, which was superior to GCN, LSTM, BPNN, SVR, GPR, MLR, KRR, SGD, T-GCN, GAT, respectively, by 121%, 284%, 289%, 363%, 426%, 426%, 632%, 716%, 753%, 795%; for Acc, the average ranking of LsOA-GCN was 11.00, which was superior to GCN, LSTM, BPNN, SVR, GPR, MLR, KRR, SGD, T-GCN, GAT, respectively, by 17%, 29%, 29%, 34%, 46%, 47%, 66%, 72%, 78%, 82%. These results once again verify the superior performance of the LsOA-GCN prediction method. The calculation method is as follows:

$$percentage = \frac{rank_i - rank_{LsOA-GCN}}{rank_{LsOA-GCN}} \tag{29}$$

where $rank_i$ represents the ranking of the prediction method i, i is the prediction method used for comparison, and $rank_{LsOA-GCN}$ is the ranking of LsOA-GCN.

When comparing methods, it is often necessary to perform statistical tests on experimental results. This part used some statistical test methods for comparing methods. First, a nonparametric Friedman test was used to determine whether there are significant differences in the performance of all methods under each indicator. In order to make a reliable comparison, we considered 11 methods and tested them on 4 different indicators for 17 regions, where the indicator used in the first group was RMSE, the second group was MAE, the third group was MAPE, and the fourth group was Acc. First, the Friedman test requires the average ranking to be calculated first. Then, the Friedman test should consider the critical value obtained at the significance level (α = 0.05, 0.1) and compare the critical value with Friedman's statistical results to determine whether there is evidence that the null hypothesis is false. For the tests of these four indicators, we found that the results rejected the null hypothesis, which indicates that the performance of each group of methods has significant differences.

Table 9 Friedman test ranking results

Friedman mean rank	RMSE	MAE	MAPE	Acc
LsOA-GCN	**1.00**	**1.12**	**1.12**	**11.00**
GCN	2.82	2.47	2.47	9.18
LSTM	4.24	4.29	4.29	7.76
GAT	10.12	10.06	10.00	1.94
T-GCN	9.53	9.53	9.53	2.47
BPNN	4.29	4.41	4.35	7.76
GPR	5.82	5.65	5.88	5.94
KRR	8.29	8.24	8.18	3.76
MLR	6.12	5.88	5.88	5.88
SGD	9.00	9.18	9.12	3.06
SVR	4.76	5.18	5.18	7.24

The three types of indicators, RMSE, MAE, and MAPE, should be as small as possible, and close to 0. Acc should be as large as possible, and close to 1

After the difference was determined, a "post hoc test" was needed to determine which methods have statistical differences in performance. For this purpose, we used the Bonferroni–Dunn test (Zar 2010). This test compared the proposed method, LsOA-GCN, with the other 10 methods; that is, the average ranking difference of each method was compared with the critical difference (CD). If the difference is greater than the critical difference, it means that the method with a good average ranking is statistically superior to the method with a bad average ranking; otherwise, there is no statistical difference between the two. The calculation formula of the critical difference is as follows:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{30}$$

where k is the number of methods used for comparison, N is the number of data sets, and the commonly used value of

q_α can be obtained by looking up the table. See Appendix B for the table.

The results of the Bonferroni–Dunn test are shown in Fig. 12. In order to facilitate observation and comparison, here we only used three indicators: RMSE, MAE, and MAPE. And the average rankings of 11 methods on these 3 indicators are all reflected in the bar chart. The value corresponding to the horizontal line is equivalent to the threshold, that is, the ranking of the compared method plus the value of CD. This part defined two thresholds at the significance level of 0.05 and 0.1, respectively. Each group used a different color to identify (the first group was blue, the second group was red, the third group was green), and the threshold was also distinguished by different colors.

It is worth noting that the method used for comparison (LsOA-GCN) can outperform those whose average ranking is higher than the threshold line (i.e., the height of the bar exceeds its corresponding line). It can be seen from Fig. 12 that in these three groups, LsOA-GCN is better than all methods, with an average ranking of 1.00, 1.12, and 1.12, respectively. In Group 1, LsOA-GCN is significantly better than the other 7 methods except GCN, LSTM and BPNN at the significance level of 0.05 and 0.1, respectively. In Group 2 and Group 3, LsOA-GCN is also significantly better than these 7 methods of GAT, T-GCN, GPR, KRR, MLR, SGD, and SVR at these two significant levels.

In addition to the Bonferroni–Dunn test, this study also considered Holm's method (Holm, 1979). This step: calculate the p value, sort the p value, and compare the p value with α/i. If $p < \alpha/i$, reject the null hypothesis, that is, the difference is significant, where α is the significance level and $\boldsymbol{i}$ is the method number. Here, we also only considered the three indicators: RMSE, MAE, and MAPE. The test results of these three groups are shown in Tables 10, 11 and 12, respectively. It can be seen that for these three indicators, LsOA-GCN is significantly better than other methods at the significance level of 0.05 and 0.1, respectively. The above test results once again verify the superiority of the LsOA-GCN model.

4 Conclusion

Aiming at the spreading characteristics of the COVID-19, this paper proposed a graph convolutional network prediction model based on the lioness optimization algorithm (LsOA-GCN). The LsOA-GCN model is mainly composed of two parts. One is the construction of the feature matrix for capturing temporal information, and the other is the graph convolution operation for capturing spatial information. First, the keyword data were sorted out through Spearman correlation analysis with a time lag, and the initial feature matrix was constructed to capture the time series information between the current number of confirmed cases and the Baidu Search Index at different periods in the past. Then LsOA was used to further filter the feature matrix to improve the accuracy of the model and avoid unnecessary feature interference. Finally, the feature matrix was used as input to obtain the spatial features of the epidemic-related data through the graph convolution network, to get the prediction results.

We used the cumulative confirmed case data of 17 regions in Hubei Province from January 24, 2020 to March 29, 2020 to test the effectiveness of the prediction model,

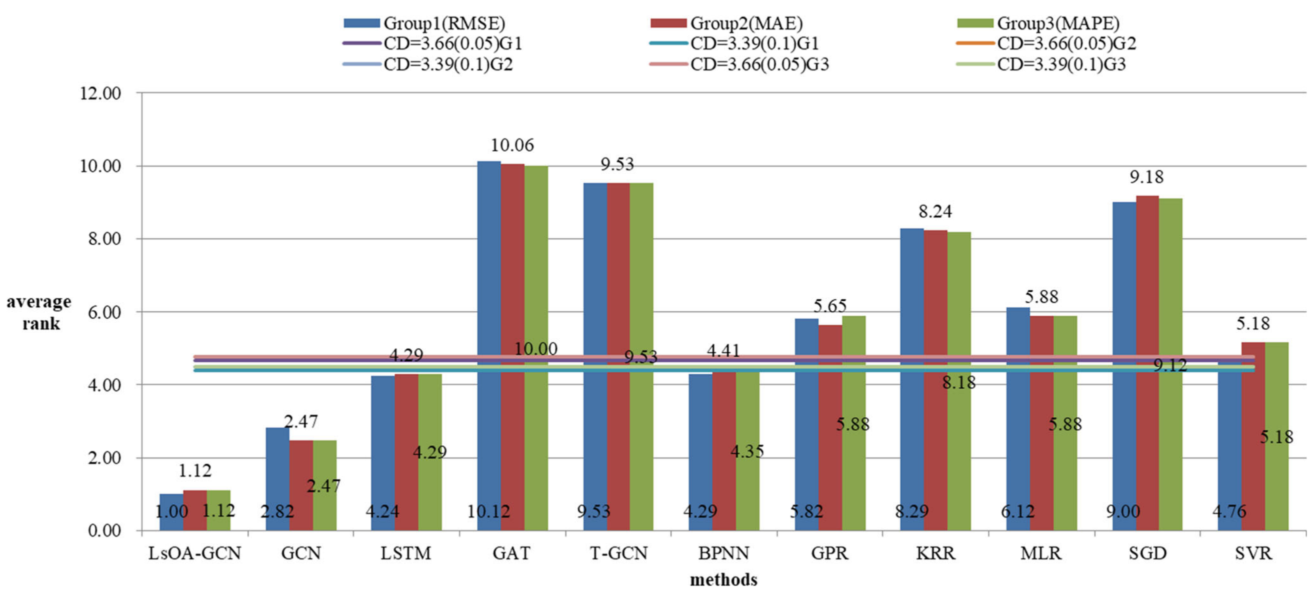


Fig. 12 Bonferroni–Dunn test for different methods and benchmark groups with α = 0.05 and α = 0.1

Table 10 Holm's method test results of the first group

LsOA-GCN vs	Rank	z-value	p-value	a/i(0.05)	a/i(0.1)
GCN	2.82	− 3.621	0.00029	0.005	0.01
LSTM	4.24	− 3.621	0.00029	0.00556	0.01111
GAT	10.12	− 3.621	0.00029	0.00625	0.0125
T-GCN	9.53	− 3.621	0.00029	0.00714	0.01429
BPNN	4.29	− 3.621	0.00029	0.00833	0.01667
GPR	5.82	− 3.621	0.00029	0.01	0.02
KRR	8.29	− 3.621	0.00029	0.0125	0.025
MLR	6.12	− 3.621	0.00029	0.01667	0.03333
SGD	9.00	− 3.621	0.00029	0.025	0.05
SVR	4.76	− 3.621	0.00029	0.05	0.1

Table 11 Holm's method test results of the second group

LsOA-GCN vs	Rank	z-value	*p*-value	a/i(0.05)	a/i(0.1)
GCN	2.47	− 3.621	0.00029	0.005	0.01
LSTM	4.29	− 3.621	0.00029	0.00556	0.01111
GAT	10.06	− 3.621	0.00029	0.00625	0.0125
T-GCN	9.53	− 3.621	0.00029	0.00714	0.01429
BPNN	4.41	− 3.621	0.00029	0.00833	0.01667
KRR	8.24	− 3.621	0.00029	0.01	0.02
SGD	9.18	− 3.621	0.00029	0.0125	0.025
SVR	5.18	− 3.621	0.00029	0.01667	0.03333
GPR	5.65	− 3.527	0.00042	0.025	0.05
MLR	5.88	− 3.527	0.00042	0.05	0.1

Table 12 Holm's method test results of the third group

LsOA-GCN vs	Rank	z-value	*p*-value	a/i(0.05)	a/i(0.1)
GCN	2.47	− 3.621	0.00029	0.005	0.01
LSTM	4.29	− 3.621	0.00029	0.00556	0.01111
GAT	10.00	− 3.621	0.00029	0.00625	0.0125
T-GCN	9.53	− 3.621	0.00029	0.00714	0.01429
BPNN	4.35	− 3.621	0.00029	0.00833	0.01667
KRR	8.18	− 3.621	0.00029	0.01	0.02
SGD	9.12	− 3.527	0.00029	0.0125	0.025
SVR	5.18	− 3.527	0.00029	0.01667	0.03333
GPR	5.88	− 3.621	0.00035	0.025	0.05
MLR	5.88	− 3.621	0.00035	0.05	0.1

and used multiple indicators to evaluate the prediction model, namely RMSE, MAE, MAPE, Accuracy, Var and R2. And these results were compared with 10 typical prediction methods, including multiple linear regression, support vector regression, backpropagation neural network, kernel ridge regression, Gaussian process regression, stochastic gradient descent regression, GAT, LSTM and T-GCN. By analyzing the experimental results, we can draw the following conclusions:

(1) Time-delay analysis can improve the accuracy of epidemic prediction. Firstly, by observing the trend graph of the features and the number of confirmed cases (Fig. 9), it can be found that the number of confirmed cases has a certain time lag relative to the search index. Then, the optimal lag days corresponding to each feature were determined by the correlation analysis (Table 1). Finally, we sorted out the feature data and found that the prediction results based on the time-delay feature data were significantly better than the original feature data (Tables 3 and 4).

(2) GCN can capture spatial information from feature data, which is helpful for prediction models to improve prediction accuracy. According to the characteristics of the epidemic transmission, this paper used the population migration index to simulate the population mobility during the epidemic, and constructed an adjacency matrix based on this. Finally, GCN was used to establish the spatial connection of various regions in Hubei Province, and to realize the prediction of the COVID-19 in various regions of Hubei Province. Through experiments, we found that compared with other prediction models, the GCN and LsOA-GCN models achieved better results in various indicators (Tables 7 and 8).

(3) There is also a significant influence between the selection of prediction features and the prediction performance of the model. Compared with the single GCN model, LsOA-GCN performed better in all indicators. It can be seen that feature selection is helpful to improve the prediction accuracy of the model.

In the future, we will consider using the LsOA-GCN model for financial market forecasting, traffic forecasting, and environmental pollution forecasting, etc., to determine that this prediction model can also have excellent performance in other fields.

Appendix A

See Tables 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28.

Table 13 Correlation analysis between keywords and the cumulative number of confirmed cases—Enshi

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	**0.3237**	0.2770	0.2437	0.2224	0.1610	0.1188	0.1063	0.1163	0.0809	0.0589	0.0351	0.0058	0.0724	0.0533	0.0522
Red Cross Society(S)	0.1399	0.0867	0.0537	0.0108	0.0334	0.0960	0.1552	0.2167	0.2609	0.3201	0.3950	0.4491	0.5355	0.6126	**0.6954**
Spring Festival(S)	0.7663	0.7712	0.7875	0.7949	0.7854	0.7765	0.7962	**0.8015**	0.7757	0.7762	0.7508	0.7548	0.7200	0.7345	0.7264
Health Commission(S)	0.2105	0.1429	0.0856	0.0532	0.0088	0.0330	0.1001	0.1813	0.2524	0.3022	0.3374	0.4149	0.4909	0.4870	**0.5003**
vaccine(S)	0.3071	0.2645	0.2108	0.1769	0.1539	0.1238	0.0634	0.0331	0.0157	0.0834	0.1550	0.1971	0.2885	0.3881	**0.3954**
SARS(S)	0.1259	0.1911	0.2494	0.3237	0.4018	0.4849	0.5659	0.6570	0.7322	0.8226	0.9207	0.9255	0.9298	0.9278	**0.9338**
hand washing(S)	0.2000	**0.2014**	0.1964	0.1444	0.1514	0.1490	0.1142	0.1191	0.1178	0.1100	0.0822	0.0296	0.0705	0.1219	0.0724
children(S)	0.6004	0.6214	0.6559	0.6562	0.7348	0.7266	0.7277	0.7332	**0.7373**	0.7185	0.7226	0.7267	0.6920	0.7366	0.7147
mask(S)	0.0208	0.0847	0.1458	0.1917	0.2589	0.3310	0.4201	0.4952	0.5788	0.6689	0.7589	0.8439	0.8597	**0.8843**	0.8800
coronavirus(S)	0.1105	0.1764	0.2462	0.3150	0.3892	0.4586	0.5426	0.6337	0.7304	0.8266	0.9257	0.9488	0.9449	0.9510	**0.9552**
novel(S)	0.1115	0.0568	0.0054	0.0367	0.1020	0.1569	0.2075	0.2778	0.3638	0.4575	0.5304	0.6209	**0.6540**	0.6105	0.5721
disinfection(S)	0.0209	0.0126	0.0572	0.0615	0.0897	0.1562	0.2167	0.2663	0.2958	0.3195	0.3514	0.3872	**0.4616**	0.4330	0.4214
lockdown of the city(S)	0.0488	0.0028	0.0574	0.0935	0.1261	0.1683	0.2369	0.2852	0.3230	0.3804	0.4288	0.4837	0.5396	**0.6162**	0.6126
Wuhan(S)	0.1004	0.1595	0.2221	0.2898	0.3520	0.4186	0.4960	0.5864	0.6698	0.7573	0.8562	0.8760	0.8795	0.8904	**0.9068**
Zhong Nanshan(S)	0.0094	0.0497	0.1132	0.1730	0.2276	0.2998	0.3869	0.4652	0.5348	0.6108	0.7044	0.7192	0.7164	0.7234	**0.7251**
epidemic(S)	0.8986	0.8896	0.8898	**0.8989**	0.8976	0.8986	0.8924	0.8557	0.8434	0.8359	0.8362	0.8127	0.7951	0.7663	0.7356
flu(S)	0.1138	0.1488	0.2081	0.2690	0.3367	0.3863	0.4329	0.5283	0.6038	0.7011	0.7992	0.8537	0.8677	0.8734	**0.8995**
diarrhea(S)	0.0489	0.0013	0.0413	0.0870	0.1532	0.2012	0.2099	0.2562	0.3260	0.3925	0.4492	0.4858	0.5615	**0.5758**	0.5673
dry cough(S)	0.0980	0.1534	0.2158	0.2839	0.3501	0.3957	0.4620	0.5332	0.6147	0.7045	0.7850	0.8179	0.8164	**0.8209**	0.8201
NCP(S)	**0.6958**	0.6789	0.6468	0.6117	0.5859	0.5689	0.5436	0.5157	0.4965	0.4673	0.4386	0.4159	0.3877	0.3441	0.2999
novel coronary pneumonia(S)	0.9213	0.9189	0.9204	**0.9220**	0.9195	0.9219	0.9153	0.9155	0.9216	0.9184	0.9210	0.9033	0.8978	0.8708	0.8508
COVID-19(S)	**0.7571**	0.7346	0.7228	0.7040	0.6897	0.6868	0.6532	0.6181	0.6106	0.5890	0.5785	0.5696	0.5419	0.5099	0.4810
novel coronavirus pneumonia(S)	0.0949	0.0396	0.0179	0.0827	0.1513	0.2250	0.3033	0.3795	0.4580	0.5335	0.6107	0.7021	0.7563	0.8184	**0.8233**
asymptomatic infection(S)	**0.5755**	0.5617	0.5457	0.5085	0.4689	0.4577	0.4375	0.4303	0.3861	0.3714	0.3486	0.3462	0.3305	0.2875	0.2335
aerosol transmission(S)	**0.3040**	0.2676	0.2470	0.2115	0.2107	0.1769	0.1350	0.0918	0.0529	0.0106	0.0082	0.0370	0.0587	0.1012	0.1440
pneumonia(S)	0.0406	0.1065	0.1721	0.2390	0.3068	0.3781	0.4587	0.5505	0.6380	0.7277	0.8220	0.9348	0.9433	0.9538	**0.9617**
cough(S)	0.1876	0.1763	0.1308	0.0739	0.0104	0.0456	0.0840	0.1221	0.1658	0.2248	0.3018	0.3541	0.4197	**0.4229**	0.3913
fever(S)	0.1696	0.1498	0.1726	0.2230	0.2727	0.3341	0.4036	0.4415	0.4922	0.5547	0.5422	0.6466	0.7075	**0.7612**	0.7349
dyspnea(S)	0.2037	0.2300	0.2489	0.2780	0.3380	0.3938	0.4485	0.4939	0.5738	0.6096	0.6578	0.7223	0.7447	**0.7568**	0.7229
high temperature(S)	0.1508	0.2116	0.2538	0.3130	0.3517	0.4087	0.4949	0.5348	0.5703	0.5842	0.6543	0.7360	0.7614	**0.7862**	0.7839
novel coronavirus(S)	0.0633	0.0044	0.0716	0.1291	0.1955	0.2654	0.3437	0.4327	0.5199	0.6039	0.6922	0.7926	0.8749	0.9216	**0.9257**
SARS-CoV-2(S)	0.3584	0.3674	**0.3767**	0.3400	0.3505	0.3212	0.3329	0.2725	0.2802	0.3066	0.3005	0.3046	0.3042	0.3107	0.2976
Spring Festival(X)	0.8245	0.8293	0.8361	0.8405	0.8498	**0.8519**	0.8234	0.8250	0.8339	0.8216	0.8063	0.7999	0.7858	0.7819	0.7882
Red Cross Society(X)	0.2608	0.1989	0.1430	0.0956	0.0515	0.0168	0.0063	0.0644	0.1592	0.2471	0.2857	0.3430	0.3891	**0.4409**	0.4342
government(X)	0.2272	0.2154	0.2048	0.1831	0.1578	0.1304	0.2128	0.2992	0.3046	0.2948	0.3355	**0.3392**	0.2931	0.3245	0.3392

Table 13 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0115	0.0852	0.0866	0.1236	0.1801	0.2616	0.3120	0.3981	0.4769	0.5501	0.6049	0.6442	0.6878	**0.7100**	0.6875
medical(X)	0.1648	0.1071	0.0632	0.0307	0.0030	0.0479	0.0750	0.1436	0.1928	0.2449	0.2803	0.3235	0.3864	0.4286	**0.4351**
vaccine(X)	0.2126	0.1872	0.1723	0.1473	0.1208	0.0751	0.0441	0.0004	0.0479	0.1020	0.1764	0.2464	0.3365	**0.3723**	0.3553
SARS(X)	0.0709	0.1464	0.2098	0.2727	0.3383	0.3876	0.4290	0.5170	0.5970	0.6744	0.7619	0.7738	**0.7785**	0.7620	0.7477
children(X)	0.5049	0.4898	0.5237	0.5064	0.5366	0.5671	0.5990	0.5832	0.5656	0.5984	0.6330	**0.6694**	0.6451	0.6197	0.5990
mask(X)	0.0837	0.0245	0.0004	0.0336	0.0738	0.0969	0.1338	0.2130	0.2787	0.3318	0.3756	0.4381	**0.5052**	0.5051	0.4861
infection(X)	0.3101	0.3351	0.3469	0.3617	0.3832	0.4057	0.4317	0.4564	0.4810	0.5068	0.5340	0.5625	**0.5926**	0.5573	0.5171
coronavirus(X)	**0.6522**	0.6120	0.5973	0.5751	0.5537	0.5444	0.5554	0.5130	0.4623	0.4347	0.4076	0.3781	0.3392	0.3345	0.2925
disinfection(X)	0.4771	0.4875	0.4984	0.5097	0.5216	0.5340	0.5469	0.5605	0.5747	0.5897	0.6054	0.6219	0.6393	0.6576	**0.6770**
nucleic acid(X)	**0.5088**	0.4744	0.4670	0.4350	0.4056	0.4147	0.4210	0.3634	0.2985	0.2519	0.2210	0.1932	0.1701	0.1317	0.0819
lockdown of the city(X)	0.5063	0.5199	0.5396	0.5512	0.5603	0.5691	0.5785	0.5908	0.5824	0.5974	0.6132	0.6299	0.6477	0.6666	**0.6867**
Wuhan(X)	0.0173	0.0265	0.0831	0.1615	0.2280	0.2789	0.3194	0.3952	0.4650	0.5405	0.6143	0.6334	0.6488	**0.6764**	0.6649
Zhong Nanshan(X)	0.1785	0.1177	0.0839	0.0544	0.0005	0.0723	0.1071	0.1709	0.2299	0.2987	0.3702	0.4614	0.5405	0.5323	**0.5463**
epidemic(X)	**0.5325**	0.4863	0.4533	0.4208	0.3815	0.3573	0.3554	0.3022	0.2364	0.1925	0.1549	0.1085	0.0595	0.0290	0.0326
flu(X)	**0.4801**	0.4305	0.4019	0.3980	0.4113	0.4049	0.3944	0.3688	0.3359	0.3030	0.2570	0.2159	0.1678	0.1401	0.1378
diarrhea(X)	0.0599	0.0558	0.0482	0.0402	0.0095	0.0023	0.0147	0.0184	0.0108	0.0029	0.0094	0.0183	0.0760	0.0829	**0.0903**
stuffy nose(X)	0.0522	0.0081	0.0445	0.0667	0.0963	0.1350	0.1517	0.1926	0.2229	0.2283	0.2685	0.3300	0.4097	0.4380	**0.5207**
dry cough(X)	0.2052	0.1602	0.1053	0.0431	0.0232	0.0863	0.0967	0.1626	0.2411	0.3265	0.4059	**0.4852**	0.4416	0.4203	0.4187
NCP(X)	**0.2270**	0.1987	0.1489	0.1337	0.1284	0.1504	0.1738	0.1137	0.0431	0.0154	0.0492	0.0360	0.0353	0.0460	0.0880
close contact(X)	**0.8081**	0.7922	0.7893	0.7727	0.7497	0.7511	0.7850	0.7556	0.7172	0.6975	0.6868	0.6793	0.6711	0.6726	0.6305
suspected case(X)	**0.5266**	0.4850	0.4632	0.4259	0.3915	0.3464	0.3534	0.3217	0.2582	0.2127	0.1856	0.1579	0.1097	0.0812	0.0250
pneumonia(X)	**0.5749**	0.5312	0.5144	0.4860	0.4571	0.4410	0.4447	0.3957	0.3382	0.3007	0.2672	0.2301	0.1826	0.1580	0.1199
cough(X)	0.3457	0.4151	0.4588	0.4968	0.4867	0.4583	0.4679	0.4962	0.5309	0.5307	0.5967	0.6719	0.7062	0.7268	**0.7364**
dyspnea(X)	0.1800	0.1277	0.0982	0.0612	0.0303	0.0334	0.0465	0.0060	0.0047	0.0002	0.0304	0.0927	0.1247	0.1635	**0.2040**
virus pneumonia(X)	0.5401	0.5466	0.5183	0.5236	0.5360	0.5790	0.5719	0.6141	0.6469	0.6569	**0.6576**	0.6337	0.6342	0.6121	0.5970
novel coronavirus(X)	**0.7072**	0.6667	0.6511	0.6290	0.6073	0.5949	0.6239	0.5868	0.5229	0.4892	0.4660	0.4417	0.4110	0.4059	0.3642

Optimal values are shown in bold in the table

Table 14 Correlation analysis between keywords and the cumulative number of confirmed cases—Ezhou

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.0168	0.0281	0.0481	0.0660	0.0792	0.1027	0.1166	0.1398	0.1826	0.2229	0.2811	0.3063	0.2780	**0.3155**	0.2801
Red Cross Society(S)	0.0167	0.0030	0.0472	0.0987	0.1421	0.1985	0.2038	0.2480	0.2961	0.3389	0.3823	0.4332	0.4984	0.5641	**0.5680**
Spring Festival(S)	0.7893	0.7941	0.8063	0.8224	**0.8229**	0.8186	0.8105	0.8139	0.7999	0.7852	0.7741	0.7636	0.7444	0.7230	0.7167
Health Commission(S)	0.1318	0.0787	0.0236	0.0205	0.0715	0.1166	0.1856	0.2420	0.2873	0.3491	0.4262	0.5028	0.5035	**0.5139**	0.4889
vaccine(S)	0.1600	0.1449	0.1154	0.0732	0.0282	0.0183	0.0469	0.1039	0.1295	0.1887	0.2539	0.2773	0.3257	0.3534	**0.3580**
SARS(S)	0.2443	0.3114	0.3677	0.4376	0.5175	0.5707	0.6497	0.7397	0.8167	**0.8964**	0.9147	0.9185	**0.9199**	0.9159	0.9017
hand washing(S)	**0.0994**	0.0860	0.0927	0.0916	0.0837	0.0806	0.0774	0.0740	0.0357	0.0236	0.0147	0.0251	0.0117	0.0024	0.0066
children(S)	0.2256	0.2237	0.2212	0.2530	**0.2707**	0.2490	0.2133	0.1946	0.2306	0.2250	0.2052	0.1723	0.1520	0.1256	0.1515
mask(S)	0.0306	0.0836	0.1487	0.1995	0.2490	0.3239	0.3955	0.4616	0.5295	0.6040	0.7031	**0.7643**	0.7457	0.7328	0.7204
coronavirus(S)	0.1599	0.2137	0.2711	0.3356	0.3967	0.4592	0.5248	0.5994	0.6810	0.7786	**0.8789**	0.8714	0.8601	0.8553	0.8434
novel(S)	0.0689	0.1003	0.1486	0.2103	0.2664	0.3126	0.3782	0.4432	0.5101	0.5675	0.6317	0.6730	**0.7424**	0.7202	0.7010
disinfection(S)	0.1245	0.1410	0.1480	0.1951	0.2198	0.2254	0.2413	0.2651	0.2860	0.3033	0.3264	0.3490	0.3747	**0.3964**	0.3469
lockdown of the city(S)	0.1007	0.0800	0.0507	0.0259	0.0140	0.0568	0.1102	0.1515	0.1825	0.2367	0.2892	0.3542	0.4216	**0.4861**	0.4611
Wuhan(S)	0.1335	0.1900	0.2554	0.3198	0.3847	0.4597	0.5432	0.6262	0.7146	0.8123	0.9072	0.9136	0.9187	**0.9210**	0.9067
Zhong Nanshan(S)	0.0960	0.1555	0.2224	0.2742	0.3294	0.4029	0.4747	0.5457	0.6185	0.7187	**0.8156**	0.8125	0.7930	0.7785	0.7593
epidemic(S)	0.9057	**0.9162**	0.9144	0.9055	0.8977	0.8986	0.8860	0.8739	0.8600	0.8620	0.8608	0.8551	0.8479	0.8332	0.8217
flu(S)	0.1302	0.1784	0.2393	0.2999	0.3669	0.4260	0.4920	0.5649	0.6457	0.7453	0.8060	0.8862	0.8963	**0.8997**	0.8946
diarrhea(S)	0.0943	0.1347	0.1780	0.2171	0.2679	0.3225	0.3813	0.4375	0.4859	0.5406	0.5976	0.6635	0.6918	**0.6941**	0.6885
dry cough(S)	0.2608	0.3084	0.3449	0.3954	0.4422	0.4994	0.5480	0.6100	0.6757	0.7368	0.7986	0.8748	**0.8788**	0.8671	0.8528
NCP(S)	**0.5195**	0.4977	0.4718	0.4630	0.4472	0.4243	0.4027	0.3809	0.3497	0.3304	0.3065	0.2803	0.2481	0.2115	0.1761
novel coronary pneumonia(S)	**0.9234**	0.9183	0.9199	0.9188	0.9172	0.9149	0.9020	0.9032	0.8952	0.8836	0.8742	0.8604	0.8473	0.8373	0.8380
COVID-19(S)	0.5771	**0.5842**	0.5753	0.5632	0.5429	0.5312	0.5344	0.5121	0.5007	0.5023	0.4951	0.4720	0.4670	0.4569	0.4362
novel coronavirus pneumonia(S)	0.2868	0.2523	0.2264	0.1878	0.1524	0.1170	0.0647	0.0098	0.0503	0.1045	0.1528	0.2164	0.2883	**0.3283**	0.3043
asymptomatic infection(S)	**0.4011**	0.3851	0.3605	0.3438	0.3217	0.2889	0.2632	0.2276	0.1960	0.1855	0.1623	0.1448	0.1309	0.1102	0.0596
aerosol transmission(S)	**0.2472**	0.2243	0.1940	0.1739	0.1465	0.1360	0.1039	0.0734	0.0580	0.0319	0.0166	0.0054	0.0273	0.0480	0.0822
pneumonia(S)	0.0894	0.1492	0.2122	0.2776	0.3417	0.4125	0.4923	0.5719	0.6609	0.7541	0.8505	0.9527	0.9642	0.9692	**0.9693**
cough(S)	0.0266	0.0124	0.0745	0.1216	0.1616	0.2136	0.2691	0.3290	0.3878	0.4671	0.5352	0.5205	0.5070	**0.5875**	0.5730
fever(S)	0.3552	0.3369	0.3733	0.4224	0.4120	0.4752	0.5236	0.5676	0.6176	0.6711	0.7261	0.7142	**0.7446**	0.7359	0.7271
dyspnea(S)	0.1875	0.2445	0.2743	0.3002	0.3475	0.3904	0.4250	0.4776	0.5238	0.5677	0.6272	0.6833	0.7172	**0.7360**	0.7277
high temperature(S)	0.2199	0.2490	0.2892	0.3294	0.3650	0.4139	0.4514	0.5242	0.5747	0.6221	0.6717	**0.7273**	0.7146	0.6983	0.6909
novel coronavirus(S)	0.0634	0.1205	0.1831	0.2471	0.3135	0.3871	0.4648	0.5465	0.6335	0.7192	0.8164	0.9220	0.9587	**0.9670**	0.9579
SARS-CoV-2(S)	0.2497	0.2479	0.2520	**0.2562**	0.2222	0.2196	0.2234	0.2272	0.1909	0.1768	0.1886	0.1993	0.1880	0.1799	0.1526
Spring Festival(X)	0.7922	0.7826	0.7875	**0.7946**	0.7939	0.7811	0.7750	0.7695	0.7832	0.7806	0.7818	0.7756	0.7620	0.7458	0.7324
Red Cross Society(X)	0.1895	0.1684	0.1338	0.0883	0.0267	0.0238	0.0673	0.1224	0.1799	0.2176	0.2746	0.3412	0.4133	**0.4703**	0.4611
government(X)	0.3444	0.3551	0.3277	0.2777	0.2421	0.2455	0.3234	0.3802	0.3860	**0.3861**	0.3791	0.3510	0.3240	0.3370	0.3272

Table 14 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.1056	0.1434	0.1680	0.2294	0.3045	0.3563	0.3885	0.4522	0.5326	0.6050	0.6884	0.7120	**0.7486**	0.7422	0.7201
medical(X)	0.1187	0.0993	0.0648	0.0265	0.0129	0.0473	0.0859	0.1387	0.1825	0.2311	0.2846	0.3439	0.4184	**0.4563**	0.4420
vaccine(X)	0.2192	0.2079	0.1917	0.1563	0.1107	0.0633	0.0168	0.0461	0.1000	0.1537	0.2208	0.2903	0.3571	**0.3764**	0.3430
SARS(X)	0.1632	0.2022	0.2590	0.3199	0.3837	0.4460	0.4964	0.5511	0.6319	0.7090	0.8018	**0.8217**	0.8148	0.8024	0.7940
children(X)	0.4266	0.4081	0.4320	0.4571	0.5035	0.5350	0.5617	0.5427	0.5760	0.6100	0.6442	**0.6806**	0.6593	0.6397	0.6763
mask(X)	0.0876	0.0624	0.0214	0.0311	0.0893	0.1361	0.1851	0.2380	0.2979	0.3538	0.4190	0.5004	**0.5895**	0.5829	0.5607
infection(X)	0.2880	0.3106	0.3191	0.3384	0.3600	0.3799	0.4007	0.4225	0.4453	0.4694	0.4947	0.5231	**0.5494**	0.5105	0.4661
coronavirus(X)	0.5795	**0.5824**	0.5651	0.5358	0.5040	0.4834	0.4643	0.4207	0.3974	0.3682	0.3322	0.2988	0.2570	0.2182	0.1828
disinfection(X)	0.4820	0.4927	0.5039	0.5156	0.5278	0.5406	0.5540	0.5680	0.5827	0.5981	0.6144	**0.6314**	0.6256	0.6137	0.6045
nucleic acid(X)	**0.5021**	0.4973	0.4749	0.4322	0.3895	0.3627	0.3365	0.2950	0.2508	0.2158	0.1757	0.1298	0.0826	0.0251	0.0462
lockdown of the city(X)	0.5894	0.6028	0.6037	0.6129	0.6260	0.6414	0.6547	0.6694	0.6872	0.7037	0.7220	0.7258	0.7259	0.7305	**0.7337**
Wuhan(X)	0.0130	0.0329	0.0836	0.1344	0.1868	0.2429	0.3178	0.3881	0.4619	0.5294	0.5943	0.5982	0.6106	**0.6455**	0.6281
Zhong Nanshan(X)	0.1143	0.0528	0.0095	0.0334	0.0920	0.1595	0.2271	0.2887	0.3528	0.4377	0.5145	0.5971	**0.6176**	0.6093	0.6019
epidemic(X)	**0.4186**	0.4153	0.3898	0.3497	0.3085	0.2796	0.2500	0.2039	0.1526	0.1153	0.0791	0.0404	0.0162	0.0725	0.1090
flu(X)	**0.4764**	0.4740	0.4652	0.4571	0.4329	0.3963	0.3596	0.3181	0.3015	0.2708	0.2365	0.1919	0.1429	0.0906	0.0930
diarrhea(X)	0.0736	0.0854	0.0934	0.1018	0.1126	0.1282	0.1227	0.1348	0.1498	0.1609	0.1725	0.1847	0.1975	0.2111	**0.2253**
stuffy nose(X)	0.1002	0.0848	0.0504	0.0092	0.0159	0.0400	0.0525	0.1051	0.1419	0.1893	0.2569	0.3110	0.3834	0.4078	**0.4873**
dry cough(X)	0.1648	0.1506	0.1132	0.0645	0.0129	0.0289	0.0630	0.1169	0.1733	0.2295	0.2940	**0.3581**	0.3198	0.3101	0.2997
NCP(X)	**0.2111**	0.2079	0.1700	0.1284	0.0982	0.0711	0.0571	0.0460	0.0380	0.0154	0.0077	0.0193	0.0351	0.0495	0.0748
close contact(X)	**0.7652**	0.7549	0.7376	0.7168	0.6961	0.6830	0.6748	0.6604	0.6414	0.6139	0.5995	0.5774	0.5559	0.5302	0.4984
suspected case(X)	**0.4923**	0.4622	0.4440	0.4081	0.3746	0.3471	0.3270	0.2931	0.2580	0.2210	0.1912	0.1454	0.0908	0.0281	0.0288
pneumonia(X)	0.4894	**0.4918**	0.4717	0.4381	0.4025	0.3789	0.3578	0.3140	0.2703	0.2383	0.2027	0.1651	0.1163	0.0708	0.0969
cough(X)	0.3218	0.3518	0.4024	0.4576	0.4575	0.4397	0.4285	0.4385	0.4650	0.4799	0.5507	0.6293	0.6794	0.6925	**0.6978**
dyspnea(X)	0.2056	0.1868	0.1551	0.1258	0.0865	0.0584	0.0001	0.0496	0.0430	0.0381	0.0970	0.1599	0.1861	0.2312	**0.2528**
virus pneumonia(X)	0.6049	0.5914	0.5974	0.6067	0.6213	0.6277	0.6076	0.6084	0.6244	0.6350	**0.6351**	0.6345	0.6146	0.6080	0.5930
novel coronavirus(X)	**0.6216**	0.6207	0.6030	0.5746	0.5430	0.5200	0.4953	0.4638	0.4382	0.4055	0.3696	0.3346	0.2927	0.2497	0.2090

Optimal values are shown in bold in the table

Table 15 Correlation analysis between keywords and the cumulative number of confirmed cases—Huanggang

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.1299	0.0954	0.0454	0.0049	0.0485	0.1074	0.1605	0.2167	0.2956	0.3570	0.4134	0.5136	0.5221	**0.5344**	0.5081
Red Cross Society(S)	0.2596	0.2178	0.1694	0.1167	0.0732	0.0214	0.0191	0.0801	0.1300	0.1864	0.2456	0.3189	0.3803	0.4625	**0.5519**
Spring Festival(S)	0.6785	0.7261	0.7618	0.8213	0.8816	0.9108	**0.9190**	0.9111	0.9082	0.9051	0.9158	0.9066	0.9080	0.9014	0.8878
Health Commission(S)	0.0180	0.0621	0.1082	0.1582	0.2167	0.2714	0.3320	0.3942	0.4809	0.5640	0.6668	0.7546	**0.7954**	0.7948	0.7760
vaccine(S)	0.1731	0.1264	0.0973	0.0433	0.0008	0.0492	0.1206	0.1754	0.2547	0.3391	0.3911	0.4331	0.5137	**0.5612**	0.5441
SARS(S)	0.1809	0.2464	0.3158	0.3883	0.4645	0.5473	0.6325	0.7258	0.8232	0.9240	0.9747	0.9812	0.9820	0.9824	**0.9837**
hand washing(S)	0.1512	0.1119	0.0750	0.1122	0.0799	0.0377	0.0054	0.0353	0.0606	0.0820	0.0970	0.1258	0.1400	0.1646	**0.1977**
children(S)	0.6428	0.6386	0.6288	0.6213	0.6516	0.6458	0.6562	**0.6630**	0.6500	0.6414	0.6394	0.6099	0.5806	0.5786	0.5428
mask(S)	0.0872	0.1443	0.1903	0.2445	0.2955	0.3601	0.4270	0.5070	0.5915	0.6807	0.7826	**0.8475**	0.8414	0.8333	0.8224
coronavirus(S)	0.1362	0.1987	0.2619	0.3319	0.4069	0.4853	0.5696	0.6577	0.7510	0.8520	0.9564	0.9772	0.9746	**0.9777**	0.9766
novel(S)	0.0603	0.1217	0.1779	0.2338	0.2939	0.3535	0.4180	0.4857	0.5752	0.6535	0.7648	**0.7929**	0.7881	0.7891	0.7664
disinfection(S)	0.1311	0.1056	0.0734	0.0296	0.0154	0.0772	0.1127	0.1548	0.2260	0.2679	0.3286	0.3883	0.3892	0.4284	**0.4615**
lockdown of the city(S)	0.1662	0.1294	0.0764	0.0433	0.0100	0.0486	0.1124	0.1706	0.2289	0.2840	0.3647	0.4456	0.5277	**0.6096**	0.5621
Wuhan(S)	0.1277	0.1865	0.2481	0.3161	0.3878	0.4609	0.5382	0.6240	0.7222	0.8185	0.9331	0.9556	0.9665	**0.9679**	0.9659
Zhong Nanshan(S)	0.1051	0.1604	0.2313	0.2970	0.3687	0.4407	0.5174	0.6047	0.6856	0.7786	0.8690	0.9014	0.9000	**0.9090**	0.9073
epidemic(S)	0.8731	0.8730	0.8784	**0.8812**	0.8699	0.8601	0.8397	0.8231	0.7989	0.7857	0.7595	0.7482	0.7318	0.7054	0.6863
flu(S)	0.0954	0.1571	0.2088	0.2684	0.3370	0.4077	0.4794	0.5715	0.6641	0.7642	0.8627	0.9083	**0.9299**	0.9291	0.9295
diarrhea(S)	0.1043	0.0621	0.0272	0.0146	0.0512	0.1152	0.1578	0.2364	0.2895	0.3523	0.4278	0.4842	0.5429	0.6056	**0.6181**
dry cough(S)	0.1227	0.1683	0.2327	0.2983	0.3576	0.4284	0.5065	0.5858	0.6761	0.7646	0.8486	0.8732	**0.8758**	0.8677	0.8567
NCP(S)	**0.6990**	0.6771	0.6584	0.6313	0.6221	0.5939	0.5733	0.5477	0.5145	0.4840	0.4486	0.4231	0.3848	0.3564	0.3169
novel coronary pneumonia(S)	0.9133	0.9179	0.9162	0.9179	0.9221	**0.9289**	0.9213	0.9221	0.9066	0.9021	0.8911	0.8872	0.8803	0.8678	0.8654
COVID-19(S)	**0.7744**	0.7644	0.7677	0.7656	0.7652	0.7571	0.7348	0.7157	0.6901	0.6690	0.6484	0.6449	0.6219	0.6082	0.5902
novel coronavirus pneumonia(S)	0.2566	0.2148	0.1735	0.1280	0.0813	0.0244	0.0254	0.0907	0.1547	0.2183	0.3008	0.3813	0.4620	**0.5396**	0.5234
asymptomatic infection(S)	**0.6947**	0.6629	0.6477	0.6321	0.6175	0.5998	0.5832	0.5559	0.5269	0.4890	0.4643	0.4369	0.4251	0.4041	0.3660
aerosol transmission(S)	**0.3725**	0.3480	0.3278	0.2961	0.2589	0.2183	0.1813	0.1464	0.1240	0.0963	0.0786	0.0548	0.0258	0.0126	0.0639
pneumonia(S)	0.0720	0.1308	0.1912	0.2540	0.3236	0.3962	0.4763	0.5579	0.6524	0.7399	0.8475	0.9512	0.9648	0.9684	**0.9687**
cough(S)	0.1527	0.2004	0.2663	0.3387	0.3256	0.3959	0.3778	0.4444	0.4285	0.5015	0.5727	0.6644	**0.7062**	0.6920	0.6875
fever(S)	0.2157	0.2431	0.2679	0.3293	0.3754	0.4255	0.4871	0.5344	0.6138	0.6984	0.7416	0.7570	**0.7574**	0.7534	0.7395
dyspnea(S)	0.0021	0.0340	0.0817	0.1446	0.2000	0.2386	0.3138	0.3638	0.4328	0.4920	0.5583	0.6260	0.7151	0.7284	**0.7344**
high temperature(S)	0.0606	0.1303	0.1753	0.2349	0.2925	0.3601	0.4133	0.4957	0.5672	0.6590	0.7498	**0.7823**	0.7743	0.7727	0.7759
novel coronavirus(S)	0.0249	0.0830	0.1417	0.2055	0.2750	0.3454	0.4239	0.5059	0.5945	0.6851	0.7846	0.8823	0.9271	0.9590	**0.9601**
SARS-CoV-2(S)	**0.4572**	0.4306	0.4159	0.3922	0.3688	0.3557	0.3519	0.3229	0.3240	0.3196	0.3158	0.3196	0.2895	0.2650	0.2219
Spring Festival(X)	0.7969	0.7969	0.8055	0.8007	**0.8126**	0.8122	0.8115	0.8051	0.8005	0.7885	0.8006	0.8066	0.8053	0.8008	0.7877
Red Cross Society(X)	0.2368	0.2040	0.1686	0.1351	0.0941	0.0401	0.0298	0.0947	0.1661	0.2298	0.2951	0.3587	0.4269	**0.4961**	0.4842
government(X)	0.2891	0.2645	0.2154	0.2003	0.1912	0.1644	0.2159	0.2857	0.3286	0.3331	**0.3332**	0.3058	0.2879	0.3132	0.3072

Table 15 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0475	0.1061	0.1355	0.2039	0.2669	0.3367	0.3885	0.4571	0.5400	0.6073	0.7026	0.7062	0.7526	**0.7566**	0.7502
medical(X)	0.1034	0.0720	0.0334	0.0127	0.0269	0.0654	0.1106	0.1602	0.2056	0.2571	0.3092	0.3801	0.4559	0.4938	**0.4972**
vaccine(X)	0.1989	0.1813	0.1405	0.1274	0.0977	0.0521	0.0020	0.0682	0.1303	0.2020	0.2738	0.3516	0.4167	**0.4395**	0.4200
SARS(X)	0.1688	0.2311	0.2930	0.3389	0.4096	0.4682	0.5344	0.6007	0.6835	0.7711	0.8696	0.8898	**0.8957**	0.8931	0.8809
children(X)	0.4165	0.4024	0.4315	0.4660	0.5125	0.5488	0.5318	0.5075	0.5590	0.5325	0.5251	**0.5615**	0.5271	0.5123	0.4754
mask(X)	0.0702	0.0503	0.0185	0.0035	0.0562	0.0980	0.1605	0.2131	0.2760	0.3227	0.3902	0.4453	**0.5135**	0.4962	0.4810
infection(X)	0.3208	0.3333	0.3593	0.3757	0.3993	0.4219	0.4455	0.4702	0.4962	0.5235	0.5523	0.5826	**0.6145**	0.5830	0.5473
coronavirus(X)	**0.5996**	0.5813	0.5509	0.5561	0.5331	0.5081	0.4760	0.4396	0.4042	0.3691	0.3367	0.2892	0.2560	0.2161	0.1638
disinfection(X)	0.5800	0.5931	0.6069	0.6212	0.6362	0.6519	0.6684	0.6857	0.7038	0.7229	0.7429	0.7640	0.7863	0.7936	**0.7971**
nucleic acid(X)	**0.4791**	0.4717	0.4594	0.4541	0.4318	0.3939	0.3475	0.2981	0.2708	0.2385	0.2048	0.1654	0.1246	0.0692	0.0108
lockdown of the city(X)	0.5486	0.5642	0.5704	0.5630	0.5759	0.5835	0.5947	0.6119	0.6260	0.6423	0.6596	0.6779	0.6906	**0.6909**	0.6902
Wuhan(X)	0.0333	0.0805	0.1408	0.1767	0.2323	0.2732	0.3351	0.3908	0.4743	0.5504	0.6270	0.6282	0.6292	**0.6740**	0.6435
Zhong Nanshan(X)	0.0531	0.0148	0.0431	0.0725	0.1320	0.1840	0.2551	0.3346	0.4127	0.4900	0.5669	0.6727	0.6831	**0.6857**	0.6690
epidemic(X)	0.2732	0.2378	0.1891	0.1734	0.1288	0.0774	0.0234	0.0363	0.0978	0.1558	0.2217	0.2982	0.3644	0.4401	**0.5141**
flu(X)	**0.4246**	0.4149	0.4029	0.4064	0.3937	0.3861	0.3532	0.3059	0.2524	0.2139	0.1805	0.1417	0.1148	0.0692	0.0888
diarrhea(X)	0.0514	0.0572	0.0141	0.0116	0.0215	0.0270	0.0348	0.0435	0.0523	0.0900	0.1296	0.1639	0.1900	0.2072	**0.2164**
stuffy nose(X)	0.0195	0.0119	0.0506	0.0624	0.0849	0.1183	0.1460	0.1921	0.2289	0.2737	0.3284	0.3901	0.4408	0.4972	**0.5336**
dry cough(X)	0.1687	0.1254	0.0693	0.0446	0.0087	0.0554	0.1168	0.1740	0.2355	0.2956	0.3600	**0.4381**	0.4052	0.4140	0.4124
NCP(X)	**0.2803**	0.2563	0.2533	0.2459	0.2137	0.1814	0.1409	0.1001	0.0814	0.0626	0.0508	0.0455	0.0382	0.0038	0.0214
close contact(X)	**0.7574**	0.7460	0.7422	0.7525	0.7441	0.7286	0.7079	0.6857	0.6829	0.6585	0.6421	0.6106	0.6054	0.5789	0.5573
suspected case(X)	**0.4694**	0.4463	0.4112	0.4005	0.3815	0.3544	0.3241	0.2889	0.2624	0.2236	0.1875	0.1390	0.1086	0.0584	0.0026
pneumonia(X)	**0.4100**	0.3842	0.3406	0.3334	0.2960	0.2670	0.2245	0.1800	0.1292	0.0878	0.0346	0.0240	0.0735	0.1326	0.1259
cough(X)	0.3545	0.4022	0.4695	0.5374	0.5329	0.5268	0.5256	0.5437	0.5637	0.5688	0.6292	0.7026	0.7240	0.7323	**0.7454**
dyspnea(X)	0.1702	0.1454	0.1197	0.1010	0.0727	0.0413	0.0006	0.0489	0.0668	0.0792	0.1311	0.1928	0.2526	**0.3004**	0.2941
virus pneumonia(X)	0.5668	0.5621	0.5693	0.5604	0.6003	0.6119	0.6208	0.6166	0.6307	**0.6415**	0.6266	0.6103	0.5946	0.6003	0.5722
novel coronavirus(X)	**0.6409**	0.6226	0.5955	0.5984	0.5748	0.5502	0.5184	0.4839	0.4516	0.4194	0.3892	0.3443	0.3116	0.2689	0.2155

Optimal values are shown in bold in the table

Table 16 Correlation analysis between keywords and the cumulative number of confirmed cases—Huangshi

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.1050	0.0698	0.0516	0.0218	0.0380	0.0991	0.1204	0.1440	0.2080	0.2512	0.2784	0.3288	**0.3733**	0.3632	0.3105
Red Cross Society(S)	0.0065	0.0364	0.0821	0.1100	0.1501	0.2024	0.2651	0.3214	0.3983	0.4324	0.4888	0.5699	0.6369	0.7151	**0.7339**
Spring Festival(S)	0.7827	0.8068	0.8202	0.8376	0.8369	**0.8392**	0.8288	0.8265	0.8051	0.7924	0.7918	0.7897	0.7876	0.7848	0.7710
Health Commission(S)	0.0196	0.0324	0.0818	0.1407	0.1935	0.2617	0.3099	0.3767	0.4584	0.5435	0.6025	0.6890	0.6943	**0.7102**	0.6880
vaccine(S)	0.1915	0.1630	0.1294	0.0875	0.0210	0.0290	0.0820	0.1428	0.2031	0.2546	0.3326	0.3919	0.4197	0.4190	**0.4628**
SARS(S)	0.2036	0.2694	0.3389	0.4108	0.4692	0.5462	0.6278	0.7170	0.8034	0.8970	**0.9610**	0.9560	0.9568	0.9505	0.9535
hand washing(S)	0.1611	0.1145	0.1215	0.0985	0.0785	0.0296	0.0062	0.0288	0.0234	0.0478	0.0550	0.0639	0.0891	0.1496	**0.1637**
children(S)	0.3862	0.4047	0.4128	0.3945	0.4181	0.4983	0.5005	0.4799	0.4552	0.4913	0.4734	**0.5033**	0.4750	0.4986	0.4915
mask(S)	0.0760	0.1313	0.1752	0.2438	0.3001	0.3775	0.4520	0.5470	0.6295	0.7093	0.8024	**0.8747**	0.8641	0.8562	0.8438
coronavirus(S)	0.1506	0.2111	0.2771	0.3439	0.4187	0.4975	0.5782	0.6670	0.7628	0.8627	0.9660	**0.9719**	0.9644	0.9646	0.9609
novel(S)	0.1032	0.1577	0.2004	0.2559	0.3174	0.3687	0.4263	0.5087	0.5671	0.6442	0.7174	**0.7394**	0.7281	0.7242	0.7235
disinfection(S)	0.0119	0.0083	0.0213	0.0707	0.0938	0.1145	0.1607	0.2000	0.2510	0.2628	0.3130	0.3081	0.3075	**0.3651**	0.3410
lockdown of the city(S)	0.1228	0.0713	0.0317	0.0226	0.0816	0.1358	0.1996	0.2588	0.3243	0.4051	0.4628	0.5399	0.6240	**0.7075**	0.7024
Wuhan(S)	0.1394	0.2027	0.2652	0.3299	0.3949	0.4702	0.5535	0.6475	0.7399	0.8365	0.9349	0.9468	0.9472	**0.9509**	0.9452
Zhong Nanshan(S)	0.1076	0.1783	0.2411	0.3021	0.3755	0.4587	0.5370	0.6162	0.6998	0.8057	0.8991	0.9027	0.9048	**0.9104**	0.8982
epidemic(S)	**0.8768**	0.8693	0.8574	0.8479	0.8417	0.8407	0.8132	0.7871	0.7671	0.7603	0.7592	0.7553	0.7474	0.7132	0.6743
flu(S)	0.1593	0.2087	0.2680	0.2935	0.3383	0.4015	0.4607	0.5370	0.6186	0.7110	0.8054	0.8928	0.9001	0.9017	**0.9045**
diarrhea(S)	0.0300	0.0063	0.0141	0.0235	0.0765	0.1335	0.1912	0.2403	0.3088	0.3691	0.4351	0.4958	0.5736	0.6335	**0.6486**
dry cough(S)	0.1627	0.2100	0.2596	0.3245	0.3916	0.4689	0.5292	0.5979	0.6672	0.7591	**0.8358**	0.8288	0.8339	0.8222	0.7951
NCP(S)	**0.5531**	0.5277	0.5011	0.4619	0.4441	0.4323	0.4144	0.4000	0.3882	0.3584	0.3246	0.2874	0.2620	0.2373	0.1972
novel coronary pneumonia(S)	**0.9093**	0.9080	0.8967	0.8803	0.8744	0.8670	0.8674	0.8504	0.8444	0.8437	0.8347	0.8240	0.7999	0.7974	0.7745
COVID-19(S)	**0.7714**	0.7529	0.7393	0.7245	0.7033	0.6874	0.6537	0.6289	0.6095	0.6044	0.5954	0.5913	0.5816	0.5518	0.5251
novel coronavirus pneumonia(S)	**0.3518**	0.3295	0.3013	0.2620	0.2327	0.1879	0.1388	0.0919	0.0338	0.0160	0.0668	0.1286	0.1966	0.2620	0.2841
asymptomatic infection(S)	**0.5724**	0.5702	0.5445	0.5231	0.4941	0.4688	0.4571	0.4349	0.3847	0.3482	0.3033	0.2866	0.2598	0.2479	0.1993
aerosol transmission(S)	**0.3396**	0.3047	0.2708	0.2390	0.2396	0.2150	0.1871	0.1608	0.1385	0.1187	0.0808	0.0398	0.0062	0.0229	0.0611
pneumonia(S)	0.0789	0.1365	0.1990	0.2651	0.3324	0.4023	0.4835	0.5684	0.6607	0.7489	0.8479	0.9535	**0.9564**	0.9500	0.9556
cough(S)	0.0893	0.0558	0.0249	0.0379	0.0614	0.1119	0.0956	0.1589	0.2217	0.2648	0.3577	0.3989	0.4010	0.4099	**0.4158**
fever(S)	0.2737	0.2749	0.2804	0.3076	0.2819	0.3567	0.4153	0.4905	0.5741	0.5789	0.6366	0.6718	0.6836	**0.6958**	0.6729
dyspnea(S)	0.0262	0.0717	0.1125	0.1621	0.2093	0.2648	0.3212	0.3786	0.4362	0.4617	0.5148	0.5624	**0.6401**	0.6334	0.6199
high temperature(S)	0.0293	0.0727	0.1247	0.1836	0.2322	0.2534	0.3267	0.3838	0.4431	0.4445	0.5001	0.5818	0.6331	0.6380	**0.6381**
novel coronavirus(S)	0.0132	0.0737	0.1341	0.1928	0.2583	0.3276	0.4000	0.4859	0.5728	0.6570	0.7504	0.8464	0.9029	**0.9337**	0.9325
SARS-CoV-2(S)	**0.3035**	0.2838	0.2455	0.2480	0.2490	0.2463	0.2397	0.2241	0.1939	0.2028	0.1934	0.1753	0.1721	0.1642	0.1298
Spring Festival(X)	0.7885	0.7817	0.7860	**0.8009**	0.7955	0.7798	0.7703	0.7648	0.7765	0.7786	0.7742	0.7778	0.7564	0.7337	0.7301
Red Cross Society(X)	0.2044	0.1803	0.1515	0.0971	0.0437	0.0185	0.0290	0.0919	0.1485	0.1847	0.2154	0.2615	0.3312	0.3942	**0.3964**
government(X)	0.3155	0.3131	0.3021	0.2509	0.2212	0.2307	0.3161	**0.3770**	0.3687	0.3644	0.3702	0.3198	0.2840	0.2999	0.2849

Table 16 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0947	0.1324	0.1436	0.2079	0.2703	0.3168	0.3348	0.4052	0.4935	0.5546	0.6236	0.6488	**0.6806**	0.6583	0.6491
medical(X)	0.1997	0.1636	0.1388	0.0944	0.0727	0.0521	0.0227	0.0398	0.1215	0.1826	0.2293	0.2758	0.3184	0.3580	**0.3718**
vaccine(X)	0.2159	0.2053	0.1935	0.1598	0.1115	0.0578	0.0035	0.0757	0.1307	0.1828	0.2489	0.3176	0.3958	**0.4201**	0.4033
SARS(X)	0.1453	0.1842	0.2370	0.2996	0.3518	0.4050	0.4471	0.5089	0.6002	0.6833	0.7692	**0.7924**	0.7742	0.7430	0.7408
children(X)	0.3785	0.3583	0.3871	0.4194	0.4642	0.4968	0.5206	0.5489	0.5836	0.6137	0.6566	**0.7168**	0.6970	0.6752	0.6535
mask(X)	0.0951	0.0681	0.0350	0.0108	0.0519	0.0781	0.1314	0.1960	0.2522	0.2926	0.3525	0.3980	**0.4709**	0.4593	0.4494
infection(X)	0.2861	0.3080	0.3300	0.3611	0.3861	0.4075	0.4298	0.4532	0.4778	0.5037	0.5309	0.5596	**0.5898**	0.5560	0.5175
coronavirus(X)	0.5883	**0.5892**	0.5784	0.5474	0.5263	0.5340	0.5127	0.4663	0.4398	0.4158	0.3742	0.3586	0.3238	0.2998	0.2581
disinfection(X)	0.5571	0.5697	0.5829	0.5967	0.6111	0.6262	0.6420	0.6586	0.6759	0.6942	0.7134	**0.7138**	0.7114	0.7058	0.7040
nucleic acid(X)	0.5163	**0.5255**	0.5032	0.4502	0.4114	0.3911	0.3762	0.3621	0.3238	0.3053	0.2769	0.2264	0.1921	0.1548	0.0684
lockdown of the city(X)	0.5565	0.5736	0.6005	0.6125	0.6278	0.6432	0.6506	0.6633	0.6800	0.6975	0.7161	0.7174	0.7185	0.7252	**0.7298**
Wuhan(X)	0.0349	0.0462	0.0978	0.1588	0.1982	0.2442	0.3123	0.3727	0.4481	0.5298	0.5941	0.6041	0.5865	**0.6104**	0.5893
Zhong Nanshan(X)	0.1258	0.0668	0.0318	0.0189	0.0717	0.1192	0.1664	0.2356	0.3158	0.3904	0.4558	0.5381	**0.5555**	0.5536	0.5494
epidemic(X)	**0.4479**	0.4437	0.4247	0.3772	0.3460	0.3442	0.3081	0.2560	0.2155	0.1778	0.1346	0.1120	0.0624	0.0215	0.0311
flu(X)	**0.4803**	0.4692	0.4558	0.4392	0.4269	0.3997	0.3838	0.3497	0.3127	0.2720	0.2559	0.2160	0.1826	0.1357	0.1348
diarrhea(X)	0.0052	0.0085	0.0040	0.0201	0.0516	0.0767	0.0699	0.0543	0.0944	0.0914	0.0986	0.1274	0.1589	0.1701	**0.1821**
stuffy nose(X)	0.0810	0.0923	0.0965	0.1226	0.1563	0.1669	0.1735	0.2269	0.2909	0.3016	0.3746	0.4035	0.4337	0.4968	**0.5396**
dry cough(X)	0.1775	0.1307	0.0917	0.0326	0.0160	0.0638	0.1118	0.1628	0.2205	0.2886	0.3652	**0.4444**	0.4082	0.4106	0.4069
NCP(X)	**0.2382**	0.2340	0.1878	0.1373	0.1040	0.0786	0.0700	0.0729	0.0674	0.0486	0.0461	0.0076	0.0128	0.0259	0.0623
close contact(X)	0.7599	**0.7777**	0.7594	0.7388	0.7126	0.6978	0.6968	0.6808	0.6562	0.6407	0.6399	0.6108	0.6045	0.5983	0.5654
suspected case(X)	**0.5137**	0.4917	0.4793	0.4433	0.4115	0.4019	0.3895	0.3434	0.2917	0.2642	0.2341	0.1954	0.1629	0.1103	0.0660
pneumonia(X)	**0.5000**	0.4971	0.4853	0.4444	0.4160	0.3974	0.3628	0.3186	0.2873	0.2532	0.2174	0.1797	0.1234	0.0858	0.0576
cough(X)	0.2899	0.3306	0.3621	0.4054	0.3997	0.3818	0.3602	0.3637	0.3959	0.4062	0.4586	0.5272	0.5586	0.5942	**0.6390**
dyspnea(X)	**0.2590**	0.2403	0.2172	0.1885	0.1543	0.1334	0.0921	0.0380	0.0178	0.0195	0.0318	0.0690	0.0881	0.1533	0.2217
virus pneumonia(X)	0.5645	0.5506	0.5589	0.5755	0.5945	0.5891	0.5803	0.5953	0.6165	0.6211	0.6236	**0.6251**	0.6008	0.5928	0.5962
novel coronavirus(X)	0.6364	**0.6372**	0.6233	0.5971	0.5771	0.5843	0.5797	0.5380	0.4991	0.4682	0.4321	0.4045	0.3816	0.3642	0.3223

Optimal values are shown in bold in the table

Table 17 Correlation analysis between keywords and the cumulative number of confirmed cases—Jingmen

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.0810	0.0358	0.0089	0.0491	0.1052	0.1376	0.1130	0.1705	0.2241	0.3044	0.2701	0.2800	0.2884	**0.3668**	0.3334
Red Cross Society(S)	0.1650	0.1177	0.1049	0.0945	0.0542	0.0003	0.0569	0.1281	0.1959	0.2509	0.3037	0.3899	0.4712	0.5323	**0.5439**
Spring Festival(S)	0.7382	0.7275	0.7654	**0.8046**	0.8019	0.7789	0.7696	0.7724	0.7589	0.7389	0.7104	0.6809	0.6666	0.6895	0.6366
Health Commission(S)	0.1311	0.1009	0.0632	0.0205	0.0377	0.0909	0.1465	0.2017	0.2813	0.3214	0.3711	0.4637	**0.5415**	0.5178	0.5329
vaccine(S)	**0.2932**	0.2736	0.2492	0.2325	0.2025	0.1878	0.1097	0.0307	0.0091	0.0569	0.1352	0.1809	0.2012	0.2612	0.2582
SARS(S)	0.1871	0.2559	0.3288	0.4054	0.4788	0.5545	0.6396	0.7365	0.8353	0.9017	0.9514	0.9596	0.9628	**0.9656**	0.9615
hand washing(S)	0.2357	0.2122	**0.2425**	0.1912	0.1974	0.1933	0.1593	0.1329	0.1795	0.1160	0.0308	0.0642	0.0402	0.0118	0.0606
children(S)	0.3497	0.3204	0.3180	0.3451	0.3616	**0.3814**	0.3249	0.3392	0.3584	0.2760	0.2102	0.2235	0.1602	0.1370	0.1482
mask(S)	0.0157	0.0576	0.1102	0.1731	0.2290	0.2733	0.3462	0.4260	0.5165	0.5985	0.6437	0.7257	0.7141	**0.7448**	0.7107
coronavirus(S)	0.1316	0.1967	0.2619	0.3364	0.4064	0.4802	0.5631	0.6448	0.7343	0.8352	0.9359	0.9301	0.9332	0.9372	**0.9387**
novel(S)	0.0643	0.1035	0.1700	0.2249	0.2467	0.2968	0.3791	0.4232	0.4920	0.5905	0.6373	0.6925	**0.7132**	0.6829	0.6515
disinfection(S)	0.0267	0.0598	0.1195	0.1272	0.1324	0.1866	0.2268	0.2648	0.3308	0.3565	0.4354	**0.4543**	0.4144	0.3925	0.3438
lockdown of the city(S)	0.1184	0.0638	0.0219	0.0065	0.0415	0.0686	0.1293	0.2022	0.2368	0.2851	0.3552	0.4378	0.5131	0.5884	**0.5897**
Wuhan(S)	0.1145	0.1754	0.2310	0.2951	0.3579	0.4270	0.5020	0.5916	0.6797	0.7651	0.8683	**0.8892**	0.8752	0.8787	0.8744
Zhong Nanshan(S)	0.0275	0.0873	0.1535	0.2215	0.2857	0.3653	0.4612	0.5674	0.6452	0.7228	0.8264	0.8655	0.8565	**0.8757**	0.8441
epidemic(S)	**0.8886**	0.8803	0.8830	0.8848	0.8770	0.8616	0.8365	0.8133	0.7888	0.7946	0.7706	0.7761	0.7630	0.7556	0.7262
flu(S)	0.1440	0.2040	0.2482	0.3166	0.3640	0.4169	0.4777	0.5638	0.6316	0.7294	0.8025	0.8532	0.8798	**0.9027**	0.8995
diarrhea(S)	0.0095	0.0366	0.0880	0.1438	0.1987	0.2602	0.3242	0.3965	0.4554	0.5295	0.6070	0.6827	0.7622	0.7599	**0.7805**
dry cough(S)	0.0475	0.1132	0.1762	0.2296	0.2913	0.3512	0.4507	0.5315	0.5945	0.6815	0.7701	**0.8845**	0.8529	0.8690	0.8339
NCP(S)	**0.5602**	0.5335	0.4999	0.4686	0.4349	0.3970	0.3856	0.3671	0.3679	0.3469	0.3736	0.3469	0.3208	0.2892	0.2752
novel coronary pneumonia(S)	**0.9293**	0.9253	0.9214	0.9200	0.9221	0.9145	0.9080	0.8965	0.8941	0.8729	0.8621	0.8589	0.8355	0.8129	0.8065
COVID-19(S)	**0.7525**	0.7477	0.7453	0.7447	0.7357	0.7446	0.7227	0.6882	0.6587	0.6383	0.6573	0.6573	0.6604	0.6745	0.7178
novel coronavirus pneumonia(S)	0.3966	**0.4014**	0.3725	0.3347	0.3004	0.2650	0.2268	0.1832	0.1313	0.0750	0.0234	0.0484	0.1160	0.1900	0.2440
asymptomatic infection(S)	**0.6621**	0.6553	0.6367	0.6187	0.6116	0.5851	0.5800	0.5646	0.5190	0.4554	0.4513	0.4518	0.4146	0.3317	0.3089
aerosol transmission(S)	**0.2966**	0.2739	0.2485	0.2154	0.1844	0.1810	0.1613	0.1474	0.1423	0.1224	0.0874	0.0698	0.0275	0.0267	0.0666
pneumonia(S)	0.0566	0.1172	0.1784	0.2355	0.3062	0.3725	0.4447	0.5226	0.6232	0.6980	0.8009	0.8907	**0.8973**	0.8899	0.8921
cough(S)	0.0288	0.0422	0.0759	0.1724	0.2275	0.2652	0.2948	0.3336	0.3529	0.4516	0.5537	0.6096	0.5939	0.6329	**0.6878**
fever(S)	0.0961	0.1462	0.2359	0.3018	0.2319	0.2647	0.3179	0.3446	0.3973	0.4488	0.5008	0.4937	0.5569	**0.5764**	0.5628
dyspnea(S)	0.1235	0.1993	0.2358	0.3096	0.3510	0.4229	0.5010	0.5844	0.6345	0.6603	0.7625	**0.8059**	0.7901	0.7568	0.7257
high temperature(S)	0.0656	0.1166	0.1429	0.2237	0.2363	0.2972	0.3021	0.3462	0.4521	0.5069	0.5208	0.5407	**0.5623**	0.5131	0.4440
novel coronavirus(S)	0.0022	0.0642	0.1322	0.2029	0.2670	0.3345	0.4100	0.4854	0.5657	0.6460	0.7528	0.8365	0.8835	0.9373	**0.9437**
SARS-CoV-2(S)	**0.4834**	0.4653	0.4426	0.4456	0.4537	0.4362	0.4094	0.3695	0.3318	0.3167	0.3222	0.3194	0.3043	0.3056	0.3065
Spring Festival(X)	0.7888	0.7854	0.8104	0.7917	0.8026	0.7954	**0.8368**	0.7823	0.7684	0.7583	0.7725	0.7659	0.7645	0.7450	0.7423
Red Cross Society(X)	0.1735	0.1511	0.0979	0.0782	0.0299	0.0194	0.1228	0.1366	0.2173	0.2755	0.3497	0.3795	0.4365	0.4897	**0.5266**
government(X)	0.3107	0.3080	0.2909	0.2507	0.2623	0.2502	0.2584	0.3645	0.4012	0.3943	**0.4150**	0.3691	0.3129	0.3570	0.3409

Table 17 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0136	0.0697	0.1018	0.1709	0.2393	0.2896	0.3815	0.4175	0.4961	0.5242	0.6572	0.6514	**0.7098**	0.6788	0.6926
medical(X)	0.1145	0.0985	0.0683	0.0392	0.0002	0.0446	0.0914	0.1336	0.1892	0.2439	0.2852	0.3114	0.3861	0.3841	**0.4236**
vaccine(X)	0.2461	0.2246	0.1783	0.1603	0.1379	0.1216	0.0192	0.0067	0.0469	0.1188	0.2043	0.2519	0.3214	**0.3708**	0.3236
SARS(X)	0.1225	0.1809	0.2521	0.2930	0.3525	0.3821	0.4781	0.4998	0.5654	0.6637	0.7680	0.7710	0.7613	**0.7752**	0.7045
children(X)	0.3833	0.3928	0.3889	0.3563	0.3923	0.4242	0.5024	0.5102	0.5115	0.5433	0.5847	**0.6170**	0.5945	0.5643	0.5341
mask(X)	0.1209	0.1006	0.0709	0.0586	0.0145	0.0189	0.1158	0.1324	0.1776	0.2400	0.3402	0.3674	0.3981	**0.4192**	0.4120
infection(X)	0.3281	0.2940	0.3182	0.3515	0.4023	0.3871	0.4480	0.4743	0.4691	0.5319	0.5684	0.5990	**0.6233**	0.5984	0.5644
coronavirus(X)	**0.5842**	0.5546	0.5177	0.5162	0.5007	0.4896	0.4552	0.4210	0.4187	0.3904	0.3507	0.3117	0.2821	0.2641	0.2169
disinfection(X)	0.4721	0.4836	0.4929	0.5017	0.5035	0.5150	0.5357	0.5474	0.5750	0.5873	0.6023	0.6171	**0.6671**	0.6458	0.6380
nucleic acid(X)	0.4991	**0.5058**	0.4854	0.4643	0.4642	0.4239	0.3343	0.3388	0.3492	0.3106	0.2941	0.2837	0.2660	0.2073	0.1208
lockdown of the city(X)	0.5523	0.5790	0.5948	0.6359	0.6184	0.6333	0.6184	0.6581	0.6678	0.6555	0.6656	0.6882	**0.7089**	0.7033	0.6917
Wuhan(X)	0.0367	0.0840	0.1329	0.1601	0.2144	0.2720	0.3413	0.3951	0.4990	0.5969	0.6592	0.6492	0.6757	**0.7086**	0.6709
Zhong Nanshan(X)	0.1293	0.0854	0.0288	0.0015	0.0428	0.0748	0.1912	0.2301	0.2855	0.3703	0.4331	0.5123	0.5312	**0.5840**	0.5043
epidemic(X)	**0.4150**	0.3857	0.3547	0.3236	0.3166	0.2827	0.2331	0.1596	0.1437	0.1030	0.0431	0.0215	0.0647	0.0962	0.1663
flu(X)	**0.4611**	0.4511	0.4264	0.4075	0.4054	0.4170	0.3977	0.3384	0.3104	0.2939	0.2868	0.2422	0.1937	0.1300	0.1949
diarrhea(X)	0.0556	0.0828	0.0730	0.0828	**0.1072**	0.0937	0.0417	0.0749	0.0513	0.0424	0.0289	0.0301	0.0721	0.0005	0.0126
stuffy nose(X)	0.1024	0.0776	0.0262	0.0070	0.0526	0.0643	0.1286	0.1443	0.1863	0.2602	0.2888	0.3530	0.3966	**0.5099**	0.5015
dry cough(X)	0.1803	0.1468	0.0722	0.0490	0.0041	0.0233	0.1185	0.1224	0.1823	0.2316	0.3360	**0.3967**	0.3674	0.3513	0.3890
NCP(X)	**0.1607**	0.1538	0.1322	0.1125	0.0684	0.0348	0.0059	0.0314	0.0354	0.0610	0.0212	0.0167	0.0250	0.0961	0.0750
close contact(X)	**0.7448**	0.7287	0.7208	0.7292	0.7010	0.6636	0.6239	0.6039	0.6007	0.5981	0.5843	0.5589	0.5850	0.5882	0.5661
suspected case(X)	**0.5152**	0.5058	0.4599	0.4061	0.4109	0.3573	0.2717	0.2664	0.2653	0.2032	0.1760	0.1313	0.1482	0.0863	0.0368
pneumonia(X)	**0.4836**	0.4654	0.4026	0.4045	0.3822	0.3824	0.2911	0.2921	0.2744	0.2437	0.1821	0.1390	0.0999	0.0730	0.0614
cough(X)	0.3059	0.3511	0.4283	0.4694	0.4519	0.4275	0.4570	0.4171	0.4566	0.4631	0.5601	0.6258	0.6730	0.6681	**0.7276**
dyspnea(X)	0.2276	**0.2563**	0.2373	0.2002	0.1587	0.1559	0.0836	0.0907	0.0808	0.0746	0.0133	0.0301	0.0888	0.1827	0.2150
virus pneumonia(X)	0.5483	0.5342	0.5454	0.5454	0.5847	0.6235	0.6586	0.6135	0.6333	0.6453	**0.6615**	0.6188	0.5992	0.6141	0.6555
novel coronavirus(X)	**0.6553**	0.6241	0.5881	0.5543	0.5330	0.5099	0.4672	0.4594	0.4632	0.4338	0.3969	0.3618	0.3317	0.2922	0.2423

Optimal values are shown in bold in the table

Table 18 Correlation analysis between keywords and the cumulative number of confirmed cases—Jingzhou

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.0972	0.1061	0.0681	0.0318	0.0088	0.0574	0.1386	0.1825	0.2238	0.3029	0.3872	0.4680	0.5113	**0.5175**	0.4839
Red Cross Society(S)	0.2385	0.1859	0.1325	0.1022	0.0587	0.0123	0.0263	0.0896	0.1312	0.2025	0.2720	0.3457	0.3942	0.4797	**0.5573**
Spring Festival(S)	0.7395	0.7628	0.8038	0.8187	0.8308	0.8521	**0.8584**	0.8519	0.8377	0.8360	0.8525	0.8419	0.8312	0.8400	0.8397
Health Commission(S)	0.0015	0.0553	0.1088	0.1626	0.2280	0.3054	0.3798	0.4382	0.5044	0.5967	0.6936	0.7653	0.7995	0.8114	**0.8187**
vaccine(S)	0.2360	0.1913	0.1671	0.1381	0.0918	0.0302	0.0031	0.0692	0.1135	0.1952	0.2859	0.3649	0.4149	0.4676	**0.4757**
SARS(S)	0.1707	0.2381	0.3100	0.3838	0.4603	0.5403	0.6282	0.7210	0.8187	0.9198	0.9711	0.9756	0.9767	**0.9800**	0.9785
hand washing(S)	**0.2960**	0.2606	0.2289	0.1922	0.1663	0.1643	0.1216	0.1063	0.0829	0.0431	0.0238	0.0389	0.0165	0.0137	0.0391
children(S)	0.4189	0.4044	0.3780	0.3959	0.4220	0.4341	**0.4634**	0.4264	0.4097	0.3769	0.3656	0.3207	0.2707	0.3264	0.3517
mask(S)	0.1038	0.1553	0.2175	0.2696	0.3309	0.3864	0.4599	0.5435	0.6207	0.7087	0.8052	**0.8544**	0.8427	0.8412	0.8340
coronavirus(S)	0.1322	0.1946	0.2619	0.3305	0.4052	0.4862	0.5694	0.6564	0.7499	0.8516	0.9548	0.9717	0.9688	0.9678	**0.9743**
novel(S)	0.0306	0.0799	0.1339	0.1861	0.2457	0.3226	0.3924	0.4768	0.5639	0.6544	0.7585	0.7873	**0.7895**	0.7732	0.7691
disinfection(S)	0.0327	0.0117	0.0326	0.0828	0.1153	0.1609	0.2313	0.2733	0.3038	0.3572	0.4241	0.4480	0.4749	**0.5079**	0.4996
lockdown of the city(S)	0.1070	0.0541	0.0135	0.0384	0.0832	0.1504	0.1994	0.2543	0.3014	0.3702	0.4485	0.5301	0.6024	**0.6870**	0.6599
Wuhan(S)	0.1249	0.1848	0.2483	0.3125	0.3890	0.4674	0.5450	0.6356	0.7268	0.8296	0.9367	0.9596	0.9635	**0.9670**	0.9660
Zhong Nanshan(S)	0.0876	0.1445	0.2064	0.2723	0.3520	0.4253	0.5053	0.5834	0.6775	0.7670	0.8632	0.8686	0.8871	**0.8963**	0.8918
epidemic(S)	0.8964	**0.9001**	0.8941	0.8937	0.8934	0.8758	0.8623	0.8458	0.8467	0.8325	0.8293	0.8008	0.7934	0.7743	0.7555
flu(S)	0.1302	0.1704	0.1994	0.2684	0.3408	0.4031	0.4809	0.5702	0.6599	0.7511	0.8468	0.8995	0.9405	0.9453	**0.9528**
diarrhea(S)	0.0430	0.0013	0.0400	0.0989	0.1714	0.2106	0.2700	0.3287	0.4087	0.4834	0.5792	0.6435	0.7347	**0.7718**	0.7560
dry cough(S)	0.1946	0.2555	0.3192	0.3847	0.4333	0.5044	0.5135	0.6051	0.6948	0.7680	0.8598	**0.8973**	0.8908	0.8825	0.8773
NCP(S)	**0.7113**	0.6914	0.6802	0.6547	0.6319	0.6111	0.5834	0.5575	0.5302	0.4968	0.4648	0.4351	0.4010	0.3582	0.3234
novel coronary pneumonia(S)	0.9258	**0.9330**	0.9280	0.9303	0.9325	0.9292	0.9277	0.9139	0.9127	0.8991	0.8984	0.8760	0.8660	0.8644	0.8530
COVID-19(S)	**0.7992**	0.7946	0.7970	0.7834	0.7683	0.7577	0.7493	0.7218	0.7089	0.6887	0.6817	0.6605	0.6317	0.5980	0.5858
novel coronavirus pneumonia(S)	0.2258	0.1876	0.1399	0.0928	0.0450	0.0119	0.0674	0.1235	0.1864	0.2606	0.3351	0.4314	0.5134	**0.5602**	0.5440
asymptomatic infection(S)	**0.7379**	0.7172	0.6877	0.6628	0.6608	0.6373	0.6076	0.5893	0.5599	0.5449	0.5022	0.4647	0.4342	0.4174	0.3759
aerosol transmission(S)	**0.5028**	0.4743	0.4489	0.4155	0.3787	0.3393	0.3092	0.2777	0.2351	0.1843	0.1472	0.1068	0.0756	0.0243	0.0230
pneumonia(S)	0.0614	0.1206	0.1818	0.2475	0.3195	0.3965	0.4746	0.5589	0.6490	0.7449	0.8427	0.9516	0.9707	0.9706	**0.9725**
cough(S)	0.1385	0.1038	0.0397	0.0071	0.0636	0.1137	0.1844	0.2439	0.3207	0.3734	0.4585	0.5677	**0.6198**	0.6042	0.5901
fever(S)	0.2194	0.2036	0.2329	0.2734	0.3329	0.3930	0.4336	0.4538	0.5395	0.6131	0.6877	0.7577	0.7765	**0.7802**	0.7540
dyspnea(S)	0.1462	0.2118	0.2682	0.2528	0.3123	0.3890	0.4484	0.5259	0.5904	0.6689	0.7562	0.8056	0.8253	0.8243	**0.8378**
high temperature(S)	0.0004	0.0301	0.0953	0.1457	0.2080	0.2751	0.3535	0.4147	0.4620	0.5360	0.6137	0.6747	**0.7122**	0.6840	0.6713
novel coronavirus(S)	0.0015	0.0609	0.1206	0.1832	0.2523	0.3283	0.4043	0.4868	0.5728	0.6615	0.7582	0.8638	0.9187	0.9525	**0.9585**
SARS-CoV-2(S)	0.4581	**0.4716**	0.4476	0.4543	0.4195	0.4032	0.4066	0.3842	0.3769	0.3604	0.3720	0.3730	0.3374	0.2999	0.2752
Spring Festival(X)	0.7844	0.7843	0.7896	**0.7950**	0.7828	0.7839	0.7852	0.7832	0.7700	0.7509	0.7464	0.7565	0.7543	0.7401	0.7254
Red Cross Society(X)	0.1417	0.0963	0.0625	0.0121	0.0209	0.0800	0.1484	0.2326	0.3028	0.3416	0.3755	0.4318	0.5047	0.5814	**0.5816**
government(X)	0.3433	0.3564	0.3140	0.2605	0.2408	0.2567	0.3073	0.3538	0.3768	0.4119	**0.4144**	0.3754	0.3560	0.3779	0.3788

Table 18 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0498	0.0990	0.1268	0.1906	0.2696	0.3365	0.3615	0.4424	0.5176	0.6061	0.6765	0.7082	0.7432	**0.7498**	0.7363
medical(X)	0.1542	0.1247	0.0751	0.0185	0.0112	0.0613	0.1068	0.1505	0.1933	0.2436	0.2925	0.3579	0.4270	**0.4510**	0.4452
vaccine(X)	0.2128	0.1889	0.1719	0.1292	0.1095	0.0861	0.0374	0.0260	0.0793	0.1275	0.1946	0.2708	0.3455	**0.3806**	0.3615
SARS(X)	0.1083	0.1669	0.2182	0.2795	0.3242	0.3905	0.4388	0.5174	0.5934	0.6638	0.7484	**0.7775**	0.7728	0.7655	0.7546
children(X)	0.3436	0.3556	0.4009	0.4241	0.4535	0.5130	0.4898	0.4672	0.4775	0.5151	0.5739	**0.6178**	0.5841	0.5480	0.5488
mask(X)	0.0503	0.0152	0.0196	0.0602	0.1001	0.1599	0.2132	0.2725	0.3226	0.3821	0.4442	0.5256	0.5992	**0.5997**	0.5830
infection(X)	0.3250	0.3409	0.3531	0.3786	0.4052	0.4287	0.4533	0.4791	0.5062	0.5346	0.5646	0.5961	**0.6294**	0.5993	0.5656
coronavirus(X)	**0.6675**	0.6395	0.6190	0.5906	0.5516	0.5242	0.4871	0.4513	0.4146	0.3830	0.3462	0.3068	0.2546	0.2062	0.1676
disinfection(X)	0.5670	0.5798	0.5931	0.6070	0.6216	0.6368	0.6528	0.6695	0.6871	0.7055	0.7250	0.7454	0.7669	0.7897	**0.7951**
nucleic acid(X)	**0.5120**	0.4889	0.4719	0.4528	0.4107	0.3734	0.3293	0.2804	0.2427	0.2118	0.1767	0.1171	0.0717	0.0077	0.0601
lockdown of the city(X)	0.5268	0.5402	0.5572	0.5661	0.5753	0.5881	0.6016	0.6158	0.6308	0.6467	0.6634	0.6811	0.6998	0.7197	**0.7209**
Wuhan(X)	0.0275	0.0192	0.0774	0.1332	0.2065	0.2629	0.3240	0.3917	0.4651	0.5445	0.6154	0.6138	0.6205	**0.6678**	0.6548
Zhong Nanshan(X)	0.1314	0.0988	0.0565	0.0064	0.0370	0.0935	0.1470	0.2152	0.2907	0.3650	0.4335	0.5190	**0.5961**	0.5907	0.5764
epidemic(X)	**0.4958**	0.4684	0.4364	0.3966	0.3451	0.3156	0.2729	0.2160	0.1556	0.1171	0.0785	0.0180	0.0534	0.1124	0.1644
flu(X)	**0.4355**	0.4239	0.4177	0.4100	0.4140	0.4077	0.3904	0.3502	0.2987	0.2720	0.2480	0.2255	0.1788	0.1357	0.1460
diarrhea(X)	0.0194	0.0352	0.0478	0.0553	0.0725	0.0888	0.0859	0.0834	0.1014	0.1200	0.1325	0.1456	0.1594	0.1740	**0.1895**
stuffy nose(X)	0.0882	0.0542	0.0226	0.0201	0.0390	0.0756	0.0987	0.1500	0.2044	0.2068	0.2763	0.3381	0.4049	0.4684	**0.5347**
dry cough(X)	0.1696	0.1167	0.0610	0.0064	0.0235	0.0828	0.1466	0.2177	0.2799	0.3508	0.4161	**0.4920**	0.4699	0.4767	0.4747
NCP(X)	**0.2420**	0.2145	0.1826	0.1740	0.1726	0.1323	0.0914	0.0547	0.0179	0.0037	0.0140	0.0359	0.0544	0.0723	0.1267
close contact(X)	**0.7843**	0.7668	0.7541	0.7493	0.7622	0.7423	0.7196	0.6914	0.6853	0.6845	0.6635	0.6344	0.6068	0.6015	0.5750
suspected case(X)	**0.5190**	0.4831	0.4643	0.4350	0.4234	0.3912	0.3487	0.3122	0.2854	0.2624	0.2209	0.1788	0.1356	0.1058	0.0555
pneumonia(X)	**0.5164**	0.4889	0.4605	0.4241	0.4213	0.3953	0.3584	0.3023	0.2540	0.2229	0.1908	0.1387	0.0834	0.0373	0.0392
cough(X)	0.3198	0.3666	0.4249	0.4901	0.4634	0.4467	0.4377	0.4575	0.4922	0.4981	0.5598	0.6490	**0.6808**	0.6774	0.6722
dyspnea(X)	0.2021	0.1729	0.1514	0.1235	0.1004	0.0648	0.0298	0.0026	0.0019	0.0181	0.0684	0.1228	0.1783	0.2296	**0.2684**
virus pneumonia(X)	0.5512	0.5490	0.5213	0.5438	0.5432	0.5762	0.5780	0.5895	0.6101	**0.6159**	0.6029	0.5867	0.5698	0.5552	0.5419
novel coronavirus(X)	**0.6930**	0.6723	0.6544	0.6309	0.5930	0.5696	0.5423	0.5105	0.4663	0.4385	0.4130	0.3773	0.3313	0.2869	0.2507

Optimal values are shown in bold in the table

Table 19 Correlation analysis between keywords and the cumulative number of confirmed cases—Qianjiang

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	**0.0865**	0.0834	0.0791	0.0721	0.0665	0.0592	0.0512	0.0493	0.0408	0.0332	0.0239	0.0141	0.0062	0.0043	0.0112
Red Cross Society(S)	0.0571	0.0089	0.0116	0.0550	0.0953	0.1427	0.1727	0.2203	0.2534	0.3135	0.3568	0.4148	0.4621	0.5221	**0.5523**
Spring Festival(S)	0.6851	0.7290	0.7552	**0.7556**	0.7491	0.7371	0.7099	0.6997	0.6856	0.6662	0.6434	0.6396	0.6206	0.6137	0.5832
Health Commission(S)	0.0200	0.0338	0.0734	0.1206	0.1490	0.2003	0.2506	0.3184	0.3835	0.4425	0.4949	0.5513	**0.6329**	0.6167	0.6133
vaccine(S)	0.1707	0.1418	0.1083	0.0759	0.0440	0.0083	0.0471	0.1009	0.1233	0.1703	0.2367	0.2888	0.3291	0.3624	**0.3789**
SARS(S)	0.1996	0.2650	0.3367	0.4112	0.4920	0.5657	0.6342	0.7201	0.8105	0.9011	0.9232	0.9333	**0.9338**	0.9329	0.9246
hand washing(S)	**0.1386**	0.1345	0.1331	0.1316	0.1285	0.1299	0.1251	0.0817	0.0787	0.0773	0.0689	0.0637	0.0581	0.0504	0.0441
children(S)	0.2831	0.2557	0.2543	0.2549	0.2555	0.2540	0.2530	0.2597	0.3052	0.3051	0.3093	0.3086	0.3105	0.3112	**0.3143**
mask(S)	0.1035	0.0388	0.0051	0.0325	0.0935	0.1606	0.2107	0.2555	0.3339	0.3976	0.4752	**0.5673**	0.5476	0.5170	0.4822
coronavirus(S)	0.0913	0.1585	0.2255	0.3017	0.3762	0.4572	0.5445	0.6344	0.7311	0.8320	0.9294	0.9632	0.9666	0.9674	**0.9675**
novel(S)	0.0502	0.0692	0.1226	0.1632	0.2246	0.2557	0.3075	0.3769	0.4353	0.4971	0.5421	0.5916	**0.6529**	0.6383	0.6184
disinfection(S)	0.0543	0.0532	0.0521	0.0509	0.0473	0.0506	0.0469	0.0454	0.0439	0.0423	0.0406	0.0173	0.0808	**0.0856**	0.0681
lockdown of the city(S)	0.1248	0.1808	0.2035	0.2286	0.2539	0.2823	0.3154	0.3440	0.3794	0.4116	0.4502	0.4915	0.5526	**0.5999**	0.5628
Wuhan(S)	0.0978	0.1711	0.2314	0.3044	0.3678	0.4415	0.5252	0.6106	0.7038	0.8074	0.8909	0.8920	**0.9037**	0.8993	0.8931
Zhong Nanshan(S)	0.0677	0.1371	0.2028	0.2652	0.3420	0.4173	0.5065	0.5959	0.6903	0.7793	0.8878	**0.8946**	0.8906	0.8754	0.8673
epidemic(S)	**0.8756**	0.8523	0.8361	0.8347	0.8268	0.8184	0.8197	0.8102	0.7845	0.7725	0.7599	0.7304	0.6969	0.6688	0.6333
flu(S)	0.2278	0.2494	0.2934	0.3396	0.3839	0.4309	0.4707	0.4968	0.5668	0.6358	0.7433	**0.8053**	0.7991	0.8040	0.7887
diarrhea(S)	0.1210	0.1807	0.2380	0.2981	0.3289	0.3648	0.4014	0.4603	0.5060	0.5636	0.6125	0.6214	0.6741	0.6905	**0.6979**
dry cough(S)	0.2520	0.3006	0.3320	0.3647	0.4249	0.4602	0.5015	0.5603	0.6183	0.6658	0.7204	0.7778	**0.7867**	0.7757	0.7616
NCP(S)	**0.3920**	0.3593	0.3318	0.3047	0.3026	0.2705	0.2370	0.1975	0.1594	0.1216	0.1078	0.0920	0.0537	0.0133	0.0066
novel coronary pneumonia(S)	0.8860	0.8911	0.8897	0.8964	0.8912	0.8873	0.8955	**0.8967**	0.8835	0.8721	0.8718	0.8712	0.8508	0.8326	0.8155
COVID-19(S)	0.6170	**0.6299**	0.6208	0.6114	0.6237	0.6076	0.5979	0.5733	0.5852	0.5616	0.5308	0.5394	0.5127	0.4832	0.4669
novel coronavirus pneumonia(S)	0.2969	0.2478	0.2058	0.1688	0.1177	0.0684	0.0176	0.0262	0.0939	0.1461	0.2154	0.2991	0.3862	**0.4328**	0.4177
asymptomatic infection(S)	0.4758	**0.4832**	0.4557	0.4366	0.4051	0.4112	0.3887	0.3557	0.3340	0.3029	0.2697	0.2627	0.2593	0.2189	0.1771
aerosol transmission(S)	**0.1829**	0.1501	0.1448	0.1375	0.1298	0.0890	0.0539	0.0158	0.0330	0.0491	0.0681	0.0905	0.1143	0.1361	0.1575
pneumonia(S)	0.0120	0.0822	0.1569	0.2226	0.2887	0.3631	0.4367	0.5188	0.6128	0.7101	0.8121	0.9166	0.9363	0.9366	**0.9384**
cough(S)	0.0434	0.0920	0.0913	0.1138	0.1085	0.1663	0.2178	0.2902	0.3475	0.3487	0.3749	0.4535	0.4631	0.4951	**0.5201**
fever(S)	0.5331	0.5444	0.5461	0.5881	0.5682	0.5701	0.5530	0.5985	0.6171	0.6671	0.7196	0.7452	**0.7711**	0.7683	0.7487
dyspnea(S)	0.1877	0.2046	0.2449	0.2668	0.2868	0.3088	0.3336	0.3593	0.3837	0.4120	0.4669	0.4956	**0.5334**	0.5234	0.4819
high temperature(S)	0.3142	0.3377	0.3601	0.3888	0.4160	0.4429	0.4714	0.5088	0.5403	0.5740	0.6130	**0.6530**	0.6304	0.5981	0.5686
novel coronavirus(S)	0.0824	0.0162	0.0473	0.1092	0.1792	0.2541	0.3321	0.4115	0.5004	0.5909	0.6855	0.7887	0.8871	**0.9288**	0.9249
SARS-CoV-2(S)	0.2069	0.2111	0.2155	0.2201	0.2248	0.2298	0.2350	0.2404	0.2461	**0.2520**	0.2000	0.1453	0.1438	0.1421	0.1403
Spring Festival(X)	0.7968	0.8059	0.8177	**0.8236**	0.8202	0.8218	0.8147	0.7944	0.8039	0.7963	0.7849	0.7695	0.7563	0.7382	0.7322
Red Cross Society(X)	0.1527	0.0965	0.0366	0.0282	0.0820	0.1373	0.1996	0.2403	0.3150	0.3680	0.4145	0.4629	0.5514	**0.6063**	0.6053
government(X)	0.2483	0.2515	0.2234	0.2048	0.2012	0.1856	0.2298	0.3182	0.3592	0.3585	0.3545	0.3560	0.3608	**0.3761**	0.3672

Table 19 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0568	0.1310	0.1454	0.1886	0.2429	0.2983	0.3253	0.3751	0.4583	0.5447	0.6341	0.6353	**0.6837**	0.6673	0.6547
medical(X)	0.1603	0.1030	0.0533	0.0011	0.0409	0.0866	0.1277	0.1595	0.2235	0.2753	0.3227	0.3773	0.4332	**0.4709**	0.4584
vaccine(X)	0.2381	0.1906	0.1632	0.1357	0.1045	0.0766	0.0311	0.0102	0.0283	0.0796	0.1421	0.2180	0.2888	0.3216	**0.3256**
SARS(X)	0.0823	0.1522	0.2217	0.2950	0.3701	0.4420	0.4991	0.5391	0.6275	0.7081	0.8026	0.8239	**0.8245**	0.8157	0.8082
children(X)	0.4820	0.4601	0.4908	0.5139	0.5426	0.5739	0.6029	0.5858	0.6162	0.6533	0.6900	**0.7238**	0.7043	0.6873	0.6652
mask(X)	0.1003	0.0409	0.0131	0.0528	0.0932	0.1360	0.1785	0.2161	0.2832	0.3475	0.4157	0.4810	**0.5576**	0.5423	0.5178
infection(X)	0.3136	0.3308	0.3502	0.3689	0.3869	0.4072	0.4273	0.4496	0.4745	0.4991	0.5249	0.5556	**0.5805**	0.5426	0.5000
coronavirus(X)	**0.6745**	0.6405	0.6105	0.5856	0.5546	0.5236	0.4929	0.4916	0.4595	0.4246	0.3898	0.3506	0.3107	0.2743	0.2414
disinfection(X)	0.4219	0.4309	0.4403	0.4502	0.4604	0.4711	0.4823	0.4941	0.5063	0.5192	0.5328	0.5470	0.5619	0.5777	**0.5943**
nucleic acid(X)	**0.5197**	0.4808	0.4470	0.4349	0.4050	0.3808	0.3676	0.3638	0.3152	0.2637	0.2146	0.1692	0.1275	0.0945	0.0368
lockdown of the city(X)	0.4891	0.4968	0.5094	0.5197	0.5288	0.5400	0.5510	0.5650	0.5764	0.5909	0.6065	0.6209	0.6372	0.6545	**0.6728**
Wuhan(X)	0.0180	0.0638	0.1233	0.1879	0.2711	0.3509	0.4197	0.4687	0.5501	0.6143	0.6769	0.6740	0.7027	**0.7592**	0.7471
Zhong Nanshan(X)	0.1772	0.1141	0.0521	0.0018	0.0386	0.0966	0.1695	0.2076	0.2778	0.3437	0.4245	0.5159	0.5728	**0.5761**	0.5571
epidemic(X)	**0.5792**	0.5358	0.4991	0.4648	0.4275	0.3887	0.3448	0.3366	0.2951	0.2451	0.1993	0.1516	0.1000	0.0514	0.0037
flu(X)	**0.5131**	0.4752	0.4337	0.4086	0.3939	0.3827	0.3689	0.3624	0.3491	0.3273	0.2917	0.2388	0.1899	0.1428	0.1680
diarrhea(X)	0.0097	0.0192	0.0253	0.0317	0.0383	0.0511	0.0526	0.0539	0.0681	0.0765	0.0853	0.0945	0.1042	0.1168	**0.1254**
stuffy nose(X)	0.0610	0.0075	0.0623	0.1072	0.1078	0.1192	0.1110	0.1510	0.2063	0.1866	0.2454	0.2859	0.3033	0.3875	**0.4320**
dry cough(X)	0.2420	0.2070	0.1620	0.1109	0.0493	0.0099	0.0681	0.1036	0.1760	0.2486	0.3165	**0.3822**	0.3464	0.3467	0.3233
NCP(X)	**0.2050**	0.1687	0.1650	0.1589	0.1245	0.1213	0.1155	0.1094	0.0642	0.0219	0.0227	0.0729	0.0878	0.1044	0.1278
close contact(X)	**0.7952**	0.7761	0.7588	0.7514	0.7284	0.7075	0.6986	0.7087	0.6866	0.6609	0.6305	0.6109	0.5954	0.5718	0.5445
suspected case(X)	**0.5455**	0.5005	0.4627	0.4345	0.3899	0.3511	0.3119	0.2975	0.2582	0.2103	0.1616	0.1226	0.0862	0.0325	0.0189
pneumonia(X)	**0.5972**	0.5558	0.5203	0.4894	0.4512	0.4141	0.3709	0.3653	0.3246	0.2760	0.2305	0.1798	0.1292	0.0797	0.0597
cough(X)	0.2695	0.3375	0.4260	0.4826	0.4785	0.4720	0.4743	0.4575	0.4998	0.5304	0.6134	0.6932	0.7560	0.7725	**0.7844**
dyspnea(X)	0.1710	0.1246	0.0800	0.0396	0.0005	0.0372	0.0557	0.0892	0.0767	0.0598	0.1174	0.1598	0.2378	0.2727	**0.3016**
virus pneumonia(X)	0.5645	0.5924	0.5769	0.5692	0.5773	0.5903	0.6114	0.5919	0.6524	**0.6826**	0.6750	0.6615	0.6429	0.6316	0.6015
novel coronavirus(X)	**0.7223**	0.6907	0.6625	0.6407	0.6114	0.5836	0.5546	0.5553	0.5213	0.4844	0.4544	0.4180	0.3812	0.3453	0.3133

Optimal values are shown in bold in the table

Table 20 Correlation analysis between keywords and the cumulative number of confirmed cases—Shennongjia

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.1026	0.1047	0.1068	0.1090	**0.1113**	0.0934	0.0913	0.0891	0.0868	0.0844	0.0818	0.0792	0.0764	0.0735	0.0704
Red Cross Society(S)	0.2322	0.3017	0.2966	0.2913	0.2858	0.2801	0.2740	0.2677	0.2611	0.2542	0.2853	0.3130	0.3517	0.3896	**0.4410**
Spring Festival(S)	0.3336	0.3402	0.3451	0.4074	0.3831	0.3849	0.4170	0.3819	0.4203	0.4209	0.4312	0.4494	0.5003	**0.5152**	0.4996
Health Commission(S)	0	0.0038	0.0352	0.0319	0.0285	0.0250	0.0319	0.0327	0.0334	0.0388	0.0351	0.0312	0.0506	0.0548	**0.0593**
vaccine(S)	0.0602	0.0570	0.0537	0.0503	0.0467	0.0429	0.0390	0.0349	0.0306	0.0835	0.1391	0.1371	0.1349	0.1472	**0.1659**
SARS(S)	0.2742	0.3386	0.4105	0.4792	0.5391	0.5606	0.5823	0.6163	0.6554	0.6947	0.7406	0.7558	**0.7885**	0.7509	0.7215
hand washing(S)	0.1026	0.1047	0.1068	0.1090	0.1113	**0.1137**	0.0913	0.0891	0.0868	0.0844	0.0818	0.0792	0.0764	0.0735	0.0704
children(S)	0.1463	0.1492	0.1522	0.1554	0.1587	0.1621	0.1657	0.1694	0.1733	0.1774	0.1817	0.1862	**0.1910**	0.0455	0.0503
mask(S)	0.1193	0.1835	0.2063	0.2339	0.2644	0.2603	0.2940	0.3721	0.3839	0.3940	0.4487	**0.4581**	0.3912	0.3143	0.2306
coronavirus(S)	0.0061	0.0398	0.0490	0.1004	0.1512	0.2187	0.2741	0.3170	0.3872	0.4429	0.5396	0.6240	0.6119	0.6066	**0.6424**
novel(S)	0.1463	0.1439	0.1413	0.1387	0.1360	0.1331	0.1302	0.1271	0.1919	0.1901	0.1882	0.1862	0.2012	0.2064	**0.2120**
disinfection(S)	0.1026	0.1047	0.1068	0.1090	0.1113	0.1137	0.1161	0.1188	0.1215	0.1243	0.1273	0.1305	0.1337	0.1372	**0.1409**
lockdown of the city(S)	0.2101	0.2066	0.2031	0.1994	0.2381	0.2800	0.3268	0.3256	0.3348	0.3492	0.3697	0.3882	0.4032	**0.4144**	0.3317
Wuhan(S)	0.0310	0.0897	0.1360	0.1519	0.2254	0.2101	0.2401	0.3303	0.4176	0.4896	0.5343	**0.5667**	0.5512	0.5635	0.5238
Zhong Nanshan(S)	0.0379	0.0035	0.0273	0.0854	0.1125	0.1831	0.2483	0.3137	0.4284	0.5262	0.6482	0.6596	**0.6726**	0.6389	0.6186
epidemic(S)	0.7322	0.7497	0.7625	0.7578	0.7776	**0.7875**	0.7434	0.7014	0.7070	0.6819	0.6304	0.6074	0.5915	0.5775	0.5223
flu(S)	0.1413	0.1348	0.2200	0.3020	0.3226	0.3442	0.3810	0.3863	0.3944	0.4446	0.4616	**0.4713**	0.4389	0.3860	0.3145
diarrhea(S)	0.0000	0.0027	0.0054	0.0083	0.0113	0.0145	0.0177	0.0454	0.0433	0.0412	0.0389	0.0532	0.0580	0.0630	**0.0754**
dry cough(S)	0.1026	0.1009	0.0992	0.0973	0.0954	0.0934	0.0913	0.0891	0.0868	0.0844	0.0818	**0.1817**	0.1815	0.1813	0.1811
NCP(S)	0.1463	0.1492	**0.1522**	0.0083	0.0113	0.1331	0.1302	0.1271	0.1238	0.1204	0.1168	0.1131	0.1091	0.1050	0.1006
novel coronary pneumonia(S)	0.6436	0.6586	**0.6743**	0.6517	0.6297	0.6477	0.6666	0.6184	0.6387	0.6206	0.6431	0.5937	0.5586	0.5837	0.5499
COVID-19(S)	0.1026	0.1047	0.1068	0.1090	0.1113	0.1137	0.1161	0.1188	0.1215	0.1243	0.1273	0.1305	0.1337	0.1372	**0.1409**
novel coronavirus pneumonia(S)	0.2735	0.2284	0.1756	0.1066	0.0282	0.0348	0.0977	0.1338	0.1606	0.1547	0.1992	0.2503	0.3215	**0.3532**	0.3476
asymptomatic infection(S)	0.1026	0.1009	0.0992	0.0973	0.0954	0.0934	0.0913	0.0891	0.0868	0.0844	0.0818	0.0792	0.0764	0.0735	**0.1811**
aerosol transmission(S)	0.0620	0.0654	0.0556	0.0522	**0.1679**	0.1644	0.1608	0.1569	0.1529	0.1488	0.1444	0.1397	0.1349	0.1298	0.1244
pneumonia(S)	0.2923	0.2285	0.1767	0.0912	0.0294	0.0060	0.0676	0.1590	0.2500	0.3554	0.4383	**0.4751**	0.4671	0.3809	0.3320
cough(S)	0.1463	0.1492	0.1522	0.1554	0.1587	0.1621	**0.1657**	0.0212	0.0248	0.0285	0.0324	0.0366	0.0409	0.0455	0.0503
fever(S)	0.0000	0.0027	0.0054	0.0083	**0.1360**	0.1331	0.1302	0.1271	0.1238	0.1204	0.1168	0.1131	0.1091	0.1050	0.1006
dyspnea(S)	0.1026	0.1047	0.1068	0.1090	**0.1113**	0.0934	0.0913	0.0891	0.0868	0.0844	0.0818	0.0792	0.0764	0.0735	0.0704
high temperature(S)	0.1026	0.1009	0.0992	0.0973	0.0954	0.0934	0.0913	0.0891	0.0868	0.0844	0.0818	**0.1817**	0.1815	0.1813	0.1811
novel coronavirus(S)	0.1984	0.1864	0.1241	0.0698	0.0049	0.0870	0.1585	0.2494	0.3712	0.4850	0.6048	0.6774	**0.6892**	0.6771	0.6647
SARS-CoV-2(S)	0.1026	0.1047	0.1068	0.1090	0.1113	0.1137	0.1161	0.1188	0.1215	0.1243	0.1273	0.1305	0.1337	0.1372	**0.1409**
Spring Festival(X)	0.7494	0.7428	0.7666	0.7840	**0.7969**	0.7598	0.7808	0.7796	0.7910	0.7624	0.7502	0.7571	0.7459	0.7469	0.7526
Red Cross Society(X)	0.1711	0.1287	0.0962	0.0608	0.0134	0.0175	0.0021	0.0404	0.1191	0.1924	0.1998	0.2427	0.3026	0.3757	**0.4123**
government(X)	0.1518	0.1763	0.1476	0.0915	0.1130	0.1266	0.1808	0.2629	0.2665	0.3497	0.3997	0.3638	0.3313	0.4163	**0.4232**

Table 20 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0090	0.0661	0.0791	0.1241	0.1627	0.1532	0.1107	0.2173	0.3149	0.3750	0.4485	0.4952	**0.5991**	0.5660	0.5609
medical(X)	0.1439	0.1300	0.1098	0.0602	0.0315	0.0360	0.0613	0.0268	0.0950	0.1699	0.2018	0.2514	0.3268	0.3389	**0.3664**
vaccine(X)	**0.2743**	0.2419	0.2486	0.2320	0.2027	0.2085	0.2081	0.1454	0.0990	0.0507	0.0090	0.0672	0.1330	0.1282	0.0799
SARS(X)	0.0049	0.0669	0.1307	0.1929	0.2576	0.3189	0.3828	0.4721	0.5747	0.6885	0.8027	**0.8357**	0.8344	0.7853	0.7792
children(X)	0.1026	0.1009	0.0992	0.0973	0.0954	0.0934	0.0913	0.0891	0.0868	0.0844	0.0818	0.0792	0.0764	**0.1813**	0.1811
mask(X)	0.1080	0.0687	0.0665	0.0405	0.0172	0.0257	0.0381	0.1254	0.2158	0.2672	0.3155	0.3693	**0.4216**	0.4067	0.4026
infection(X)	0.1805	0.1776	0.1745	0.2199	0.2705	0.2692	0.2679	0.2786	0.2955	0.3051	0.3204	0.3918	**0.3934**	0.3052	0.1811
coronavirus(X)	**0.7398**	0.6987	0.6773	0.6437	0.6093	0.6435	0.5881	0.5316	0.4806	0.4627	0.4182	0.3708	0.3283	0.3719	0.3620
disinfection(X)	0.2367	0.2415	0.2465	0.2517	0.2572	0.2628	0.2687	0.2749	0.2814	0.2882	0.2953	0.3028	0.3107	0.3190	**0.3277**
nucleic acid(X)	**0.4908**	0.4749	0.4323	0.3835	0.3317	0.3600	0.3831	0.3159	0.2489	0.2467	0.2338	0.1895	0.1899	0.2109	0.1271
lockdown of the city(X)	0.4066	0.4153	0.4243	0.4337	0.4436	0.4539	0.4646	0.4759	0.4878	0.5002	0.5133	0.5270	0.5415	0.5568	**0.5730**
Wuhan(X)	0.0727	0.0265	0.0301	0.0798	0.1226	0.1737	0.2104	0.2715	0.3734	0.4592	0.5538	0.5829	0.5889	0.6095	**0.6223**
Zhong Nanshan(X)	0.3097	0.2751	0.2372	0.1825	0.1090	0.0388	0.0599	0.1477	0.2232	0.2730	0.3185	0.3900	**0.4699**	0.4314	0.4179
epidemic(X)	**0.7167**	0.6761	0.6539	0.6365	0.6089	0.5617	0.5970	0.5394	0.4800	0.4506	0.4141	0.3677	0.3172	0.2592	0.2247
flu(X)	**0.4929**	0.4328	0.3807	0.3705	0.3556	0.2907	0.2847	0.2219	0.1690	0.1297	0.0921	0.0743	0.0704	0.0600	0.1377
diarrhea(X)	0.1026	0.1047	0.1068	**0.1090**	0.0954	0.0934	0.0913	0.0891	0.0868	0.0844	0.0818	0.0792	0.0764	0.0735	0.0704
stuffy nose(X)	0.0479	0.0555	0.0121	0.0456	0.1074	0.1064	0.0963	0.1575	0.2063	0.2509	0.2704	0.2930	0.3613	0.3801	**0.4047**
dry cough(X)	0.2304	0.1822	0.1105	0.0618	0.0027	0.0505	0.1155	0.1459	0.1948	0.2483	0.3169	0.3976	0.3765	0.3897	**0.4006**
NCP(X)	0.2367	**0.2415**	0.1544	0.0596	0.0278	0.0228	0.0176	0.0121	0.0064	0.0004	0.0060	0.0126	0.0196	0.0269	0.0693
close contact(X)	0.7662	0.7847	0.8040	**0.8243**	0.8073	0.7706	0.7934	0.7475	0.7094	0.6852	0.6349	0.5930	0.5846	0.6135	0.5709
suspected case(X)	**0.5111**	0.4981	0.4831	0.4762	0.4322	0.4555	0.4020	0.3458	0.2933	0.2631	0.1783	0.1078	0.0515	0.0604	0.0015
pneumonia(X)	**0.6758**	0.6332	0.6061	0.5780	0.5373	0.5720	0.5165	0.4564	0.4017	0.3881	0.3446	0.2887	0.2479	0.2808	0.2743
cough(X)	0.2337	0.2972	0.3279	0.3971	0.3715	0.3591	0.3525	0.3845	0.4642	0.4588	0.5288	0.6428	**0.7309**	0.6946	0.7201
dyspnea(X)	0.1066	0.1182	0.1303	0.1388	0.1503	0.1624	0.1734	0.1887	0.1762	0.2267	**0.2437**	0.2335	0.2050	0.1770	0.1487
virus pneumonia(X)	0.4311	0.4250	0.3903	0.4239	0.4840	0.5374	0.5574	0.6127	0.6053	**0.6145**	0.6003	0.5467	0.5576	0.5712	0.6100
novel coronavirus(X)	**0.7858**	0.7448	0.7218	0.6948	0.6622	0.6129	0.5572	0.4982	0.4512	0.4406	0.4037	0.3647	0.3246	0.3553	0.3473

Optimal values are shown in bold in the table

Table 21 Correlation analysis between keywords and the cumulative number of confirmed cases—Shiyan

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.0331	0.0891	0.1379	0.1930	0.2411	0.2705	0.2855	0.3349	0.4004	0.4009	0.4398	**0.4491**	0.4106	0.4117	0.4194
Red Cross Society(S)	0.2750	0.2347	0.1968	0.1415	0.0810	0.0325	0.0234	0.0743	0.1508	0.2244	0.2886	0.3602	0.4488	0.5552	**0.6453**
Spring Festival(S)	0.7856	0.8085	0.8338	0.8469	0.8394	0.8587	**0.8674**	0.8619	0.8646	0.8616	0.8480	0.8431	0.8314	0.8212	0.8068
Health Commission(S)	0.0307	0.0232	0.0865	0.1435	0.2021	0.2776	0.3471	0.4302	0.5196	0.5935	0.6877	0.7753	0.8056	**0.8169**	0.8046
vaccine(S)	0.1787	0.1727	0.1243	0.0879	0.0474	0.0122	0.0179	0.0705	0.1208	0.1969	0.2589	0.3485	0.4104	**0.4685**	0.4537
SARS(S)	0.1753	0.2412	0.3124	0.3869	0.4644	0.5430	0.6294	0.7181	0.8195	0.9252	0.9650	0.9655	0.9682	**0.9688**	0.9647
hand washing(S)	**0.1800**	0.1389	0.1069	0.1037	0.0723	0.0690	0.0229	0.0113	0.0343	0.0488	0.0599	0.1143	0.1294	0.1796	0.1409
children(S)	0.3452	0.3485	0.3246	**0.3867**	0.3862	0.3791	0.3516	0.3067	0.2782	0.2467	0.2462	0.1903	0.1244	0.1546	0.2177
mask(S)	0.0461	0.1127	0.1742	0.2164	0.2884	0.3547	0.4274	0.5042	0.5763	0.6609	0.7454	0.8429	**0.8430**	0.8312	0.8092
coronavirus(S)	0.1209	0.1848	0.2483	0.3189	0.3876	0.4694	0.5560	0.6436	0.7386	0.8407	0.9485	0.9660	0.9689	0.9700	**0.9729**
novel(S)	0.0363	0.0810	0.1271	0.1901	0.2497	0.3186	0.3920	0.4559	0.5204	0.6108	0.6864	0.7919	0.7865	**0.8053**	0.8034
disinfection(S)	0.1104	0.1029	0.0949	0.0567	0.0001	0.0183	0.0570	0.0740	0.0969	0.1275	0.1875	**0.2421**	0.2409	0.2003	0.1714
lockdown of the city(S)	0.0626	0.0061	0.0476	0.0892	0.1353	0.2013	0.2559	0.3043	0.3777	0.4416	0.5220	0.5877	0.6880	**0.7876**	0.7793
Wuhan(S)	0.1263	0.1897	0.2517	0.3227	0.3976	0.4783	0.5632	0.6499	0.7429	0.8509	0.9436	0.9492	0.9575	0.9591	**0.9642**
Zhong Nanshan(S)	0.0560	0.1190	0.1744	0.2311	0.3011	0.3892	0.4753	0.5532	0.6346	0.7203	0.8103	0.8233	0.8189	**0.8287**	0.8176
epidemic(S)	**0.5496**	0.5098	0.4954	0.4655	0.4468	0.4183	0.3637	0.3234	0.2633	0.2182	0.1538	0.1096	0.0665	0.0112	0.0528
flu(S)	0.2272	0.2479	0.3055	0.3643	0.3915	0.4677	0.5021	0.5939	0.6737	0.7661	0.8615	**0.9035**	0.8935	0.8949	0.8939
diarrhea(S)	0.0470	0.0978	0.1492	0.1644	0.2280	0.2756	0.3328	0.3932	0.4533	0.5141	0.4818	0.5701	0.6513	0.6957	**0.7228**
dry cough(S)	0.1856	0.2428	0.2903	0.3299	0.3972	0.4791	0.5604	0.6382	0.6958	0.7656	0.8440	**0.8590**	0.8529	0.8453	0.8429
NCP(S)	**0.6761**	0.6580	0.6376	0.6142	0.5874	0.5588	0.5407	0.5119	0.4854	0.4548	0.4234	0.3846	0.3477	0.3119	0.2615
novel coronary pneumonia(S)	0.8885	**0.8941**	0.8901	0.8779	0.8744	0.8551	0.8410	0.8401	0.8364	0.8184	0.7900	0.7861	0.7703	0.7524	0.7248
COVID-19(S)	**0.8130**	0.7973	0.7828	0.7803	0.7813	0.7608	0.7545	0.7333	0.7088	0.6973	0.6734	0.6559	0.6319	0.6124	0.5908
novel coronavirus pneumonia(S)	**0.4067**	0.3911	0.3601	0.3216	0.2768	0.2394	0.2006	0.1603	0.1023	0.0631	0.0041	0.0644	0.1362	0.2148	0.2706
asymptomatic infection(S)	**0.7259**	0.7039	0.6802	0.6673	0.6369	0.6022	0.5736	0.5340	0.5103	0.4772	0.4323	0.3822	0.3243	0.2876	0.2496
aerosol transmission(S)	**0.3240**	0.3017	0.2725	0.2305	0.1943	0.1612	0.1167	0.0793	0.0441	0.0250	0.0141	0.0653	0.0945	0.1423	0.1725
pneumonia(S)	0.0720	0.1327	0.1959	0.2604	0.3269	0.4019	0.4877	0.5753	0.6661	0.7599	0.8696	0.9533	**0.9613**	0.9601	0.9555
cough(S)	0.0359	0.0608	0.0879	0.1496	0.2159	0.2782	0.3255	0.4100	0.4024	0.4057	0.4846	0.5043	0.5112	0.5371	**0.5441**
fever(S)	0.3585	0.3722	0.4072	0.4510	0.4780	0.5444	0.5913	0.6164	0.5946	0.6389	0.6664	0.6899	**0.6988**	0.6781	0.6505
dyspnea(S)	0.1255	0.1703	0.2333	0.2852	0.3300	0.3649	0.4290	0.4876	0.5551	0.6051	0.6330	0.6917	**0.7199**	0.7043	0.6836
high temperature(S)	0.0764	0.1249	0.1857	0.2539	0.2874	0.3616	0.4221	0.4727	0.5305	0.6100	0.7106	0.7323	**0.7369**	0.7245	0.7263
novel coronavirus(S)	0.0188	0.0397	0.0954	0.1735	0.2510	0.3215	0.3995	0.4811	0.5693	0.6627	0.7605	0.8630	0.9091	0.9302	**0.9322**
SARS-CoV-2(S)	0.4787	**0.4916**	0.4758	0.4524	0.4261	0.4042	0.3827	0.3894	0.3652	0.3667	0.3670	0.3378	0.3019	0.2979	0.2643
Spring Festival(X)	0.7795	0.7859	0.7831	0.7867	**0.7882**	0.7700	0.7754	0.7725	0.7640	0.7442	0.7336	0.7257	0.7243	0.7195	0.7097
Red Cross Society(X)	0.1580	0.1008	0.0578	0.0191	0.0371	0.0973	0.1580	0.2301	0.3173	0.3526	0.3783	0.4049	0.4742	**0.5483**	0.5446
government(X)	0.1193	0.1032	0.1121	0.1783	0.2180	**0.2291**	0.1667	0.1266	0.1092	0.0811	0.0497	0.0715	0.1169	0.0755	0.0954

Table 21 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.1786	0.2217	0.2318	0.2937	0.3748	0.3907	0.3887	0.4605	0.5506	0.6269	0.7219	0.7125	**0.7476**	0.7420	0.7342
medical(X)	0.0060	0.0480	0.0755	0.1208	0.1671	0.2048	0.2290	0.2733	0.3246	0.3595	0.4131	0.4636	0.5215	**0.5222**	0.5109
vaccine(X)	0.1112	0.0739	0.0387	0.0064	0.0472	0.0945	0.1293	0.1671	0.2271	0.2971	0.3734	0.4409	0.5188	**0.5486**	0.5457
SARS(X)	0.1175	0.1804	0.2392	0.3053	0.3741	0.4165	0.4758	0.5302	0.6051	0.6867	0.7705	0.7859	**0.7930**	0.7814	0.7775
children(X)	0.6002	0.5912	0.5762	0.6090	0.6435	0.6798	0.6669	0.6526	0.6585	0.6410	0.6773	**0.6866**	0.6612	0.6395	0.6139
mask(X)	0.0102	0.0277	0.0682	0.1031	0.1459	0.1846	0.2488	0.3072	0.3744	0.4344	0.4996	0.5701	**0.6463**	0.6302	0.6154
infection(X)	0.3403	0.3604	0.3813	0.4031	0.4259	0.4498	0.4749	0.5011	0.5287	0.5576	0.5881	0.6201	**0.6539**	0.6256	0.5940
coronavirus(X)	**0.5882**	0.5638	0.5341	0.5030	0.4676	0.4687	0.4363	0.3936	0.3515	0.3086	0.2735	0.2363	0.1983	0.1478	0.0961
disinfection(X)	0.4576	0.4673	0.4773	0.4878	0.4988	0.5721	0.5861	0.6009	0.6163	0.6326	0.6496	0.6676	0.6865	0.7064	**0.7275**
nucleic acid(X)	**0.3942**	0.3783	0.3501	0.3293	0.3098	0.2639	0.2158	0.1606	0.0999	0.0594	0.0192	0.0242	0.0857	0.1396	0.2147
lockdown of the city(X)	0.4971	0.5059	0.5152	0.5251	0.5354	0.5463	0.5578	0.5699	0.5827	0.5963	0.6108	0.6261	0.6424	0.6597	**0.6782**
Wuhan(X)	0.1067	0.1395	0.1970	0.2683	0.3303	0.3768	0.4520	0.5057	0.5848	0.6569	0.7289	0.7273	0.7289	**0.7615**	0.7441
Zhong Nanshan(X)	0.0239	0.0221	0.0634	0.1220	0.1898	0.2307	0.2979	0.3649	0.4449	0.5343	0.6186	0.7046	0.7180	**0.7318**	0.7312
epidemic(X)	0.1559	0.1086	0.0692	0.0176	0.0403	0.0709	0.1268	0.1961	0.2646	0.3342	0.4017	0.4715	0.5539	0.6477	**0.6544**
flu(X)	**0.3419**	0.3149	0.3038	0.2975	0.2790	0.2692	0.2494	0.2232	0.1842	0.1292	0.0844	0.0391	0.0113	0.0397	0.0285
diarrhea(X)	0.0815	0.0974	0.1068	0.1166	0.1268	0.1374	0.1486	0.1603	0.1725	0.1853	0.1988	0.2130	0.2279	0.2435	**0.2601**
stuffy nose(X)	0.1920	0.2169	0.2262	0.2241	0.2243	0.2242	0.2276	0.2720	0.3101	0.3226	0.3630	0.4424	0.4793	0.5386	**0.5674**
dry cough(X)	0.1349	0.1077	0.0578	0.0003	0.0574	0.0863	0.1484	0.2145	0.2822	0.3369	0.4039	**0.4683**	0.4383	0.4476	0.4478
NCP(X)	**0.1891**	0.1547	0.1584	0.1556	0.1509	0.1435	0.0920	0.0488	0.0081	0.0320	0.0367	0.0516	0.0699	0.0918	0.1180
close contact(X)	**0.7903**	0.7823	0.7610	0.7496	0.7445	0.7297	0.7093	0.6844	0.6566	0.6426	0.6329	0.6102	0.5829	0.5419	0.5206
suspected case(X)	**0.4277**	0.4017	0.3662	0.3356	0.2999	0.2842	0.2508	0.2001	0.1559	0.1066	0.0501	0.0080	0.0577	0.1159	0.1702
pneumonia(X)	0.1746	0.1335	0.0831	0.0286	0.0321	0.0633	0.1210	0.1950	0.2663	0.3408	0.4113	0.4862	0.5666	0.6396	**0.6414**
cough(X)	0.6119	0.6375	0.6639	0.6921	0.6925	0.6643	0.6619	0.6751	0.6875	0.7024	0.7388	0.7764	0.8002	0.8092	**0.8147**
dyspnea(X)	0.0657	0.0274	0.0099	0.0545	0.0909	0.1042	0.1303	0.1583	0.1396	0.1095	0.1847	0.2433	0.2901	0.3171	**0.3435**
virus pneumonia(X)	0.5793	0.5869	0.5720	0.5777	0.6113	0.6111	0.6516	0.6720	0.7001	**0.7208**	0.7077	0.7042	0.6895	0.6790	0.6678
novel coronavirus(X)	**0.6381**	0.6109	0.5864	0.5633	0.5308	0.5324	0.5033	0.4638	0.4251	0.3856	0.3540	0.3214	0.2883	0.2394	0.1898

Optimal values are shown in bold in the table

Table 22 Correlation analysis between keywords and the cumulative number of confirmed cases—Suizhou

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.0234	0.0065	0.0544	0.1111	0.1690	0.2241	0.2908	0.3233	0.3937	0.4328	**0.4743**	0.4503	0.4558	0.4691	0.4639
Red Cross Society(S)	0.1276	0.0963	0.0423	0.0064	0.0581	0.1154	0.1709	0.1988	0.2564	0.3324	0.4023	0.4870	0.5575	0.6432	**0.6811**
Spring Festival(S)	0.5994	0.6145	0.6554	0.6691	**0.6719**	0.6636	0.6428	0.6126	0.6119	0.6115	0.5988	0.5900	0.5829	0.5784	0.5602
Health Commission(S)	0.1771	0.1124	0.0456	0.0231	0.0803	0.1549	0.2120	0.2842	0.3695	0.4442	0.5185	0.5987	0.6794	0.7163	**0.7204**
vaccine(S)	0.2365	0.2008	0.1544	0.1163	0.0801	0.0484	0.0153	0.0193	0.0623	0.0934	0.1610	0.2173	0.2775	0.3569	**0.4135**
SARS(S)	0.1331	0.2013	0.2727	0.3487	0.4285	0.5129	0.6023	0.6961	0.7809	0.8615	0.9507	0.9569	0.9582	0.9625	**0.9663**
hand washing(S)	0.0815	0.0908	0.1005	0.1107	0.1213	0.1324	0.1439	0.1561	0.1688	0.1821	0.1961	0.2108	0.2262	0.2425	**0.2597**
children(S)	0.4765	0.5094	0.4824	0.4871	0.4796	0.4427	0.4822	0.4968	0.5260	0.5317	0.5130	0.5288	0.5454	**0.5598**	0.5330
mask(S)	0.0092	0.0349	0.1039	0.1619	0.2080	0.2723	0.3446	0.4060	0.4736	0.5418	0.6129	**0.7130**	0.6803	0.6635	0.6353
coronavirus(S)	0.0902	0.1530	0.2237	0.2999	0.3771	0.4558	0.5391	0.6248	0.7148	0.8177	0.9225	0.9272	0.9275	0.9356	**0.9385**
novel(S)	0.0154	0.0603	0.1227	0.1739	0.2406	0.2793	0.3558	0.4133	0.4742	0.5524	0.6359	0.7202	0.7763	**0.7852**	0.7617
disinfection(S)	0.0643	0.0149	0.0006	0.0522	0.1005	0.1426	0.1661	0.1909	0.2172	0.2450	0.2746	0.3060	**0.3395**	0.3238	0.2751
lockdown of the city(S)	0.0351	0.0152	0.0372	0.0604	0.1117	0.1661	0.1975	0.2308	0.2787	0.3173	0.3783	0.4351	0.4844	**0.5711**	0.5423
Wuhan(S)	0.0799	0.1428	0.2194	0.2936	0.3704	0.4486	0.5228	0.6045	0.6978	0.8012	0.9010	0.9050	**0.9057**	0.9042	0.8987
Zhong Nanshan(S)	0.0340	0.1022	0.1743	0.2474	0.3199	0.3899	0.4707	0.5600	0.6490	0.7400	0.8408	0.8695	**0.8760**	0.8665	0.8588
epidemic(S)	0.8918	**0.8980**	0.8756	0.8681	0.8707	0.8709	0.8708	0.8736	0.8798	0.8649	0.8444	0.8241	0.8153	0.7880	0.7724
flu(S)	0.0004	0.0582	0.1227	0.1983	0.2755	0.3332	0.3966	0.4788	0.5523	0.6457	0.7131	0.7996	**0.8527**	0.8469	0.8373
diarrhea(S)	0.0849	0.0525	0.0118	0.0668	0.1217	0.1661	0.2129	0.2624	0.3461	0.4042	0.4731	0.5310	0.6083	0.6509	**0.6803**
dry cough(S)	0.1317	0.1864	0.2443	0.3119	0.3510	0.4133	0.4393	0.4871	0.5702	0.6573	0.7209	**0.8028**	0.8004	0.7887	0.7986
NCP(S)	**0.5722**	0.5443	0.5164	0.4869	0.4582	0.4290	0.3954	0.3668	0.3306	0.2957	0.2681	0.2551	0.2411	0.2060	0.1634
novel coronary pneumonia(S)	0.8897	0.8942	0.8904	0.8929	**0.8949**	0.8921	0.8817	0.8632	0.8526	0.8427	0.8323	0.8347	0.8367	0.8156	0.8121
COVID-19(S)	**0.6989**	0.6770	0.6695	0.6449	0.6290	0.6059	0.5877	0.5753	0.5623	0.5689	0.5424	0.5122	0.4796	0.4531	0.4499
novel coronavirus pneumonia(S)	**0.4262**	0.3841	0.3416	0.2982	0.2718	0.2253	0.1850	0.1435	0.0952	0.0467	0.0131	0.0703	0.1366	0.2186	0.2251
asymptomatic infection(S)	**0.6322**	0.6204	0.6049	0.5764	0.5391	0.5053	0.4744	0.4331	0.3840	0.3486	0.2977	0.2559	0.1991	0.1570	0.1119
aerosol transmission(S)	**0.2804**	0.2466	0.2083	0.1680	0.1312	0.1197	0.1076	0.0739	0.0251	0.0132	0.0336	0.0746	0.1238	0.1515	0.1810
pneumonia(S)	0.0147	0.0745	0.1477	0.2191	0.2915	0.3697	0.4394	0.5164	0.6031	0.6967	0.8011	0.9065	0.9109	0.9168	**0.9184**
cough(S)	0.1429	0.0848	0.0323	0.0164	0.0655	0.1307	0.1949	0.2773	0.3426	0.4185	0.4879	0.5681	0.5785	**0.6560**	0.6450
fever(S)	0.1906	0.2208	0.2218	0.2661	0.3197	0.3829	0.4263	0.4724	0.5216	0.5532	0.5866	0.6540	**0.6722**	0.6678	0.6452
dyspnea(S)	0.1094	0.1653	0.2203	0.2501	0.2981	0.3645	0.4031	0.4440	0.5105	0.5585	0.6096	**0.6641**	0.6407	0.6294	0.6120
high temperature(S)	0.1954	0.2376	0.2642	0.3167	0.3647	0.4150	0.4752	0.5152	0.5575	0.6023	0.6651	0.6743	**0.6749**	0.6476	0.6175
novel coronavirus(S)	0.0750	0.0198	0.0454	0.1194	0.1848	0.2586	0.3398	0.4209	0.5035	0.6002	0.7008	0.8028	0.8986	0.9413	**0.9447**
SARS-CoV-2(S)	0.2926	0.2986	**0.3049**	0.2635	0.2677	0.2721	0.2304	0.2324	0.1848	0.1837	0.1826	0.1815	0.1802	0.1377	0.1332
Spring Festival(X)	0.7865	0.7714	0.7835	0.7945	**0.7969**	0.7902	0.7922	0.7841	0.7673	0.7648	0.7623	0.7459	0.7232	0.7102	0.6950
Red Cross Society(X)	0.2236	0.2067	0.1535	0.0928	0.0302	0.0254	0.0760	0.1380	0.1805	0.2490	0.3241	0.4104	0.4992	**0.5914**	0.5882
government(X)	0.2938	0.3135	0.2910	0.2557	0.2565	0.2542	0.3038	0.3565	0.3897	**0.4053**	0.3956	0.3650	0.3647	0.4021	0.3942

Table 22 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0424	0.1068	0.1429	0.2116	0.2635	0.3265	0.3618	0.4493	0.5091	0.5993	0.6986	0.7289	0.7586	**0.7733**	0.7553
medical(X)	0.2156	0.1593	0.0975	0.0376	0.0233	0.0721	0.1206	0.1704	0.2442	0.3140	0.3760	0.4393	0.5076	**0.5436**	0.5345
vaccine(X)	0.2767	0.2264	0.1744	0.1432	0.1100	0.0733	0.0384	0.0135	0.0617	0.1008	0.1645	0.2379	0.3220	0.3618	**0.3648**
SARS(X)	0.0543	0.1152	0.1820	0.2560	0.3318	0.4028	0.4643	0.5285	0.5878	0.6728	0.7610	0.7821	**0.7843**	0.7714	0.7592
children(X)	0.3368	0.3600	0.3844	0.4099	0.4596	0.4889	0.5197	0.5521	0.5862	0.5641	0.5994	**0.6368**	0.6113	0.5819	0.5534
mask(X)	0.0845	0.0673	0.0085	0.0501	0.0935	0.1343	0.1755	0.2198	0.2606	0.3270	0.3905	0.4616	**0.5274**	0.5136	0.4767
infection(X)	0.3254	0.3445	0.3643	0.3850	0.4067	0.4293	0.4529	0.4777	0.5037	0.5310	0.5596	0.5898	**0.6215**	0.5885	0.5515
coronavirus(X)	0.6851	**0.6999**	0.6662	0.6332	0.6067	0.5765	0.5447	0.5075	0.4625	0.4275	0.3824	0.3390	0.2905	0.2458	0.1984
disinfection(X)	0.4760	0.4863	0.4970	0.5082	0.5199	0.5322	0.5450	0.5584	0.5724	0.5872	0.6027	0.6190	0.6361	0.6542	**0.6732**
nucleic acid(X)	0.4747	**0.4788**	0.4439	0.4052	0.3879	0.3619	0.3378	0.3191	0.3082	0.2635	0.2104	0.1516	0.0962	0.0515	0.0090
lockdown of the city(X)	0.5136	0.5239	0.5347	0.5460	0.5578	0.5702	0.5833	0.5970	0.6114	0.6070	0.6218	0.6375	0.6541	0.6716	**0.6903**
Wuhan(X)	0.0314	0.0200	0.0302	0.0896	0.1467	0.2205	0.2970	0.3582	0.4081	0.4848	0.5447	0.5458	0.5565	**0.6137**	0.5984
Zhong Nanshan(X)	0.1943	0.1313	0.0665	0.0003	0.0457	0.0896	0.1510	0.2272	0.2729	0.3443	0.4157	0.5058	0.5810	**0.5983**	0.5959
epidemic(X)	**0.5147**	0.4737	0.4293	0.3832	0.3426	0.3000	0.2527	0.1985	0.1800	0.1275	0.0651	0.0107	0.0538	0.1208	0.1878
flu(X)	0.4335	**0.4382**	0.3982	0.3557	0.3322	0.3141	0.3016	0.2872	0.2789	0.2618	0.2382	0.1998	0.1407	0.0902	0.1072
diarrhea(X)	0.0170	0.0241	0.0315	0.0392	0.0472	0.0556	0.0644	0.0737	0.0833	0.0935	0.1041	0.1153	0.1271	0.1395	**0.1526**
stuffy nose(X)	0.1693	0.1045	0.0692	0.0169	0.0017	0.0327	0.0296	0.0658	0.0925	0.1160	0.1716	0.2092	0.2771	0.3455	**0.4304**
dry cough(X)	0.2348	0.1833	0.1334	0.0788	0.0181	0.0485	0.1188	0.1901	0.2333	0.3076	0.3859	**0.4705**	0.4399	0.4426	0.4277
NCP(X)	**0.2957**	0.2629	0.2310	0.2299	0.2288	0.1962	0.1932	0.1594	0.1542	0.1097	0.0651	0.0215	0.0257	0.0435	0.0623
close contact(X)	**0.8517**	0.8320	0.8133	0.7973	0.7918	0.7718	0.7544	0.7486	0.7339	0.7111	0.6839	0.6525	0.6352	0.6212	0.5972
suspected case(X)	**0.5151**	0.4743	0.4298	0.3845	0.3537	0.3055	0.2687	0.2260	0.2043	0.1638	0.1078	0.0516	0.0093	0.0366	0.0998
pneumonia(X)	**0.5872**	0.5513	0.5122	0.4713	0.4380	0.3993	0.3590	0.3125	0.3018	0.2592	0.2060	0.1547	0.0976	0.0413	0.0152
cough(X)	0.2399	0.2551	0.3295	0.4071	0.4057	0.4007	0.4178	0.4423	0.4500	0.4686	0.5506	0.6378	0.6779	0.6954	**0.6996**
dyspnea(X)	**0.2556**	0.2045	0.1579	0.1067	0.0708	0.0348	0.0004	0.0327	0.0246	0.0308	0.0698	0.1157	0.1754	0.2239	0.2444
virus pneumonia(X)	0.5305	0.5528	0.5601	0.5646	0.5768	0.5915	0.5839	0.6021	0.6111	0.6604	**0.6632**	0.6508	0.6316	0.6125	0.5884
novel coronavirus(X)	**0.7286**	0.6994	0.6665	0.6337	0.6073	0.5767	0.5448	0.5073	0.5034	0.4704	0.4281	0.3886	0.3434	0.2998	0.2536

Optimal values are shown in bold in the table

Table 23 Correlation analysis between keywords and the cumulative number of confirmed cases—Tianmen

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	**0.2928**	0.2216	0.1864	0.1525	0.1500	0.1066	0.0634	0.0522	0.0524	0.1141	0.0649	0.0082	0.0518	0.0098	0.0426
Red Cross Society(S)	0.1973	0.1569	0.1089	0.0636	0.0139	0.0697	0.1003	0.1619	0.2320	0.2701	0.3360	0.3672	0.4324	0.4879	**0.5608**
Spring Festival(S)	0.4613	0.4748	0.6133	0.6928	0.6996	0.6962	**0.7007**	0.6880	0.6929	0.6484	0.6735	0.6485	0.6818	0.6196	0.6566
Health Commission(S)	0.1361	0.0699	0.0465	0.0229	0.0618	0.1125	0.1891	0.2197	0.2775	0.3638	0.4322	0.5366	0.6154	**0.6231**	0.5974
vaccine(S)	0.1632	0.1196	0.0793	0.0782	0.0680	0.0121	0.0754	0.0403	0.1334	0.1069	0.2408	0.2742	0.3098	**0.3874**	0.3725
SARS(S)	0.0881	0.1480	0.2128	0.2926	0.3649	0.4298	0.5038	0.6000	0.7422	0.7693	0.8129	0.8198	**0.8745**	0.8728	0.8146
hand washing(S)	0.1451	0.1436	0.0074	0.1703	0.1899	0.1863	0.0626	0.2074	0.2291	0.2433	0.2581	0.2737	0.2901	**0.3073**	0.2235
children(S)	0.2288	0.2790	0.2337	0.2942	0.2715	0.2538	0.3234	0.2301	0.2167	0.2084	0.2100	0.2829	0.3080	**0.3929**	0.3440
mask(S)	0.2317	0.1382	0.1410	0.0495	0.0078	0.0583	0.0974	0.1626	0.2496	0.3294	0.3573	0.4062	**0.4652**	0.3915	0.3736
coronavirus(S)	0.0438	0.1123	0.1656	0.2463	0.3259	0.4111	0.4721	0.5706	0.6488	0.7431	0.8382	0.8668	0.8744	**0.9369**	0.8945
novel(S)	0.1238	0.0791	0.0611	0.0226	0.1555	0.1043	0.2058	0.1922	0.2825	0.2884	0.3411	0.3553	**0.4592**	0.3638	0.4125
disinfection(S)	0.0474	0.0446	0.0185	0.0690	0.0069	0.1177	0.0932	0.1824	0.2029	0.2642	0.2921	**0.3019**	0.2444	0.2050	0.2249
lockdown of the city(S)	0.0387	0.0152	0.0112	0.0700	0.1520	0.2197	0.2235	0.2713	0.2982	0.3434	0.3728	0.4135	0.5576	**0.6369**	0.5978
Wuhan(S)	0.0104	0.0776	0.1595	0.2354	0.2978	0.3770	0.4667	0.5360	0.6387	0.7244	0.8270	0.8773	**0.8817**	0.8802	0.8639
Zhong Nanshan(S)	0.0177	0.0603	0.1250	0.2011	0.2718	0.3546	0.4313	0.4931	0.5906	0.6543	0.7409	0.7525	0.7866	**0.8038**	0.7503
epidemic(S)	0.8447	**0.8601**	0.8272	0.8323	0.7863	0.7798	0.7790	0.7810	0.7659	0.8212	0.8501	0.8073	0.7216	0.7226	0.6933
flu(S)	0.0533	0.0370	0.0518	0.0753	0.1355	0.2419	0.2926	0.3413	0.4231	0.5413	0.6183	0.6689	0.6898	0.7547	**0.7777**
diarrhea(S)	0.1399	0.0676	0.0039	0.0516	0.1053	0.1264	0.2307	0.3264	0.3709	0.3524	0.4332	0.4936	0.5833	**0.6666**	0.5943
dry cough(S)	0.1126	0.2082	0.2316	0.2967	0.3825	0.4543	0.5418	0.5610	0.5970	0.6910	0.7599	0.7991	0.8171	0.8369	**0.8553**
NCP(S)	**0.4385**	0.4075	0.3689	0.3301	0.3354	0.4043	0.3758	0.3473	0.3447	0.3198	0.2659	0.1571	0.2196	0.1843	0.1278
novel coronary pneumonia(S)	0.8600	0.8742	**0.8817**	0.8598	0.8295	0.8300	0.8062	0.8072	0.7794	0.7743	0.7944	0.8256	0.8219	0.8176	0.7514
COVID-19(S)	0.7333	**0.7510**	0.7293	0.7155	0.6921	0.6678	0.6549	0.6429	0.6391	0.7079	0.6692	0.6799	0.6496	0.5719	0.5302
novel coronavirus pneumonia(S)	**0.4534**	0.4199	0.3765	0.3285	0.2682	0.2578	0.2154	0.1687	0.1087	0.0790	0.0378	0.0345	0.1498	0.1979	0.2032
asymptomatic infection(S)	0.5315	**0.5446**	0.4972	0.4781	0.4804	0.4414	0.4492	0.4597	0.4002	0.4009	0.2792	0.2903	0.2199	0.2217	0.1358
aerosol transmission(S)	0.1551	0.1220	**0.1704**	0.1199	0.0294	0.1241	0.1189	0.1069	0.0930	0.0507	0.0452	0.0953	0.1225	0.1437	0.1660
pneumonia(S)	0.0867	0.0283	0.0344	0.1107	0.1769	0.2673	0.3431	0.4196	0.5008	0.6010	0.6645	0.7714	0.8740	**0.8900**	0.8873
cough(S)	0.0743	0.0178	0.0279	0.0840	0.1985	0.2349	0.2316	0.2640	0.3677	0.4427	0.4871	0.5827	0.6789	0.7018	**0.7411**
fever(S)	0.1159	0.1317	0.1998	0.2708	0.3245	0.4555	0.4349	0.5531	0.6527	0.6673	0.7095	0.7089	0.7388	0.7269	**0.7604**
dyspnea(S)	0.0275	0.0533	0.1730	0.1505	0.1727	0.2453	0.2762	0.3189	0.3236	0.3913	0.4705	0.5498	0.6035	**0.6149**	0.5542
high temperature(S)	0.2033	0.2260	0.2428	0.3174	0.4294	0.4576	0.4067	0.4762	0.5886	0.5895	0.6680	0.6725	**0.7196**	0.7016	0.7033
novel coronavirus(S)	0.1206	0.0518	0.0020	0.0544	0.1220	0.2106	0.2914	0.3828	0.4715	0.5727	0.6340	0.7525	0.8682	0.9046	**0.9141**
SARS-CoV-2(S)	0.2915	0.2975	**0.3038**	0.2585	0.2109	0.1691	0.1680	0.1173	0.1030	0.0676	0.1873	0.1897	0.1237	0.0153	0.0967
Spring Festival(X)	0.7088	**0.7924**	0.7021	0.6977	0.7308	0.7283	0.7073	0.7195	0.7920	0.7628	0.7014	0.7041	0.6874	0.6736	0.6225
Red Cross Society(X)	0.2849	0.2339	0.1730	0.1090	0.0303	0.0522	0.1014	0.1642	0.2189	0.2624	0.3354	0.3986	0.5132	0.6410	**0.6515**
government(X)	0.3842	0.2799	0.3117	0.2734	0.1926	0.2583	0.3184	0.3122	0.3706	0.4060	0.3978	0.3937	0.2746	0.3251	**0.4409**

Table 23 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0530	0.0162	0.0893	0.1426	0.1915	0.3061	0.3448	0.4417	0.4619	0.5320	0.5995	0.6770	**0.7925**	0.7615	0.7333
medical(X)	0.2807	0.1074	0.1598	0.0878	0.0211	0.0817	0.1133	0.1808	0.1850	0.2751	0.3415	0.4257	0.5205	**0.6028**	0.5800
vaccine(X)	0.2265	0.2371	0.1968	0.1277	0.0610	0.0024	0.0368	0.0896	0.1878	0.1448	0.1797	0.2980	0.3984	**0.4427**	0.4394
SARS(X)	0.0235	0.1532	0.1161	0.1685	0.2639	0.3531	0.4174	0.5000	0.6445	0.6245	0.7080	0.7200	**0.7339**	0.7332	0.7275
children(X)	0.2838	0.4058	0.3821	0.4054	0.4297	0.4552	0.5099	0.5405	0.5696	0.6032	0.6418	**0.6782**	0.6451	0.5510	0.6519
mask(X)	0.1945	0.0245	0.0947	0.0538	0.0463	0.1414	0.1663	0.1723	0.2071	0.3494	0.3274	0.4069	0.5473	**0.5495**	0.5249
infection(X)	0.3221	0.2493	0.3540	0.2963	0.3988	0.4260	0.4496	0.4743	0.5002	0.5273	0.5558	0.5858	**0.6174**	0.5835	0.5454
coronavirus(X)	**0.7003**	0.6295	0.6707	0.6790	0.6237	0.6104	0.5846	0.5427	0.4008	0.4671	0.4450	0.3886	0.3017	0.2699	0.2156
disinfection(X)	0.4166	0.4255	0.4348	0.4445	0.4546	0.4652	0.4762	0.4878	0.4999	0.5126	0.5259	0.5400	0.5547	0.5703	**0.5867**
nucleic acid(X)	**0.5800**	0.4368	0.5041	0.5031	0.3893	0.4164	0.4186	0.4058	0.3809	0.2232	0.2658	0.2145	0.0641	0.0249	0.0313
lockdown of the city(X)	0.4402	0.4549	0.5296	0.5355	0.4892	0.4967	0.5076	0.5192	0.5313	0.5441	0.5576	0.5718	0.5868	0.6027	**0.6196**
Wuhan(X)	0.0965	0.0488	0.0052	0.0780	0.1742	0.2173	0.2840	0.4259	0.4910	0.5335	0.5820	0.5796	0.5944	0.6059	**0.6134**
Zhong Nanshan(X)	0.2355	0.0615	0.1396	0.0886	0.0039	0.0543	0.0828	0.1290	0.2001	0.3511	0.3113	0.3894	0.5315	**0.5608**	0.5118
epidemic(X)	**0.6251**	0.4791	0.5271	0.5139	0.4581	0.4089	0.3731	0.3326	0.2981	0.2848	0.2499	0.1769	0.0943	0.0278	0.0279
flu(X)	0.4264	**0.4350**	0.3762	0.3751	0.3803	0.3566	0.3060	0.2430	0.1742	0.2651	0.2650	0.2395	0.1427	0.0389	0.0303
diarrhea(X)	0.0968	0.0864	0.0992	0.1100	0.1276	0.1373	0.1293	0.0124	0.1453	0.1658	0.1775	0.1898	0.2027	0.2163	**0.2306**
stuffy nose(X)	0.2649	0.1739	0.2276	0.1834	0.1012	0.0701	0.0392	0.0193	0.1635	0.0912	0.1358	0.1843	0.2112	0.3142	**0.3597**
dry cough(X)	0.2411	0.1227	0.2115	0.1757	0.1181	0.0530	0.0001	0.0852	0.2637	0.3057	0.2946	**0.3687**	0.3326	0.3202	0.3116
NCP(X)	0.2160	0.2167	0.2134	0.2069	0.2121	**0.3185**	0.3076	0.2651	0.1607	0.1523	0.0988	0.0459	0.0144	0.0535	0.0313
close contact(X)	**0.7577**	0.7384	0.7552	0.7333	0.7061	0.6937	0.6765	0.7193	0.7312	0.6640	0.7344	0.6980	0.6349	0.6596	0.6276
suspected case(X)	**0.5847**	0.4578	0.5337	0.5146	0.4063	0.3794	0.3237	0.3415	0.2966	0.1903	0.2566	0.2012	0.0958	0.1013	0.0314
pneumonia(X)	0.5664	0.4719	**0.5779**	0.5631	0.5155	0.4857	0.4611	0.4144	0.2578	0.2520	0.3335	0.2815	0.1980	0.1634	0.1504
cough(X)	0.0838	0.1426	0.2345	0.3086	0.3522	0.3432	0.3430	0.3900	0.4535	0.4511	0.5281	0.6189	0.7512	0.7800	**0.7928**
dyspnea(X)	0.2718	0.2431	0.2257	0.2076	0.1426	0.0907	0.0627	0.0100	0.0606	0.0663	0.0638	0.0905	0.1697	0.2158	**0.2758**
virus pneumonia(X)	0.4841	0.6011	0.4848	0.5193	0.5658	0.5742	0.5678	0.5310	0.5845	**0.6720**	0.6010	0.6411	0.6459	0.6080	0.5825
novel coronavirus(X)	**0.6901**	0.6283	0.6523	0.6663	0.6253	0.5916	0.5628	0.5338	0.5097	0.5041	0.4762	0.4173	0.3380	0.3116	0.2577

Optimal values are shown in bold in the table

Table 24 Correlation analysis between keywords and the cumulative number of confirmed cases—Xiangyang

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.1518	0.1157	0.0678	0.0262	0.0403	0.1109	0.1951	0.2652	0.3318	0.3949	0.4745	**0.5200**	0.5011	0.4662	0.4571
Red Cross Society(S)	0.2356	0.1783	0.1275	0.0781	0.0192	0.0472	0.0836	0.1588	0.2199	0.3001	0.3775	0.4623	0.5604	0.6575	**0.7120**
Spring Festival(S)	0.7922	0.8042	0.8334	0.8526	0.8556	0.8592	**0.8683**	0.8492	0.8291	0.8257	0.8109	0.7829	0.7646	0.7475	0.7335
Health Commission(S)	0.0588	0.0087	0.0702	0.1450	0.2297	0.3080	0.3947	0.4726	0.5529	0.6446	0.7407	0.8324	0.8643	**0.8663**	0.8539
vaccine(S)	0.3382	0.2899	0.2443	0.1877	0.1303	0.0967	0.0599	0.0063	0.0523	0.1095	0.1511	0.1986	0.2706	0.3557	**0.3969**
SARS(S)	0.1218	0.1918	0.2626	0.3402	0.4223	0.5079	0.5984	0.6970	0.7969	0.8988	0.9490	0.9543	0.9585	0.9619	**0.9646**
hand washing(S)	**0.4566**	0.4147	0.3725	0.3515	0.3473	0.3244	0.2879	0.2437	0.2642	0.2378	0.1910	0.1805	0.1472	0.2220	0.2123
children(S)	**0.4075**	0.3623	0.3111	0.2652	0.3472	0.3183	0.3616	0.3479	0.3188	0.2981	0.2823	0.3319	0.3215	0.3248	0.2862
mask(S)	0.0042	0.0701	0.1382	0.2094	0.2857	0.3636	0.4427	0.5332	0.6155	0.7076	0.8130	**0.9149**	0.9004	0.8977	0.8901
coronavirus(S)	0.0752	0.1423	0.2127	0.2850	0.3655	0.4499	0.5384	0.6315	0.7284	0.8339	0.9446	0.9519	0.9548	0.9612	**0.9633**
novel(S)	0.0628	0.0019	0.0525	0.1027	0.1777	0.2545	0.3341	0.3974	0.4746	0.5698	0.6445	0.7450	0.7398	**0.7657**	0.7519
disinfection(S)	0.1339	0.0946	0.0505	0.0131	0.0806	0.1365	0.2076	0.2715	0.3354	0.4147	0.4766	0.5325	**0.6080**	0.5779	0.5455
lockdown of the city(S)	0.1600	0.1188	0.0608	0.0019	0.0693	0.1360	0.2102	0.2737	0.3340	0.4057	0.4730	0.5465	0.6328	**0.7153**	0.7086
Wuhan(S)	0.0572	0.1229	0.1932	0.2649	0.3515	0.4391	0.5261	0.6178	0.7124	0.8078	0.9126	**0.9249**	0.9212	0.9206	0.9214
Zhong Nanshan(S)	0.0213	0.0822	0.1440	0.2230	0.2999	0.3815	0.4700	0.5601	0.6475	0.7594	0.8781	**0.8895**	0.8817	0.8808	0.8823
epidemic(S)	**0.8915**	0.8829	0.8798	0.8753	0.8479	0.8215	0.8179	0.8130	0.8098	0.8175	0.8258	0.8040	0.7828	0.7590	0.7472
flu(S)	0.0793	0.1442	0.2122	0.2592	0.3238	0.4057	0.4929	0.5726	0.6578	0.7545	0.8357	0.8949	0.8978	0.8946	**0.9050**
diarrhea(S)	0.0994	0.0334	0.0207	0.0899	0.1626	0.1657	0.2361	0.3052	0.3766	0.4415	0.5184	0.6024	0.7006	**0.8014**	0.7756
dry cough(S)	0.0065	0.0654	0.1386	0.2045	0.2762	0.3631	0.4404	0.5195	0.6144	0.7185	0.7989	0.8514	0.8821	**0.8824**	0.8783
NCP(S)	**0.8058**	0.7822	0.7565	0.7297	0.7019	0.6729	0.6425	0.6145	0.5817	0.5450	0.5080	0.4763	0.4340	0.3931	0.3555
novel coronary pneumonia(S)	0.9053	0.9100	0.9136	**0.9164**	0.9000	0.8942	0.8967	0.8963	0.8771	0.8728	0.8744	0.8677	0.8595	0.8442	0.8254
COVID-19(S)	0.8169	0.8365	**0.8571**	0.8368	0.8182	0.7956	0.7720	0.7519	0.7352	0.7234	0.6954	0.6733	0.6652	0.6306	0.5993
novel coronavirus pneumonia(S)	0.0793	0.0149	0.0517	0.1220	0.1987	0.2756	0.3394	0.4199	0.5104	0.6012	0.6869	0.7838	0.8009	**0.8018**	0.7883
asymptomatic infection(S)	**0.7008**	0.6674	0.6476	0.6357	0.6090	0.5906	0.5719	0.5263	0.4840	0.4391	0.3918	0.3436	0.2976	0.2345	0.1669
aerosol transmission(S)	**0.6684**	0.6342	0.5979	0.5602	0.5264	0.5028	0.4625	0.4249	0.3846	0.3400	0.2845	0.2308	0.1718	0.1137	0.0614
pneumonia(S)	0.0000	0.0625	0.1299	0.1985	0.2835	0.3704	0.4551	0.5427	0.6356	0.7303	0.8343	0.9281	**0.9383**	0.9360	0.9349
cough(S)	0.0207	0.0485	0.1225	0.2038	0.2793	0.3414	0.4103	0.4931	0.5635	0.6236	0.7047	0.7088	**0.7387**	0.7384	0.7346
fever(S)	0.2455	0.2433	0.2848	0.2895	0.3367	0.3551	0.4317	0.4845	0.5593	0.6466	0.7250	0.8230	**0.8494**	0.8399	0.8342
dyspnea(S)	0.0057	0.0521	0.1236	0.1941	0.2664	0.3431	0.4142	0.4913	0.5557	0.6330	0.7337	0.8306	0.8186	0.8297	**0.8344**
high temperature(S)	0.0344	0.0207	0.0864	0.1621	0.2421	0.3134	0.3844	0.4747	0.5541	0.6359	**0.7396**	0.7240	0.7122	0.6998	0.7082
novel coronavirus(S)	0.0508	0.0136	0.0780	0.1437	0.2228	0.3026	0.3859	0.4742	0.5658	0.6621	0.7645	0.8763	0.9255	0.9518	**0.9552**
SARS-CoV-2(S)	0.4929	0.5035	0.5146	**0.5262**	0.5048	0.4722	0.4544	0.4325	0.4030	0.3811	0.3473	0.3101	0.2684	0.2328	0.2018
Spring Festival(X)	0.7912	**0.8003**	0.7877	0.7704	0.7771	0.7781	0.7747	0.7656	0.7661	0.7524	0.7273	0.7261	0.7145	0.6902	0.6609
Red Cross Society(X)	0.2085	0.1568	0.1037	0.0818	0.0197	0.0517	0.1217	0.1845	0.2499	0.2890	0.3031	0.3765	0.4664	0.5633	**0.5672**
government(X)	0.1556	0.1799	0.1497	0.1370	0.1323	0.1177	0.2041	0.2927	0.3081	0.2932	0.3367	0.3459	0.3250	0.3356	**0.3468**

Table 24 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0325	0.0775	0.1110	0.1799	0.2656	0.3344	0.3422	0.4095	0.4944	0.5908	0.6591	0.6632	0.7113	**0.7152**	0.6903
medical(X)	0.2129	0.1744	0.1050	0.0344	0.0408	0.1162	0.1808	0.2356	0.3021	0.3685	0.4582	0.5458	0.6212	**0.6475**	0.6403
vaccine(X)	0.1675	0.1307	0.1181	0.1049	0.0501	0.0147	0.0184	0.0603	0.0987	0.1543	0.1902	0.2333	0.2950	**0.3311**	0.3165
SARS(X)	0.0584	0.1181	0.1839	0.2566	0.3365	0.4167	0.4844	0.5511	0.6368	0.7250	0.8026	**0.8329**	0.8306	0.8244	0.8188
children(X)	0.4808	0.4642	0.4410	0.4727	0.4899	0.5256	0.4981	0.4709	0.4629	0.4984	0.5591	**0.6196**	0.5874	0.5530	0.5401
mask(X)	0.1026	0.0549	0.0049	0.0286	0.0937	0.1617	0.2066	0.2529	0.3015	0.3528	0.4024	0.4798	**0.5547**	0.5376	0.4995
infection(X)	0.3829	0.4054	0.4289	0.4534	0.4790	0.5058	0.5338	0.5632	0.5940	0.6263	0.6603	0.6961	**0.7337**	0.7100	0.6841
coronavirus(X)	**0.7232**	0.6935	0.6570	0.6196	0.5773	0.5356	0.5007	0.4610	0.4187	0.3704	0.3611	0.3203	0.2621	0.2098	0.1567
disinfection(X)	0.4472	0.4565	0.4663	0.4764	0.4871	0.5508	0.5640	0.5779	0.5924	0.6077	0.6237	0.6406	0.6583	0.6770	**0.6967**
nucleic acid(X)	**0.5088**	0.4652	0.4713	0.4351	0.3897	0.3467	0.3300	0.3025	0.2746	0.2567	0.2454	0.1947	0.1318	0.0639	0.0085
lockdown of the city(X)	0.4611	0.4677	0.4765	0.4838	0.4922	0.5011	0.5105	0.5204	0.5309	0.5421	0.5539	0.5473	0.5425	0.5556	**0.5692**
Wuhan(X)	0.0137	0.0172	0.0660	0.0957	0.1618	0.2333	0.3030	0.3767	0.4635	0.5358	0.6131	0.6193	0.6187	**0.6695**	0.6495
Zhong Nanshan(X)	0.1415	0.0822	0.0133	0.0100	0.0821	0.1608	0.2138	0.2650	0.3363	0.4247	0.4815	0.5653	**0.5838**	0.5806	0.5801
epidemic(X)	**0.4563**	0.4146	0.3660	0.3636	0.3107	0.2514	0.2025	0.1500	0.0904	0.0247	0.0047	0.0630	0.1443	0.2174	0.2805
flu(X)	**0.4386**	0.4222	0.4261	0.3838	0.3346	0.2894	0.2581	0.2363	0.2195	0.1988	0.1846	0.1646	0.1328	0.0882	0.0917
diarrhea(X)	**0.2397**	0.1957	0.1496	0.1483	0.1468	0.1032	0.0966	0.0936	0.0864	0.0801	0.0735	0.0665	0.0591	0.0513	0.0429
stuffy nose(X)	0.1962	0.1517	0.1107	0.0801	0.0236	0.0347	0.0691	0.1347	0.1883	0.2335	0.2752	0.3412	0.3950	0.4478	**0.5084**
dry cough(X)	0.2573	0.2052	0.1502	0.1352	0.0799	0.0188	0.0461	0.1190	0.1957	0.2709	0.3190	**0.4012**	0.3776	0.3828	0.3754
NCP(X)	**0.3854**	0.3502	0.3135	0.2792	0.2458	0.2130	0.2091	0.1748	0.1394	0.1325	0.1248	0.0745	0.0253	0.0219	0.0716
close contact(X)	0.8166	0.8095	**0.8278**	0.8067	0.7820	0.7656	0.7587	0.7356	0.7104	0.6935	0.7031	0.6779	0.6443	0.6108	0.5900
suspected case(X)	**0.5657**	0.5312	0.4916	0.4455	0.3937	0.3409	0.3029	0.2474	0.2006	0.1509	0.1230	0.0761	0.0085	0.0557	0.1057
pneumonia(X)	**0.5407**	0.5038	0.5126	0.4707	0.4198	0.3725	0.3338	0.2873	0.2373	0.1819	0.1630	0.1145	0.0454	0.0187	0.0029
cough(X)	0.2886	0.3416	0.4142	0.4419	0.4504	0.4598	0.4508	0.4600	0.4977	0.5228	0.5774	0.6638	0.7109	0.7243	**0.7386**
dyspnea(X)	**0.2395**	0.2049	0.1558	0.0987	0.0463	0.0111	0.0481	0.0879	0.0636	0.0441	0.0753	0.1226	0.1707	0.2159	0.2191
virus pneumonia(X)	0.5558	0.5838	0.5490	0.5317	0.5673	0.5799	0.5762	0.5742	0.6000	0.6326	0.6329	**0.6429**	0.6379	0.6225	0.5946
novel coronavirus(X)	**0.6971**	0.6670	0.6771	0.6891	0.6527	0.6152	0.5889	0.5557	0.5203	0.4802	0.4764	0.4430	0.3929	0.3475	0.3020

Optimal values are shown in bold in the table

Table 25 Correlation analysis between keywords and the cumulative number of confirmed cases—Xianning

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.1306	0.0368	0.0047	0.0347	0.0408	0.0611	0.1314	0.2077	0.2949	0.3177	**0.4007**	0.2970	0.2877	0.3280	0.3368
Red Cross Society(S)	0.4881	0.4365	0.3791	0.3101	0.2717	0.2140	0.1405	0.0358	0.0587	0.1095	0.1936	0.2161	0.3561	0.4848	**0.5517**
Spring Festival(S)	0.5408	0.5754	0.6284	0.6840	0.7160	0.7341	0.7396	0.7162	0.7071	0.6904	0.6892	**0.7657**	0.7068	0.7583	0.7361
Health Commission(S)	0.2971	0.1916	0.1133	0.0273	0.0038	0.0426	0.1264	0.2297	0.2779	0.3497	0.3818	0.4273	0.5281	0.6279	**0.7132**
vaccine(S)	0.3997	0.3951	0.4460	**0.4538**	0.3718	0.3359	0.1832	0.1068	0.0631	0.0800	0.1205	0.0560	0.0692	0.0888	0.0497
SARS(S)	0.0942	0.0176	0.0598	0.1484	0.2436	0.3216	0.4067	0.5013	0.5937	0.7091	0.7696	0.7772	0.7860	0.7914	**0.7971**
hand washing(S)	**0.3271**	0.2872	0.2866	0.1675	0.1653	0.1114	0.1068	0.1019	0.1172	0.0542	0.0338	0.0808	0.0913	0.1054	0.0493
children(S)	0.4579	0.3781	0.3332	0.3376	0.3820	**0.4592**	0.3823	0.2932	0.2975	0.2688	0.3054	0.2839	0.2808	0.3005	0.2717
mask(S)	0.2326	0.1771	0.0849	0.0504	0.0715	0.1560	0.2334	0.3347	0.4102	0.5113	0.5285	0.6026	0.6234	0.5902	**0.6712**
coronavirus(S)	0.1649	0.0762	0.0041	0.0818	0.1513	0.2194	0.3114	0.3934	0.5065	0.6270	0.7207	0.7662	0.7660	0.7649	**0.7836**
novel(S)	0.2660	0.2453	0.2289	0.1710	0.1195	0.0502	0.0794	0.1176	0.1396	0.1950	0.2303	0.3174	**0.3985**	0.3625	0.3218
disinfection(S)	0.1238	0.0651	0.0132	0.0187	0.0710	0.0497	0.0667	0.1494	0.2421	0.2636	0.3164	0.2876	**0.3770**	0.3688	0.2566
lockdown of the city(S)	0.2303	0.1874	0.1401	0.1384	0.1087	0.0800	0.0248	0.0795	0.2505	0.3789	0.4227	0.5335	0.5833	**0.6276**	0.5573
Wuhan(S)	0.1657	0.0965	0.0250	0.0424	0.1244	0.2264	0.3373	0.4464	0.5483	0.6398	0.6982	0.7143	0.7050	0.7454	**0.7666**
Zhong Nanshan(S)	0.1797	0.0947	0.0010	0.0960	0.1502	0.2331	0.3123	0.4151	0.5213	0.5921	0.6894	0.6894	0.7134	**0.7649**	0.7072
epidemic(S)	0.7211	0.6953	0.6831	0.6973	0.7503	0.7743	**0.7890**	0.7395	0.6902	0.6637	0.7175	0.7841	0.7875	0.7577	0.6053
flu(S)	0.1930	0.1373	0.0474	0.0194	0.0824	0.1480	0.2215	0.3311	0.4525	0.5472	0.6490	0.6991	0.7129	**0.7478**	0.7403
diarrhea(S)	0.2217	0.1804	0.1547	0.1190	0.0307	0.0195	0.0989	0.1507	0.2167	0.2900	0.3375	0.4848	0.5439	0.6738	**0.7222**
dry cough(S)	0.0496	0.0001	0.0240	0.0558	0.1266	0.2268	0.2924	0.4005	0.4596	0.5097	0.5929	0.6266	**0.7140**	0.7024	0.6511
NCP(S)	0.5631	0.5747	0.5353	0.4982	0.4256	0.4993	0.5684	0.6392	**0.7070**	0.6650	0.6348	0.5819	0.5367	0.4889	0.4432
novel coronary pneumonia(S)	0.7155	0.6980	0.6993	0.6887	0.6995	0.7161	0.7506	0.7328	0.7462	0.7669	0.7742	**0.7944**	0.7754	0.7895	0.7049
COVID-19(S)	0.5149	0.5269	0.5395	0.5526	0.5664	0.5809	0.5453	0.5411	0.5101	0.5924	0.6211	0.7185	**0.8142**	0.7419	0.7202
novel coronavirus pneumonia(S)	0.4487	0.3869	0.3349	0.2836	0.2183	0.1640	0.0923	0.0061	0.0947	0.1899	0.2889	0.4054	0.4875	**0.5104**	0.5084
asymptomatic infection(S)	0.5238	0.5746	**0.5890**	0.4824	0.4802	0.4526	0.4604	0.5174	0.4642	0.3666	0.2987	0.2927	0.2544	0.2092	0.1134
aerosol transmission(S)	0.3587	0.3991	0.4292	**0.4756**	0.4299	0.4256	0.4154	0.3877	0.3834	0.3259	0.2598	0.2050	0.1466	0.0657	0.0740
pneumonia(S)	0.2345	0.1506	0.0783	0.0051	0.0672	0.1429	0.2332	0.3385	0.4549	0.5617	0.6392	0.7353	0.7354	0.7249	**0.7544**
cough(S)	0.1451	0.0844	0.0540	0.0598	0.1175	0.2169	0.2777	0.3764	0.3336	0.4055	0.4126	0.4723	0.5131	0.5681	**0.6108**
fever(S)	0.1246	0.0443	0.0212	0.0666	0.1369	0.1717	0.2834	0.2319	0.2514	0.3765	0.4652	**0.5328**	0.4885	0.4969	0.4852
dyspnea(S)	0.0370	0.0712	0.1087	0.1451	0.1903	0.1700	0.2337	0.3791	0.4462	0.5833	0.6387	0.6465	0.6382	**0.6877**	0.6788
high temperature(S)	0.0659	0.0824	0.0119	0.0121	0.0993	0.2111	0.2253	0.2979	0.3135	0.3365	0.3302	0.4056	0.4167	0.4273	**0.4350**
novel coronavirus(S)	0.2912	0.2300	0.1663	0.1100	0.0642	0.0170	0.0933	0.2052	0.3399	0.4120	0.4766	0.5592	0.5985	0.6586	**0.7014**
SARS-CoV-2(S)	0.1843	0.1880	0.1918	0.1958	0.1999	0.2042	0.2087	0.2134	0.1586	0.1641	0.1755	0.2400	0.2422	0.1875	**0.2473**
Spring Festival(X)	0.5344	0.6303	0.6696	0.6795	**0.7244**	0.6648	0.6330	0.6581	0.6492	0.7095	0.7036	0.7094	0.7239	0.6518	0.6588
Red Cross Society(X)	**0.5348**	0.4053	0.3594	0.3184	0.2533	0.2956	0.2555	0.1559	0.0509	0.1112	0.1411	0.1963	0.2455	0.2185	0.1823
government(X)	0.2896	0.1951	0.1725	0.1482	0.0492	0.1478	0.1995	0.2311	0.2212	0.2453	0.2466	0.2344	0.2034	0.2200	**0.3133**

Table 25 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.2383	0.0493	0.0108	0.0224	0.0613	0.0805	0.1250	0.2747	0.3639	0.3930	0.3980	0.3826	0.5333	0.5954	**0.6321**
medical(X)	0.4103	0.2620	0.2471	0.1944	0.1002	0.1547	0.0523	0.1039	0.2459	0.2738	0.2857	0.2951	0.2946	0.4494	**0.5393**
vaccine(X)	0.3158	0.1677	0.1963	0.2095	0.1467	0.1534	0.0143	0.1127	0.1904	0.1802	0.1880	0.2582	0.3444	0.4164	**0.4537**
SARS(X)	0.1427	0.0287	0.0670	0.0850	0.1557	0.1400	0.2332	0.3525	0.4711	0.5593	**0.6385**	0.6338	0.6216	0.5954	0.6113
children(X)	0.2428	0.3023	0.3840	0.4625	0.5935	0.6253	0.6084	0.5917	0.6579	0.6918	**0.7244**	0.6966	0.6749	0.6502	0.6332
mask(X)	0.3208	0.1648	0.1143	0.0700	0.0352	0.0231	0.0974	0.2108	0.3631	0.4432	0.4342	0.4765	0.5116	0.4703	**0.5358**
infection(X)	0.0142	0.0426	0.0626	0.1684	0.2786	0.3884	0.4100	0.4800	0.5033	0.5235	0.4563	0.4868	0.5189	0.4767	**0.5274**
coronavirus(X)	0.5655	0.4988	0.5202	0.5886	0.6067	**0.7312**	0.7309	0.6734	0.5173	0.5088	0.4924	0.4627	0.5102	0.4977	0.4364
disinfection(X)	0.2867	0.2927	0.2989	0.3054	0.3121	0.3192	0.3265	0.3342	0.3423	0.3507	0.3596	0.3688	0.3786	0.3889	**0.3997**
nucleic acid(X)	**0.5996**	0.5274	0.5451	0.5150	0.3974	0.4374	0.4380	0.3967	0.4235	0.4125	0.3675	0.3151	0.1883	0.1282	0.0009
lockdown of the city(X)	0.3143	0.3252	0.3992	0.4384	0.4443	0.4528	**0.4547**	0.4040	0.4103	0.4170	0.3739	0.3821	0.3908	0.4001	0.4100
Wuhan(X)	0.2322	0.1044	0.0467	0.0452	0.1408	0.0780	0.1234	0.2340	0.3789	0.5385	**0.6379**	0.6258	0.5659	0.4850	0.4538
Zhong Nanshan(X)	0.3753	0.2128	0.1652	0.1154	0.0138	0.0858	0.0576	0.0363	0.2108	0.3517	0.3695	**0.4458**	0.4457	0.3802	0.4057
epidemic(X)	**0.6177**	0.5192	0.5382	0.5338	0.4925	0.5424	0.5126	0.4431	0.4026	0.3756	0.3506	0.3079	0.2137	0.1669	0.0765
flu(X)	0.4444	0.3139	0.2567	0.2807	0.3588	0.4898	**0.4986**	0.4092	0.2286	0.1812	0.1864	0.2335	0.2703	0.1615	0.0930
diarrhea(X)	0.2681	0.2694	**0.3842**	0.2474	0.0454	0.0224	0.1346	0.0121	0.1000	0.0855	0.0704	0.0605	0.1900	0.2053	0.2214
stuffy nose(X)	0.2962	0.1611	0.0991	0.0673	0.0013	0.0774	0.1426	0.0969	0.0842	0.0807	0.1207	0.1734	0.2211	0.1798	**0.3046**
dry cough(X)	**0.4662**	0.3283	0.3053	0.2759	0.2112	0.2586	0.1994	0.1241	0.0500	0.0307	0.1354	0.2489	0.2290	0.2014	0.2047
NCP(X)	0.2455	0.2505	0.2022	0.1512	0.1066	0.2236	0.2979	**0.4068**	0.3658	0.3523	0.2006	0.1742	0.1702	0.1610	0.0933
close contact(X)	0.5275	0.5398	0.5528	0.5663	0.5444	0.5593	0.5780	0.6154	0.5861	0.5777	0.6518	0.6457	0.6958	**0.8141**	0.7654
suspected case(X)	**0.6602**	0.5668	0.5968	0.5900	0.4970	0.5018	0.4079	0.3418	0.3440	0.3585	0.3794	0.3439	0.2807	0.2505	0.1570
pneumonia(X)	0.5309	0.4647	0.4966	0.5519	0.5188	0.5763	**0.6085**	0.5478	0.5015	0.3581	0.3335	0.2958	0.2088	0.2935	0.2533
cough(X)	0.1222	0.2561	0.2910	0.3331	0.3502	0.1928	0.1424	0.2064	0.3576	0.4677	0.5261	**0.5713**	0.5388	0.4548	0.4995
dyspnea(X)	**0.4577**	0.2856	0.2484	0.2027	0.1410	0.1226	0.0812	0.0335	0.0739	0.1068	0.1335	0.1500	0.2065	0.1989	0.2685
virus pneumonia(X)	0.3238	0.2885	0.2755	0.3359	0.4292	0.5190	0.5639	0.5479	0.5394	0.5194	0.4681	0.4972	0.5619	0.5980	**0.6976**
novel coronavirus(X)	0.6502	0.5922	0.6028	0.6645	0.7038	0.7555	**0.7620**	0.7090	0.6852	0.6783	0.6676	0.6376	0.5579	0.4770	0.4044

Optimal values are shown in bold in the table

Table 26 Correlation analysis between keywords and the cumulative number of confirmed cases—Xiantao

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.0048	0.0417	0.0577	0.0727	0.0909	0.1485	0.1553	0.2056	0.2642	0.2937	0.3235	**0.3783**	0.3451	0.2913	0.3252
Red Cross Society(S)	0.2052	0.1609	0.1227	0.0704	0.0227	0.0007	0.0458	0.1070	0.1532	0.2110	0.2651	0.3398	0.4025	0.4663	**0.5547**
Spring Festival(S)	0.6092	0.6173	0.6296	0.6645	0.6680	**0.6719**	0.6576	0.6479	0.6198	0.6106	0.6081	0.6011	0.5876	0.5543	0.5251
Health Commission(S)	0.1098	0.0621	0.0075	0.0805	0.1616	0.2221	0.2595	0.3274	0.3738	0.4372	0.5235	0.6240	**0.6535**	0.6285	0.6212
vaccine(S)	0.2401	0.2282	0.2068	0.1502	0.1069	0.0561	0.0055	0.0527	0.1113	0.1774	0.2164	0.2460	0.3307	0.3823	**0.4295**
SARS(S)	0.1203	0.1922	0.2633	0.3398	0.4226	0.5075	0.5992	0.6949	0.7761	0.8457	0.9385	0.9496	0.9564	0.9558	**0.9575**
hand washing(S)	0.1660	0.1744	0.1832	0.1923	0.2018	0.2117	0.2220	0.2328	0.2440	0.2558	0.2682	0.2811	0.2947	**0.3090**	0.1979
children(S)	0.4079	**0.4491**	0.4183	0.4207	0.4050	0.3707	0.3903	0.3555	0.3331	0.3312	0.3281	0.3239	0.3021	0.2581	0.3169
mask(S)	0.0696	0.0294	0.0192	0.0882	0.1415	0.2079	0.2845	0.3440	0.4045	0.4745	0.5415	**0.5684**	0.5597	0.5454	0.4944
coronavirus(S)	0.0601	0.1318	0.2027	0.2748	0.3580	0.4397	0.5260	0.6195	0.7170	0.8124	0.9186	0.9379	0.9422	**0.9463**	0.9412
novel(S)	0.0405	0.0204	0.0202	0.0584	0.1200	0.1777	0.2467	0.2848	0.3231	0.3671	0.4534	0.5340	**0.5351**	0.5143	0.4701
disinfection(S)	0.0773	0.0701	0.0653	0.0257	0.0269	0.0381	0.0487	0.0632	0.0768	0.0912	0.1064	0.1512	0.2115	**0.2337**	0.1771
lockdown of the city(S)	0.0891	0.0516	0.0337	0.0093	0.0535	0.1125	0.1549	0.1833	0.2537	0.2920	0.3279	0.3838	0.4530	**0.5029**	0.4550
Wuhan(S)	0.0252	0.0868	0.1588	0.2307	0.3154	0.4006	0.4814	0.5551	0.6398	0.7282	0.8258	**0.8700**	0.8514	0.8412	0.8408
Zhong Nanshan(S)	0.0501	0.1085	0.1735	0.2547	0.3231	0.4038	0.4929	0.5826	0.6741	0.7787	0.8961	**0.9047**	0.8924	0.8876	0.8895
epidemic(S)	**0.8340**	0.8324	0.8247	0.8121	0.7809	0.7553	0.7572	0.7436	0.7398	0.7273	0.7272	0.6971	0.6688	0.6467	0.6193
flu(S)	0.0171	0.0929	0.1702	0.2468	0.3198	0.3984	0.4729	0.5549	0.6016	0.6773	0.7339	0.7937	**0.7989**	0.7857	0.7941
diarrhea(S)	0.0097	0.0416	0.0994	0.1537	0.2109	0.2910	0.3454	0.4078	0.5003	0.5875	0.6286	0.6939	0.7534	**0.7629**	0.7558
dry cough(S)	0.1254	0.1771	0.2254	0.2939	0.3662	0.4402	0.5068	0.5687	0.6201	0.6886	0.7580	0.7981	0.7952	**0.8042**	0.7990
NCP(S)	**0.5207**	0.4986	0.4664	0.4327	0.4003	0.3643	0.3399	0.3062	0.2781	0.2502	0.2092	0.1693	0.1303	0.1140	0.0978
novel coronary pneumonia(S)	0.9045	0.9102	0.9158	0.9174	0.9139	0.9123	0.9122	0.9171	0.9200	**0.9264**	0.9094	0.9001	0.8955	0.8717	0.8653
COVID-19(S)	0.6834	0.6992	**0.7156**	0.6919	0.6687	0.6424	0.6272	0.6052	0.5873	0.5777	0.5851	0.5757	0.5813	0.5543	0.5322
novel coronavirus pneumonia(S)	0.2980	0.2439	0.1868	0.1308	0.0670	0.0001	0.0581	0.1273	0.1859	0.2483	0.3333	0.4109	0.4801	**0.5622**	0.5528
asymptomatic infection(S)	**0.5446**	0.5167	0.4866	0.4652	0.4432	0.4018	0.3700	0.3252	0.2776	0.2388	0.2249	0.1828	0.1269	0.0835	0.0316
aerosol transmission(S)	**0.2759**	0.2359	0.1978	0.1592	0.1135	0.0768	0.0659	0.0544	0.0408	0.0028	0.0548	0.1039	0.1507	0.1763	0.2028
pneumonia(S)	0.0214	0.0427	0.1063	0.1687	0.2507	0.3365	0.4154	0.5020	0.5972	0.6890	0.7846	0.8914	0.9113	0.9130	**0.9136**
cough(S)	0.0251	0.0490	0.1004	0.1635	0.2265	0.3023	0.3314	0.4088	0.4879	0.5389	0.6183	0.6551	0.7279	0.7310	**0.7597**
fever(S)	0.0903	0.1552	0.1980	0.2276	0.2877	0.3251	0.3824	0.4260	0.4821	0.5610	0.6148	0.5957	0.6535	**0.6641**	0.6459
dyspnea(S)	0.1876	0.2137	0.2603	0.3356	0.4073	0.4201	0.4790	0.5421	0.5206	0.5858	0.6386	**0.7100**	0.6949	0.6660	0.6560
high temperature(S)	0.3397	0.3898	0.4200	0.4729	0.5303	0.5995	0.6408	0.6357	0.6802	0.7262	0.7747	**0.7764**	0.7577	0.7372	0.7144
novel coronavirus(S)	0.1362	0.0712	0.0147	0.0517	0.1332	0.2146	0.2925	0.3719	0.4674	0.5585	0.6561	0.7670	0.8670	0.9058	**0.9070**
SARS-CoV-2(S)	0.2786	0.2843	0.2902	0.2963	**0.3027**	0.2544	0.2072	0.2104	0.2132	0.1653	0.1620	0.1640	0.1604	0.1139	0.1068
Spring Festival(X)	0.7891	0.8005	0.8254	0.8195	0.8302	0.8342	0.8340	0.8280	**0.8346**	0.8275	0.8120	0.8176	0.8109	0.7909	0.7690
Red Cross Society(X)	0.2069	0.1563	0.1017	0.0418	0.0228	0.0963	0.1741	0.2420	0.3096	0.3596	0.3856	0.4539	0.5363	0.6367	**0.6522**
government(X)	0.3477	0.3641	0.3312	0.2924	0.2768	0.2438	0.3327	**0.4250**	0.4211	0.3992	0.3843	0.3788	0.3552	0.3638	0.3686

Table 26 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0002	0.0447	0.0801	0.1474	0.2309	0.3003	0.3155	0.3814	0.4665	0.5642	0.6309	0.6480	0.7008	**0.7129**	0.6937
medical(X)	0.1814	0.1325	0.1183	0.0509	0.0222	0.0862	0.1467	0.1957	0.2512	0.3092	0.3579	0.4326	0.5011	**0.5240**	0.5108
vaccine(X)	0.1864	0.1487	0.1132	0.0969	0.0404	0.0019	0.0321	0.0759	0.1156	0.1723	0.2108	0.2544	0.3123	**0.3459**	0.3359
SARS(X)	0.0707	0.1289	0.1594	0.2292	0.3084	0.3863	0.4493	0.5124	0.5968	0.6812	0.7545	**0.7804**	0.7789	0.7713	0.7627
children(X)	0.4480	0.4737	0.4519	0.4781	0.5055	0.5342	0.5949	0.6293	0.6117	0.6474	0.6839	**0.7223**	0.7015	0.6781	0.6610
mask(X)	0.1371	0.0884	0.0707	0.0031	0.0652	0.1278	0.1783	0.2235	0.2710	0.3201	0.3679	0.4493	**0.5362**	0.5264	0.4973
infection(X)	0.3200	0.3374	0.3555	0.3744	0.3941	0.4147	0.4362	0.4587	0.4822	0.5069	0.5328	0.5600	**0.5886**	0.5502	0.5064
coronavirus(X)	**0.7256**	0.6993	0.6647	0.6291	0.5890	0.5495	0.5161	0.4778	0.4408	0.3965	0.3896	0.3475	0.2894	0.2387	0.1870
disinfection(X)	0.4535	0.4632	0.4733	0.4838	0.4948	0.5063	0.5183	0.5309	0.5441	0.5579	0.5724	0.5877	0.6037	0.6206	**0.6385**
nucleic acid(X)	**0.5297**	0.4848	0.4358	0.4371	0.3956	0.3481	0.3325	0.2996	0.2666	0.2443	0.2321	0.1804	0.1173	0.0493	0.0073
lockdown of the city(X)	0.4720	0.4806	0.4897	0.4992	0.5091	0.5196	0.5305	0.5421	0.5543	0.5671	0.5806	0.5950	0.5909	0.6050	**0.6224**
Wuhan(X)	0.0001	0.0358	0.0609	0.1147	0.1862	0.2588	0.3300	0.4214	0.5186	0.5997	0.6553	0.6639	0.6746	**0.7267**	0.7131
Zhong Nanshan(X)	0.2193	0.1625	0.1467	0.0795	0.0063	0.0655	0.1181	0.1662	0.2342	0.3201	0.4209	0.5052	0.5359	0.5363	**0.5416**
epidemic(X)	**0.5344**	0.4975	0.4537	0.4568	0.4096	0.3565	0.3134	0.2688	0.2188	0.1614	0.1409	0.0822	0.0081	0.0597	0.1170
flu(X)	**0.4134**	0.3944	0.3976	0.3997	0.3525	0.3066	0.2782	0.2586	0.2452	0.2319	0.2173	0.1959	0.1704	0.1318	0.1362
diarrhea(X)	0.0330	0.1009	0.1123	0.1187	0.1309	0.1418	0.1531	0.1650	0.1774	0.1904	0.2041	0.2184	0.2334	0.2492	**0.2658**
stuffy nose(X)	0.0973	0.0591	0.0569	0.0763	0.0229	0.0334	0.1075	0.1321	0.1976	0.2251	0.2481	0.3197	0.3666	0.4077	**0.4583**
dry cough(X)	0.2678	0.2216	0.1639	0.1498	0.0962	0.0375	0.0249	0.0993	0.1742	0.2474	0.2971	**0.3776**	0.3532	0.3740	0.3641
NCP(X)	**0.3211**	0.2836	0.2422	0.2072	0.1674	0.1628	0.1597	0.1538	0.1489	0.1437	0.1382	0.0906	0.0464	0.0010	0.0458
close contact(X)	0.7973	0.7880	0.8056	**0.8266**	0.8052	0.7856	0.7804	0.7587	0.7377	0.7323	0.7453	0.7214	0.6911	0.6601	0.6416
suspected case(X)	**0.5967**	0.5633	0.5256	0.4818	0.4322	0.3819	0.3475	0.2941	0.2484	0.2010	0.1765	0.1314	0.0674	0.0035	0.0428
pneumonia(X)	**0.5825**	0.5500	0.5606	0.5215	0.4737	0.4298	0.3920	0.3490	0.3067	0.2559	0.2415	0.1909	0.1245	0.0630	0.0537
cough(X)	0.2030	0.2670	0.2970	0.3265	0.3463	0.3569	0.3514	0.3617	0.3970	0.4294	0.5305	0.6197	0.6585	0.6749	**0.6863**
dyspnea(X)	0.1337	0.0889	0.0750	0.0603	0.0060	0.0506	0.0800	0.1255	0.1229	0.1168	0.1314	0.1830	0.2135	0.2593	**0.2913**
virus pneumonia(X)	0.5053	0.5290	0.5319	0.5526	0.5932	0.6120	0.6057	0.6029	0.6263	**0.6655**	0.6468	0.6579	0.6605	0.6404	0.6086
novel coronavirus(X)	**0.7274**	0.7006	0.7122	0.6797	0.6401	0.6045	0.5746	0.5401	0.5061	0.4647	0.4589	0.4193	0.3651	0.3166	0.2674

Optimal values are shown in bold in the table

Table 27 Correlation analysis between keywords and the cumulative number of confirmed cases—Xiaogan

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	0.1706	0.1457	0.1004	0.0354	0.0242	0.0570	0.1093	0.1472	0.2257	0.3002	0.3649	0.4346	0.4927	0.5131	**0.5508**
Red Cross Society(S)	0.1681	0.1163	0.0627	0.0059	0.0586	0.1103	0.1737	0.2348	0.2977	0.3578	0.4226	0.4880	0.5786	0.6754	**0.6948**
Spring Festival(S)	0.8011	0.8235	0.8467	**0.8484**	0.8410	0.8304	0.8374	0.8402	0.8234	0.8265	0.8293	0.8262	0.8091	0.7877	0.7738
Health Commission(S)	0.0399	0.0157	0.0850	0.1377	0.1966	0.2599	0.3342	0.4098	0.4951	0.5750	0.6558	0.7608	0.8060	**0.8072**	0.8072
vaccine(S)	0.2336	0.1937	0.1510	0.1077	0.0748	0.0442	0.0051	0.0474	0.1135	0.1913	0.2509	0.3270	0.3942	0.4672	**0.5356**
SARS(S)	0.1686	0.2350	0.3070	0.3828	0.4611	0.5457	0.6328	0.7264	0.8191	0.9230	0.9624	0.9679	0.9678	0.9731	**0.9771**
hand washing(S)	0.1803	0.1730	**0.2071**	0.1689	0.1617	0.1342	0.1248	0.1168	0.1083	0.0755	0.0622	0.0518	0.0133	0.0078	0.0201
children(S)	0.3337	0.3857	0.3672	0.3737	0.3827	**0.3967**	0.3773	0.3459	0.3144	0.2732	0.2737	0.2155	0.1672	0.1284	0.0833
mask(S)	0.0671	0.1331	0.1793	0.2310	0.2897	0.3524	0.4188	0.4843	0.5677	0.6553	0.7396	**0.7947**	0.7805	0.7730	0.7513
coronavirus(S)	0.1115	0.1758	0.2438	0.3136	0.3902	0.4710	0.5549	0.6457	0.7411	0.8379	0.9458	0.9685	**0.9714**	0.9703	0.9698
novel(S)	0.0276	0.0695	0.1335	0.1846	0.2498	0.3141	0.3923	0.4608	0.5395	0.6229	0.7132	0.8174	**0.8205**	0.8105	0.8056
disinfection(S)	0.0088	0.0656	0.1006	0.1397	0.2084	0.2712	0.3206	0.3615	0.4343	0.4991	0.5492	0.5951	**0.6408**	0.6393	0.6389
lockdown of the city(S)	0.1226	0.0768	0.0384	0.0074	0.0600	0.1065	0.1654	0.2358	0.3078	0.3807	0.4699	0.5612	0.6434	**0.7440**	0.7264
Wuhan(S)	0.1064	0.1698	0.2371	0.3070	0.3718	0.4493	0.5333	0.6261	0.7191	0.8172	0.9203	0.9452	0.9506	0.9564	**0.9565**
Zhong Nanshan(S)	0.0866	0.1438	0.2083	0.2797	0.3514	0.4348	0.5205	0.6044	0.6923	0.7906	0.9029	**0.9039**	0.8908	0.8865	0.8979
epidemic(S)	0.8707	0.8624	0.8662	0.8725	0.8752	0.8857	**0.8900**	0.8812	0.8756	0.8769	0.8797	0.8615	0.8596	0.8503	0.8407
flu(S)	0.1195	0.1824	0.2495	0.2961	0.3668	0.4409	0.5143	0.6012	0.6973	0.7944	0.8821	0.9131	**0.9380**	0.9376	0.9332
diarrhea(S)	0.1294	0.0729	0.0246	0.0361	0.0859	0.1383	0.1987	0.2725	0.3327	0.4138	0.4817	0.5601	0.6569	0.7256	**0.7688**
dry cough(S)	0.1157	0.1813	0.2357	0.2987	0.3383	0.4098	0.4929	0.5699	0.6548	0.7358	0.8273	0.8856	0.9168	**0.9210**	0.9186
NCP(S)	**0.7156**	0.6933	0.6695	0.6436	0.6206	0.5945	0.5635	0.5363	0.5048	0.4772	0.4493	0.4165	0.3779	0.3426	0.3037
novel coronary pneumonia(S)	0.9257	0.9295	0.9319	0.9310	0.9326	0.9370	0.9422	**0.9466**	0.9465	0.9431	0.9376	0.9226	0.9124	0.9062	0.8976
COVID-19(S)	**0.8302**	0.8121	0.7981	0.7873	0.7784	0.7787	0.7623	0.7552	0.7568	0.7327	0.7213	0.6988	0.6731	0.6448	0.6285
novel coronavirus pneumonia(S)	0.2162	0.1714	0.1439	0.0950	0.0583	0.0130	0.0375	0.0947	0.1616	0.2251	0.2880	0.3789	0.4294	**0.5150**	0.5087
asymptomatic infection(S)	**0.7057**	0.6781	0.6722	0.6385	0.6280	0.6165	0.5893	0.5580	0.5268	0.4897	0.4535	0.4051	0.3769	0.3316	0.2960
aerosol transmission(S)	**0.4399**	0.4132	0.3876	0.3513	0.3213	0.2906	0.2542	0.2096	0.1737	0.1366	0.1011	0.0633	0.0116	0.0318	0.0682
pneumonia(S)	0.0573	0.1212	0.1862	0.2548	0.3275	0.4027	0.4825	0.5695	0.6606	0.7547	0.8530	0.9434	**0.9477**	0.9427	0.9403
cough(S)	0.1918	0.2373	0.2985	0.2998	0.2918	0.3493	0.4456	0.5451	0.6479	0.6461	0.6457	0.6975	0.7407	**0.7531**	0.7433
fever(S)	0.2901	0.2834	0.3462	0.4065	0.4401	0.4887	0.4948	0.5387	0.6159	0.6809	0.7559	0.8074	0.8164	0.8262	**0.8263**
dyspnea(S)	0.0782	0.0197	0.0241	0.0727	0.1380	0.1879	0.2415	0.3051	0.3888	0.4533	0.5161	0.6001	0.6284	**0.6517**	0.6364
high temperature(S)	0.0221	0.0911	0.1394	0.2073	0.2654	0.3307	0.3890	0.4631	0.5346	0.6144	**0.6568**	0.6361	0.6277	0.5987	0.5867
novel coronavirus(S)	0.0167	0.0444	0.1069	0.1733	0.2444	0.3184	0.3958	0.4801	0.5687	0.6599	0.7582	0.8663	0.9261	0.9642	**0.9671**
SARS-CoV-2(S)	**0.5920**	0.5692	0.5799	0.5911	0.5787	0.5894	0.5688	0.5441	0.5261	0.5133	0.5211	0.5021	0.4767	0.4623	0.4372
Spring Festival(X)	0.8039	0.8073	0.8169	0.8210	**0.8257**	0.8222	0.8086	0.8117	0.8115	0.7973	0.7856	0.7770	0.7646	0.7632	0.7610
Red Cross Society(X)	0.2373	0.1850	0.1285	0.0805	0.0330	0.0224	0.0780	0.1418	0.2104	0.2934	0.3650	0.4421	0.5254	**0.6010**	0.5948
government(X)	0.3292	0.3165	0.3091	0.2909	0.2538	0.2245	0.3077	0.3964	0.4051	0.3909	0.4070	**0.4078**	0.3899	0.3974	0.4025

Table 27 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0088	0.0461	0.0492	0.0911	0.1516	0.2186	0.2670	0.3380	0.4195	0.4995	0.5687	0.5872	**0.6155**	0.6108	0.5840
medical(X)	0.1735	0.1188	0.0712	0.0317	0.0038	0.0379	0.0649	0.1184	0.1740	0.2157	0.2696	0.3225	0.3806	**0.4196**	0.4154
vaccine(X)	0.1955	0.1662	0.1347	0.1008	0.0677	0.0211	0.0069	0.0434	0.0977	0.1604	0.2369	0.3125	0.3971	**0.4411**	0.4372
SARS(X)	0.1058	0.1721	0.2417	0.3126	0.3871	0.4566	0.4928	0.5531	0.6306	0.7141	0.8088	0.8276	**0.8298**	0.8259	0.8114
children(X)	0.4338	0.4117	0.4406	0.4711	0.5210	0.5732	0.5750	0.5561	0.5350	0.5726	0.6299	**0.6733**	0.6517	0.6246	0.6004
mask(X)	0.0584	0.0033	0.0353	0.0751	0.1170	0.1592	0.1982	0.2613	0.3216	0.3899	0.4530	0.5260	**0.5947**	0.5856	0.5524
infection(X)	0.3539	0.3758	0.3963	0.4222	0.4470	0.4730	0.5003	0.5289	0.5589	0.5905	0.6237	0.6587	**0.6956**	0.6709	0.6436
coronavirus(X)	**0.6950**	0.6603	0.6384	0.6082	0.5780	0.5422	0.4994	0.4696	0.4263	0.3852	0.3424	0.2945	0.2511	0.2094	0.1529
disinfection(X)	0.4987	0.5097	0.5211	0.5330	0.5454	0.5585	0.5721	0.5864	0.6013	0.6170	0.6336	0.6509	0.6692	0.6885	**0.7089**
nucleic acid(X)	**0.5321**	0.4988	0.4917	0.4689	0.4537	0.4424	0.4101	0.3646	0.3180	0.2628	0.2197	0.1750	0.1396	0.0864	0.0381
lockdown of the city(X)	0.5254	0.5372	0.5496	0.5596	0.5716	0.5842	0.5782	0.5910	0.6046	0.6189	0.6340	0.6501	0.6671	0.6852	**0.7027**
Wuhan(X)	0.0489	0.0052	0.0502	0.1226	0.1955	0.2548	0.3442	0.4100	0.4807	0.5541	0.6218	0.6366	0.6565	**0.7125**	0.7029
Zhong Nanshan(X)	0.1225	0.0623	0.0232	0.0164	0.0712	0.1400	0.1754	0.2453	0.3055	0.3836	0.4700	0.5636	**0.6059**	0.6039	0.6053
epidemic(X)	**0.4747**	0.4292	0.3938	0.3568	0.3132	0.2639	0.2466	0.2036	0.1452	0.0923	0.0364	0.0278	0.0877	0.1544	0.2239
flu(X)	**0.4409**	0.4046	0.3841	0.3712	0.3622	0.3501	0.3424	0.3329	0.3075	0.2779	0.2290	0.1800	0.1351	0.1046	0.1244
diarrhea(X)	0.0618	0.0562	0.0479	0.0466	0.0377	0.0280	0.0240	0.0196	0.0087	0.0005	0.0082	0.0173	0.0573	0.0698	**0.0830**
stuffy nose(X)	0.1686	0.1221	0.0899	0.0767	0.0361	0.0107	0.0299	0.0784	0.1222	0.1751	0.2321	0.2864	0.3637	0.4296	**0.5053**
dry cough(X)	0.2177	0.1759	0.1250	0.0672	0.0057	0.0542	0.0900	0.1572	0.2211	0.2945	0.3621	**0.4401**	0.4011	0.4032	0.4004
NCP(X)	**0.2778**	0.2499	0.2188	0.1860	0.1547	0.1492	0.1460	0.0993	0.0548	0.0145	0.0304	0.0478	0.0669	0.0839	0.1075
close contact(X)	**0.8143**	0.8024	0.7997	0.7830	0.7674	0.7600	0.7765	0.7570	0.7331	0.7130	0.7000	0.6880	0.6712	0.6498	0.6181
suspected case(X)	**0.4854**	0.4454	0.4197	0.3755	0.3422	0.3026	0.2853	0.2509	0.2030	0.1556	0.1212	0.0809	0.0304	0.0211	0.0778
pneumonia(X)	**0.6024**	0.5685	0.5455	0.5132	0.4807	0.4435	0.4007	0.3588	0.3149	0.2684	0.2221	0.1702	0.1202	0.0673	0.0497
cough(X)	0.3078	0.3762	0.4317	0.4908	0.4930	0.4876	0.4609	0.4779	0.5027	0.5173	0.6031	0.6869	0.7154	0.7268	**0.7316**
dyspnea(X)	**0.2301**	0.1881	0.1596	0.1310	0.0994	0.0715	0.0507	0.0038	0.0096	0.0146	0.0258	0.0939	0.1536	0.2083	0.2245
virus pneumonia(X)	0.6014	0.6026	0.5858	0.5922	0.6070	0.6397	0.6244	0.6438	0.6775	**0.6971**	0.6907	0.6722	0.6542	0.6385	0.6095
novel coronavirus(X)	**0.6839**	0.6538	0.6342	0.6069	0.5800	0.5467	0.5437	0.5164	0.4790	0.4424	0.4057	0.3645	0.3273	0.2888	0.2422

Optimal values are shown in bold in the table

Table 28 Correlation analysis between keywords and the cumulative number of confirmed cases—Yichang

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Center for Disease Control and Prevention(S)	**0.2973**	0.2656	0.2338	0.1586	0.1058	0.0726	0.0306	0.0233	0.0574	0.1112	0.1653	0.2118	0.1995	0.2408	0.2534
Red Cross Society(S)	0.2334	0.1874	0.1335	0.0760	0.0412	0.0183	0.0726	0.1437	0.2001	0.2765	0.3441	0.4254	0.5048	0.5916	**0.6730**
Spring Festival(S)	0.8581	0.8626	0.8867	**0.8891**	0.8816	0.8811	0.8757	0.8652	0.8621	0.8593	0.8354	0.8228	0.8258	0.8107	0.8034
Health Commission(S)	0.1093	0.0363	0.0390	0.1066	0.1735	0.2507	0.3143	0.3830	0.4660	0.5456	0.6255	0.7161	0.7863	0.8001	**0.8189**
vaccine(S)	0.2283	0.1916	0.1453	0.1051	0.0504	0.0074	0.0300	0.0894	0.1369	0.1822	0.2493	0.3213	0.3649	**0.4651**	0.4628
SARS(S)	0.2104	0.2821	0.3589	0.4371	0.5213	0.6064	0.6974	0.7906	0.8597	0.9260	0.9605	0.9663	0.9665	0.9660	**0.9697**
hand washing(S)	**0.3582**	0.3251	0.3131	0.2739	0.2183	0.2115	0.1608	0.1452	0.0978	0.0608	0.0111	0.0028	0.0455	0.0054	0.0444
children(S)	0.5504	0.5269	0.5023	0.4896	0.4517	0.4889	**0.5820**	0.5613	0.5195	0.4906	0.5119	0.4565	0.4257	0.4421	0.3964
mask(S)	0.0654	0.1330	0.1999	0.2733	0.3415	0.4032	0.4718	0.5476	0.6233	0.7081	0.8058	**0.8494**	0.8323	0.8295	0.8170
coronavirus(S)	0.1046	0.1709	0.2392	0.3120	0.3895	0.4692	0.5501	0.6394	0.7388	0.8400	0.9499	0.9591	0.9617	0.9612	**0.9686**
novel(S)	0.0064	0.0524	0.1123	0.1798	0.2386	0.3143	0.4062	0.4859	0.5788	0.6819	0.7797	0.8460	**0.8510**	0.8306	0.8146
disinfection(S)	**0.2895**	0.2560	0.2028	0.1863	0.1297	0.0948	0.0768	0.0341	0.0551	0.0845	0.1376	0.2045	0.2314	0.2668	0.2318
lockdown of the city(S)	0.1438	0.1149	0.0629	0.0027	0.0471	0.0968	0.1669	0.2381	0.2979	0.3787	0.4594	0.5339	0.6335	**0.7402**	0.6874
Wuhan(S)	0.0907	0.1649	0.2370	0.3075	0.3810	0.4604	0.5361	0.6255	0.7256	0.8191	0.9192	0.9231	0.9297	0.9257	**0.9396**
Zhong Nanshan(S)	0.0678	0.1369	0.2056	0.2728	0.3493	0.4281	0.5111	0.6018	0.6936	0.7868	0.8845	**0.8943**	0.8862	0.8814	0.8714
epidemic(S)	**0.6942**	0.6585	0.6218	0.5965	0.5608	0.5411	0.5269	0.5096	0.4570	0.4126	0.3731	0.3298	0.2777	0.2305	0.1465
flu(S)	0.0838	0.1225	0.1923	0.2631	0.3360	0.4060	0.4808	0.5577	0.6560	0.7517	0.8296	0.8962	0.9283	**0.9351**	0.9291
diarrhea(S)	0.0434	0.0154	0.0805	0.1277	0.1849	0.2577	0.3182	0.3704	0.4404	0.5074	0.5726	0.6612	0.6936	0.7564	**0.7710**
dry cough(S)	0.0940	0.1602	0.2271	0.2883	0.3508	0.4174	0.4984	0.5762	0.6715	0.7603	0.8305	0.8696	**0.8762**	0.8715	0.8594
NCP(S)	**0.7519**	0.7287	0.6940	0.6567	0.6342	0.6127	0.5820	0.5585	0.5445	0.5100	0.4807	0.4428	0.4067	0.3661	0.3223
novel coronary pneumonia(S)	**0.9276**	0.9181	0.9122	0.9103	0.9074	0.9099	0.9058	0.8920	0.8885	0.8901	0.8896	0.8782	0.8585	0.8447	0.8254
COVID-19(S)	**0.8404**	0.8245	0.8043	0.7893	0.7756	0.7699	0.7401	0.7245	0.7150	0.7007	0.6761	0.6603	0.6458	0.6127	0.5867
novel coronavirus pneumonia(S)	0.2147	0.1610	0.1065	0.0701	0.0260	0.0329	0.0920	0.1602	0.2138	0.2783	0.3430	0.4306	0.4744	**0.5462**	0.5455
asymptomatic infection(S)	**0.6790**	0.6694	0.6503	0.6150	0.5787	0.5628	0.5304	0.4863	0.4438	0.4107	0.3816	0.3400	0.3089	0.2676	0.2288
aerosol transmission(S)	**0.5176**	0.4851	0.4573	0.4336	0.4031	0.3702	0.3345	0.2890	0.2481	0.2059	0.1623	0.1170	0.0670	0.0111	0.0480
pneumonia(S)	0.0663	0.1411	0.2129	0.2830	0.3592	0.4398	0.5230	0.6106	0.7126	0.8072	0.9124	0.9450	0.9506	0.9462	**0.9505**
cough(S)	0.2371	0.1942	0.1436	0.0694	0.0293	0.0156	0.0873	0.1404	0.1776	0.2319	0.2862	0.3327	0.4150	**0.4314**	0.4178
fever(S)	0.1348	0.2055	0.2808	0.2767	0.3188	0.3671	0.4319	0.4988	0.5626	0.5796	0.6058	0.6636	0.7098	**0.7297**	0.7170
dyspnea(S)	0.0455	0.1005	0.1527	0.2119	0.2989	0.3265	0.3799	0.4638	0.5157	0.6099	0.6903	0.7257	0.7180	0.7485	**0.7762**
high temperature(S)	0.1110	0.1708	0.1849	0.2514	0.3068	0.3938	0.4730	0.5361	0.6087	0.6968	0.7527	**0.7684**	0.7675	0.7428	0.7024
novel coronavirus(S)	0.0170	0.0541	0.1180	0.1868	0.2627	0.3364	0.4088	0.4954	0.5925	0.6827	0.7813	0.8881	0.9248	0.9500	**0.9555**
SARS-CoV-2(S)	**0.6330**	0.6117	0.5892	0.5653	0.5533	0.5297	0.5108	0.4707	0.4605	0.4492	0.4239	0.3950	0.3744	0.3603	0.3471
Spring Festival(X)	0.8399	0.8492	0.8579	0.8648	0.8682	0.8734	0.8725	0.8780	**0.8856**	0.8781	0.8717	0.8604	0.8551	0.8444	0.8587
Red Cross Society(X)	0.2758	0.2338	0.1888	0.1284	0.0784	0.0681	0.0262	0.0241	0.0716	0.1123	0.1727	0.2272	0.2781	**0.3391**	0.3262
government(X)	0.2940	0.2999	0.2678	0.2205	0.2130	0.2059	0.2414	0.2940	0.3445	0.3287	0.3188	0.3178	0.3300	0.3691	**0.3745**

Table 28 (continued)

Keywords	Delay days														
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Health Commission(X)	0.0630	0.1412	0.1564	0.1977	0.2644	0.3267	0.3460	0.3919	0.4968	0.5683	0.6515	0.6475	**0.7042**	0.6521	0.6471
medical(X)	0.1467	0.0957	0.0513	0.0147	0.0660	0.1037	0.1437	0.1814	0.2602	0.3074	0.3506	0.4015	**0.4808**	0.4732	0.4660
vaccine(X)	0.1968	0.1513	0.1313	0.0999	0.0552	0.0401	0.0167	0.0601	0.0892	0.1354	0.2091	0.2831	0.3329	**0.3717**	0.3554
SARS(X)	0.1007	0.1639	0.2219	0.2936	0.3750	0.4103	0.4524	0.5043	0.5992	0.6687	0.7538	**0.7736**	0.7582	0.7231	0.7092
children(X)	0.4189	0.4484	0.4778	0.4718	0.5169	**0.5513**	0.5276	0.5014	0.4943	0.4679	0.5024	0.5391	0.5032	0.4639	0.4520
mask(X)	0.1186	0.0663	0.0233	0.0182	0.0774	0.1089	0.1370	0.1999	0.2889	0.3309	0.3829	0.4557	**0.5369**	0.5033	0.4922
infection(X)	0.3301	0.3392	0.3616	0.3924	0.4207	0.4385	0.4855	0.5139	0.5269	0.5590	0.5940	0.6211	**0.6485**	0.6327	0.6013
coronavirus(X)	**0.6514**	0.6147	0.5918	0.5683	0.5279	0.5216	0.4943	0.4691	0.4433	0.4133	0.3695	0.3179	0.3011	0.2753	0.2411
disinfection(X)	0.5295	0.5412	0.5534	0.5423	0.5544	0.5670	0.5803	0.5941	0.6087	0.6160	0.6311	0.6608	0.6769	0.6930	**0.7179**
nucleic acid(X)	**0.5181**	0.4873	0.4691	0.4404	0.4024	0.3938	0.3958	0.3529	0.2937	0.2551	0.2119	0.1487	0.0976	0.0680	0.0146
lockdown of the city(X)	0.5070	0.5274	0.5433	0.5488	0.5534	0.5823	0.5898	0.5927	0.6094	0.6091	0.5991	0.6128	0.6274	0.6430	**0.6596**
Wuhan(X)	0.0200	0.0103	0.0645	0.1375	0.2188	0.2766	0.3507	0.4376	0.5156	0.5755	0.6514	0.6491	0.6542	**0.7068**	0.7033
Zhong Nanshan(X)	0.1379	0.0734	0.0131	0.0373	0.0865	0.1389	0.2118	0.3257	0.3949	0.4533	0.5357	0.6473	0.6725	**0.6818**	0.6691
epidemic(X)	**0.4238**	0.3740	0.3402	0.2927	0.2347	0.1966	0.1594	0.1385	0.0657	0.0079	0.0402	0.1060	0.1865	0.2179	0.2775
flu(X)	**0.4846**	0.4318	0.3863	0.3661	0.3540	0.3450	0.3272	0.3260	0.3107	0.2866	0.2584	0.2138	0.1634	0.1156	0.1444
diarrhea(X)	0.0405	0.0394	0.0541	0.0432	0.0045	0.0028	0.0168	0.0075	0.0041	0.0008	0.0048	0.0222	0.0296	0.0789	**0.0903**
stuffy nose(X)	0.0895	0.0185	0.0277	0.0765	0.1049	0.1404	0.1723	0.1928	0.2444	0.2931	0.3279	0.3628	0.4242	0.4747	**0.5722**
dry cough(X)	0.2406	0.1964	0.1567	0.0975	0.0251	0.0232	0.0821	0.1209	0.1842	0.2510	0.3290	**0.3952**	0.3581	0.3401	0.3318
NCP(X)	**0.4199**	0.3944	0.3566	0.3188	0.2944	0.2848	0.2694	0.2242	0.1803	0.1462	0.1009	0.0446	0.0330	0.0224	0.0044
close contact(X)	**0.8282**	0.8102	0.7950	0.7899	0.7639	0.7464	0.7535	0.7519	0.7159	0.7091	0.6914	0.6549	0.6589	0.6696	0.6315
suspected case(X)	**0.5090**	0.4664	0.4393	0.4039	0.3434	0.3054	0.2675	0.2426	0.2059	0.1583	0.1162	0.0739	0.0438	0.0018	0.0554
pneumonia(X)	**0.5050**	0.4587	0.4229	0.3985	0.3530	0.3068	0.2762	0.2618	0.1942	0.1474	0.1010	0.0361	0.0290	0.0564	0.0416
cough(X)	0.3112	0.3754	0.4428	0.5117	0.5158	0.4920	0.4970	0.4912	0.5242	0.5402	0.6163	0.7095	0.7513	0.7451	**0.7540**
dyspnea(X)	0.2093	0.1655	0.1260	0.0846	0.0402	0.0199	0.0170	0.0463	0.0374	0.0155	0.0500	0.0859	0.1603	0.1967	**0.2185**
virus pneumonia(X)	0.5725	0.5896	0.5660	0.5759	0.6024	0.5902	0.5988	0.5985	0.6370	**0.6459**	0.6435	0.6356	0.5990	0.5782	0.5573
novel coronavirus(X)	**0.7394**	0.7076	0.6838	0.6633	0.6313	0.6264	0.6168	0.5572	0.5075	0.4823	0.4454	0.3866	0.3583	0.3351	0.2955

Optimal values are shown in bold in the table

Appendix B

Critical values for the two-tailed Bonferroni–Dunn test:

Methods	2	3	4	5	6	7
q0.05	1.960	2.241	2.394	2.498	2.576	2.638
q0.10	1.645	1.960	2.128	2.241	2.326	2.394
Methods	8	9	10	11	12	13
q0.05	2.690	2.724	2.774	3.219	3.268	3.313
q0.10	2.450	2.498	2.539	2.978	3.030	3.077

where *methods* is the number of methods used for comparison.

Funding This work was supported by Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No.20JG031) and Postgraduate Innovation Fund Project of Xi'an University of Posts and Telecommunications (CXJJWY2020001).

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

Alessa A, Faezipour M (2018) A review of influenza detection and prediction through social networking sites. Theor Biol Med Model. https://doi.org/10.1186/s12976-017-0074-5

Arthur CK, Temeng VA, Ziggah YY (2020) Novel approach to predicting blast-induced ground vibration using Gaussian process regression. Eng Comput 36:29–42. https://doi.org/10.1007/s00366-018-0686-3

Askari Q, Saeed M, Younas I (2020) Heap-based optimizer inspired by corporate rank hierarchy for global optimization. Expert Syst Appl 161:113702. https://doi.org/10.1016/j.eswa.2020.113702

Aslan IH, Demir M, Wise MM et al (2022) Modeling COVID-19: forecasting and analyzing the dynamics of the outbreaks in Hubei and Turkey. Math Methods Appl Sci 45(10):6481–6494. https://doi.org/10.1002/mma.8181

Bi J, Wang XM, Hu YY et al (2021) A method for dynamic risk assessment and prediction of public health emergencies based on an improved SEIR model: novel coronavirus COVID-19 in ten European countries. J Geo-Inform Sci 23(02):259–273

Brereton RG, Lloyd GR (2010) Support vector machines for classification and regression. Anal 135(2):230–267. https://doi.org/10.1039/b918972f

Cao SL, Feng PH, Shi PP (2020) Study on the epidemic development of COVID-19 in Hubei province by a modified SEIR model. J Zhejiang Univ (med Sci) 49(02):178–184. https://doi.org/10.3785/j.issn.1008-9292.2020.02.05

Chae S, Kwon S, Lee D (2018) Predicting infectious disease using deep learning and big data. Int J Environ Res Public Health 15(8):1596. https://doi.org/10.3390/ijerph15081596

Chen S, Paul R, Janies D et al (2021) Exploring feasibility of multivariate deep learning models in predicting COVID-19 epidemic. Front Public Heal 9:661615. https://doi.org/10.3389/fpubh.2021.661615

Chen Y, Chu CW, Chen MIC et al (2018) The utility of LASSO-based models for real time forecasts of endemic infectious diseases: a cross country comparison. J Biomed Inform 81:16–30. https://doi.org/10.1016/j.jbi.2018.02.014

Ding G, Liu C, Gong J, et al. (2004) SARS epidemical forecast research in mathematical model. Chin Sci Bulletin Kexue Tongbao, 49(21): 2332–2338. https://doi.org/10.1360/04we0073.

Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In Proceedings of the sixth international symposium on micro machine and human science, MHS'95 (pp39–43)

Fang X, Liu W, Ai J, et al. (2020). Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. BMC Infect Dis https://doi.org/10.1186/s12879-020-4930-2

Faramarzi A, Heidarinejad M, Mirjalili S et al (2020) Marine predators algorithm: a nature-inspired metaheuristic. Expert Syst Appl 152:113377. https://doi.org/10.1016/j.eswa.2020.113377

Filmalter JD, Dagorn L, Cowley PD, et al. (2011). First descriptions of the behavior of silky sharks, *Carcharhinus falciformis*, around drifting fish aggregating devices in the Indian Ocean. Bulletin Marine Sci Miami, 87: 325–337. https://www.researchgate.net/publication/233524023

Ginsberg J, Mohebbi MH, Patel RS et al (2009) Detecting influenza epidemics using search engine query data. Nature 457(7232):1012–1014. https://doi.org/10.1038/nature07634

Gumaei A, Al-Rakhami M, Al Rahhal M M, et al. (2021) Prediction of COVID-19 confirmed cases using gradient boosting regression method. Cmc-Comput Mater Continua, 66(1): 315–329. https://doi.org/10.32604/cmc.2020.012045.

Guo Z, Xiao D, Li D et al (2016) Predicting and evaluating the epidemic trend of ebola virus disease in the 2014–2015 outbreak and the effects of intervention measures. PLoS ONE 11(4):e0152438. https://doi.org/10.1371/journal.pone.0152438

Guo H, Fu JD, Li ZD et al. (2022) SVM parameters and feature selection were optimized based on improved whale algorithm. J Jilin Univ (Eng Technol Ed) [J/OL]:1–22[2022–04–22]
Hao Y, Xu T, Hu HP et al (2020) Prediction and analysis of corona virus disease 2019. PLoS ONE 15(10):e0239960. https://doi.org/10.1371/journal.pone.0239960
Hasan N (2020) A methodological approach for predicting COVID-19 epidemic using EEMD-ANN hybrid model. Internet Things 11:100228. https://doi.org/10.1016/j.iot.2020.100228
Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780
Holland JH (1992) Genetic algorithms. Scientific Am 267(1):66–72. https://doi.org/10.1038/scientificamerican0792-66
Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Statistics 6:65–70. https://doi.org/10.1007/BF00139637
Huang R, Luo G, Duan Q et al (2020) Using Baidu search index to monitor and predict newly diagnosed cases of HIV/AIDS, syphilis and gonorrhea in China: estimates from a vector autoregressive (VAR) model. BMJ Open 10(3):e036098. https://doi.org/10.1136/bmjopen-2019-036098
Ighalo JO, Adeniyi AG, Marques G (2020) Application of linear regression algorithm and stochastic gradient descent in a machine-learning environment for predicting biomass higher heating value. Biofuels Bioproduct Biorefin Biofpr 14(6):1286–1295. https://doi.org/10.1002/bbb.2140
Jia HM, Xing ZK, Song WL (2019) A new hybrid seagull optimization algorithm for feature selection. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2909945
Jia W, Han K, Song Y et al (2020) Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with hunan, China. Front Med 7:169. https://doi.org/10.3389/fmed.2020.00169
Jia H M, Li Y, Sun K J (2022). Simultaneous feature selection optimization based on hybrid sooty tern optimization algorithm and genetic algorithm. Acta Automatica Sinica. https://doi.org/10.16383/j.aas.c200322.
Khishe M, Mosavi M R (2020) Chimp optimization algorithm. Expert Syst Appl 149:113338-. https://doi.org/10.1016/j.eswa.2020.113338
Kipf T N, Welling M (2017) Semi-supervised classification with graph convolutional networks [M], In Proc. of ICLR. https://doi.org/10.48550/arXiv.1609.02907
Kuperman M, Abramson G (2001) Small world effect in an epidemiological model. Phys Rev Lett 86(13):2909–2912. https://doi.org/10.1103/PhysRevLett.86.2909
Li Z, Liu T, Zhu G et al (2017) Dengue Baidu search index data can improve the prediction of local dengue epidemic: a case study in Guangzhou, China. Plos Negl Trop Dis 11(3):e0005354. https://doi.org/10.1371/journal.pntd.0005354
Liao ZF, Lan P, Liao ZN et al (2020) TW-SIR: time-window based SIR for COVID-19 forecasts. Scientific Rep 10(1):22454. https://doi.org/10.1038/s41598-020-80007-8
Liu CY, Wu XQ, Fan RG et al (2020) A new SAIR model on complex networks for analysing the 2019 novel coronavirus (COVID-19). Nonlinear Dyn 101(3):1777–1787. https://doi.org/10.1007/s11071-020-05704-5
Manuel DLS, Ibeas A, Agarwal RP (2020) On confinement and quarantine concerns on an SEIAR epidemic model with simulated parameterizations for the COVID-19 pandemic. Symmetry Basel 12(10):1646. https://doi.org/10.3390/sym12101646
Merler S, Ajelli M, Pugliese A et al (2011) Determinants of the spatiotemporal dynamics of the 2009 H1N1 pandemic in Europe: implications for real-time modelling. PLoS Comput Biol 7(9):e1002205–e1002205. https://doi.org/10.1371/journal.pcbi.1002205
Minh HL, Sang-To T, Wahab MA et al (2022) A new metaheuristic optimization based on K-means clustering algorithm and its application to structural damage identification. Knowl Based Syst 251:109189. https://doi.org/10.1016/j.knosys.2022.109189
Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69(3):46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007
Mirjalili S, Mirjalili SM, Hatamlou A (2015) Multi-verse optimizer: a nature-inspired algorithm for global optimization. Neural Comput Appl 27(2):495–513. https://doi.org/10.1007/s00521-015-1870-7
Mirjalili S, Lewies A (2016) The whale optimization algorithm. Adv Eng Softw 95:51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008
Qi C, Zhang D, Zhu Y et al (2020) SARFIMA model prediction for infectious diseases: application to hemorrhagic fever with renal syndrome and comparing with SARIMA. BMC Med Res Methodol. https://doi.org/10.1186/s12874-020-01130-8
Rajakumar BR (2012) The lion's algorithm: a new nature-inspired search algorithm. Procedia Technol 6(1):126–135. https://doi.org/10.1016/j.protcy.2012.10.016
Rao RV, Savsani VJ, Vakharia DP (2011) Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. Comput-Aid Des 43(3):303–315. https://doi.org/10.1016/j.cad.2010.12.015
Rashedi E, Nezamabadi-Pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. Inform Sci 179(13):2232–2248. https://doi.org/10.1016/j.ins.2009.03.004
Rauf HT, Lali MIU, Khan MA et al (2021) Time series forecasting of COVID-19 transmission in Asia Pacific countries using deep neural networks. Pers Ubiquitous Comput. https://doi.org/10.1007/s00779-020-01494-0
Samieiyan B, Mohammadinasab P, Mollaei MA et al (2022) Solving dimension reduction problems for classification using promoted crow search algorithm (PCSA). Computing 104:1255–1284. https://doi.org/10.1007/s00607-021-01037-2
Shareef H, Lbrahim AA, Mutlag AH (2015) Lightning search algorithm. Appl Soft Comput 36:315–333. https://doi.org/10.1016/j.asoc.2015.07.028
Sha ZJ, Hu L, Rouyendegh BD (2020) Deep learning and optimization algorithms for automatic breast cancer detection. Int J Imag Syst Technol 30(2):495–506. https://doi.org/10.1002/ima.22400
Singh S, Sundram BM, Rajendran K et al (2020) Forecasting daily confirmed COVID-19 cases in Malaysia using ARIMA models. J Infect Dev Ctries. https://doi.org/10.3855/jidc.13116
Smith M, Alvarez F (2021) Identifying mortality factors from machine learning using shapley values –a case of COVID19. Expert Syst Appl 176:114832. https://doi.org/10.1016/j.eswa.2021.114832
Turabieh H, Azwari SA, Rokaya M et al (2021) Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance. Computing 103:1417–1438. https://doi.org/10.1007/s00607-020-00894-7
Velikovi P, Cucurull G, Casanova A, et al. (2017) Graph Attention Networks. https://doi.org/10.48550/arXiv.1710.10903
Wang B, Jin XP, Cheng B (2012) Lion pride optimizer: an optimization algorithm inspired by lion pride behavior. Sci China Inform Sci. https://doi.org/10.1007/s11432-012-4548-0
Wang J, Jia P, Cuadros DF et al (2017) A remote sensing data based artificial neural network approach for predicting climate-sensitive infectious disease outbreaks: a case study of human Brucellosis. Remote Sens. https://doi.org/10.3390/rs9101018
Wang J K, Zhang H, Jia P, et al. (2022a) City-level structured prediction and simulation model of COVID-19 J Comput Aid Des Comput Gr [J/OL] 1–11[2022a–08–30].http://kns.cnki.net/kcms/detail/11.2925.tp.2022a0801.1809.002.html.

Wang L, Liang C, Wu W et al (2019) Epidemic situation of brucellosis in Jinzhou city of China and prediction using the ARIMA model. Can J Infect Dis Med Microbiol. https://doi.org/10.1155/2019/1429462

Wang L, Zeng Y, Chen T (2015) Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. Expert Syst Appl 42:855–863. https://doi.org/10.1016/j.eswa.2014.08.018

Wang P, Zheng X, Ai G, et al. (2020) Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: case studies in Russia, Peru and Iran. Chaos Solitons Fractals. https://doi.org/10.1016/j.chaos.2020.110214

Wang HS, Tao GY, Ma JL et al (2022b) Predicting the epidemics trend of COVID-19 using epidemiological-based generative adversarial networks. IEEE J Sel Topics Signal Process 16(2):276–288. https://doi.org/10.1109/JSTSP.2022.3152375

Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Transactions Evolut Comput 1(1):67–82. https://doi.org/10.1109/4235.585893

Wu JT, Leung K, Leung GM (2020) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 395(100225):689–697. https://doi.org/10.1016/S0140-6736(20)30260-9

Wu XH, Zhao PW (2020) Predicting nuclear masses with the kernel ridge regression. Phys Rev C. https://doi.org/10.1103/PhysRevC.101.051301

Xie XJ, Luo KY, Zhang Y, et al. (2021) Nonlinear combinational dynamic transmission rate model and COVID-19 epidemic analysis and prediction in China. Oper Res Transactions, 25(01):17–30. https://doi.org/10.15960/j.cnki.issn.1007-6093.2021.01.002.

Yang ZF, Zeng ZQ, Wang K, et al. (2020) Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis, 12(3): 165–174. https://doi.org/10.21037/jtd.2020.02.64.

Yazdani M, Jolai F (2016) Lion optimization algorithm (LOA): a nature-inspired metaheuristic algorithm. J Comput Des Eng. https://doi.org/10.1016/j.jcde.2015.06.003

Yi CL, Cohnstaedt LW, Scoglio CM (2021) SEIR-SEI-EnKF: a new model for estimating and forecasting dengue outbreak dynamics. IEEE Access 9:156758–156767. https://doi.org/10.1109/ACCESS.2021.3129997

Yuchi W, Gombojav E, Boldbaatar B et al (2019) Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city. Environ Pollut 245:746–753. https://doi.org/10.1016/j.envpol.2018.11.034

Zar JH (2010) Biostatistical analysis. Q Rev Biol 18:797–799

Zhang X, Hou F, Qiao Z et al (2016) Temporal and long-term trend analysis of class C notifiable diseases in China from 2009 to 2014. BMJ Open. https://doi.org/10.1136/bmjopen-2016-011038

Zhao C, Yang Y, Wu S et al (2020a) Search trends and prediction of human brucellosis using Baidu index data from 2011 to 2018 in China. Scientific Rep. https://doi.org/10.1038/s41598-020-62517-7

Zhao L, Song Y, Zhang C et al (2020b) T-GCN: a temporal graph convolutional network for traffic prediction. IEEE Transactions Intell Transport Syst 21:3848–3858. https://doi.org/10.1109/TITS.2019.2935152