

RESEARCH

Open Access



# Research on hand, foot and mouth disease incidence forecasting using hybrid model in mainland China

Daren Zhao<sup>1</sup>, Huiwu Zhang<sup>1\*</sup>, Ruihua Zhang<sup>2,3\*</sup> and Sizhang He<sup>4</sup>

## Abstract

**Background** This study aimed to construct a more accurate model to forecast the incidence of hand, foot, and mouth disease (HFMD) in mainland China from January 2008 to December 2019 and to provide a reference for the surveillance and early warning of HFMD.

**Methods** We collected data on the incidence of HFMD in mainland China between January 2008 and December 2019. The SARIMA, SARIMA-BPNN, and SARIMA-PSO-BPNN hybrid models were used to predict the incidence of HFMD. The prediction performance was compared using the mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), and correlation analysis.

**Results** The incidence of HFMD in mainland China from January 2008 to December 2019 showed fluctuating downward trends with clear seasonality and periodicity. The optimal SARIMA model was SARIMA(1,0,1)(2,1,2)<sub>[1,2]</sub>, with Akaike information criterion (AIC) and Bayesian Schwarz information criterion (BIC) values of this model were 638.72, 661.02, respectively. The optimal SARIMA-BPNN hybrid model was a 3-layer BPNN neural network with nodes of 1, 10, and 1 in the input, hidden, and output layers, and the R-squared, MAE, and RMSE values were 0.78, 3.30, and 4.15, respectively. For the optimal SARIMA-PSO-BPNN hybrid model, the number of particles is 10, the acceleration coefficients c1 and c2 are both 1, the inertia weight is 1, the probability of change is 0.95, and the values of R-squared, MAE, and RMSE are 0.86, 2.89, and 3.57, respectively.

**Conclusions** Compared with the SARIMA and SARIMA-BPNN hybrid models, the SARIMA-PSO-BPNN model can effectively forecast the change in observed HFMD incidence, which can serve as a reference for the prevention and control of HFMD.

**Keywords** Prediction model, SARIMA model, SARIMA-BPNN hybrid model, SARIMA-PSO-BPNN hybrid model

\*Correspondence:

Huiwu Zhang  
cdzhanghuiwu@163.com

Ruihua Zhang  
cdzhangrh@126.com; zhangruihua@cdutcm.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Hand, foot, and mouth disease (HFMD) is an acute infectious disease caused by EV71 and the Cox A16 enterovirus, which spreads globally and is prevalent among children under five years of age [1–3]. HFMD is transmitted primarily through contact with the gastrointestinal and respiratory tracts and close contacts, and can develop throughout the year [4]. Most children with HFMD have mild symptoms, but a small percentage of infected individuals can develop severe disease [5, 6]. HFMD is a self-limiting disease that mainly manifests as fever and herpes on the hands, feet, and mouth [7]. Few children develop complications, such as myocarditis, pulmonary edema, and meningoencephalitis [7, 8]. HFMD is also a global infectious disease, and its prevalence has been reported in most regions of the world, especially in the Asia–Pacific and Western Pacific Region [3, 9]. It has been reported that 96,900 Disability Adjusted life years per year are due to HFMD in some countries in East and Southeast Asia, and HFMD causes a more severe economic burden of disease in these countries [10].

HFMD is not only a public health issue of global concern but has also become a widespread and typical infectious disease in mainland China. Several large outbreaks of HFMD occurred in 2007 and early 2008 in China; therefore, HFMD was included in the reporting of category C infectious diseases on May 2, 2008 [11]. Since HFMD was included in the management of category C infectious diseases of the Chinese Communicable Diseases Control Law, the number of cases of HFMD incidence and deaths has been ranked at the top of the list of legally reported infectious diseases in mainland China [12]. An average of approximately 2 million cases have been reported each year in 31 provinces and municipalities of mainland China [13]. HFMD causes a greater economic burden of disease in mainland China. A study on the economic burden of HFMD in mainland China showed that the average per capita cost of HFMD cases during treatment was 600–1,000 RMB for mild outpatient cases, 3,000–5,000 RMB for general inpatient cases, and 15,000–25,000 RMB for severe cases (without considering their impact on social productivity) [14]. Moreover, it is estimated that the direct economic burden of all severe HFMD cases in Jiangsu Province, China, was RMB 16.64 million during 2017–2018 [15]. Therefore, prevention and control of HFMD continues to be an important public health issue in mainland China.

Early surveillance and warning of HFMD are of high priority and important work. If the government and related departments can effectively monitor and provide accurate early warnings of HFMD, they will be able to respond in advance and provide information for the proper allocation of medical resources [16]. Therefore,

strengthening the surveillance and prediction of HFMD epidemiological trends in China is important for implementing effective preventive and control measures. Exploring approaches to enhance early monitoring and warning capabilities has become an urgent priority for improving China's public health system.

Many scholars have conducted extensive research on predicting the incidence of HFMD. Because the incidence of HFMD presents obvious seasonal and periodic characteristics, most current studies have focused on using the traditional time series auto-regressive integrated moving average (ARIMA) model for forecasting. Although the ARIMA model has achieved a better performance in predicting the incidence of HFMD [17, 18], there is still a failure to fully mine the nonlinear information of seasonal infectious disease data [19]. Some studies have focused on using machine learning models to predict the incidence of HFMD [20–22], but they may not explain the nonlinear functions within the time-series data in practice [23]. Moreover, a few studies have combined traditional time-series ARIMA models with machine learning models to develop hybrid models that have achieved better prediction performance [16, 23, 24]. However, hybrid models only combined the advantages of the two models, and there may be insufficient optimization of the model parameters. Therefore, the prediction performance of these models needs to be further improved.

To overcome the shortcomings of a single SARIMA for nonlinear information processing and the hybrid model with insufficiently optimized parameters, in the present study, we first proposed a SARIMA-PSO-BPNN hybrid model for forecasting the incidence of HFMD between January 2008 and December 2019 in mainland China. We constructed the SARIMA and SARIMA-BPNN hybrid models based on the data characteristics of HFMD incidence in mainland China and optimized the SARIMA-BPNN hybrid model using the Particle Swarm Optimization (PSO) algorithm. Predictions from the SARIMA-PSO-BPNN hybrid model can serve as an information reference for the surveillance and early warning of HFMD in mainland China.

## Methods

### Data source

Data on monthly HFMD incidence from January 2008 to December 2018 in 31 provinces and municipalities in mainland China were obtained from the China Public Health Science Data Center website (<https://www.phsciencedata.cn/Share/index.jsp>). The total number of HFMD cases from January to December 2019 was obtained from the National Health Commission of the People's Republic of China's website ([http://www.nhc.gov.cn/jkj/pgzdt/new\\_list.shtml](http://www.nhc.gov.cn/jkj/pgzdt/new_list.shtml)). The overall population

size in 2019 was obtained from the Chinese Statistical Yearbook (<http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm>). The average population per year was calculated as the population at the beginning and end of the year.

A total of 144 data on the monthly incidence of HFMD in mainland China from 2008 to 2019 were included in this study. We divided the HFMD incidence data into training and test sets. HFMD incidence data from January 2008 to December 2018 were used as the training set to construct the models, and data from January to December 2019 were used as the test set to evaluate the generalization capability of the models.

**SARIMA model**

Auto-regressive integrated moving average (ARIMA) model is a well-known time-series forecasting method proposed by Box and Jenkins in the early 1970s, also known as the Box-Jenkins model [25]. If the time series contains significant seasonal characteristics, the model can be identified as a SARIMA model. The SARIMA model is expressed as SARIMA (p, d, q) (P, D, Q)<sub>s</sub> and can be expressed as [26, 27]:

$$\nabla^d \nabla_S^D Y_t = \frac{\theta_q(B) \Theta Q(B^S)}{\phi_p(B) \Phi P(B^S)} \varepsilon_t \tag{1}$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p \tag{2}$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots - \theta_q B^q \tag{3}$$

$$\Phi P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2s} - \Phi_3 B^{3s} - \dots - \Phi_P B^{Ps} \tag{4}$$

$$\Theta Q(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2s} - \Theta_3 B^{3s} - \dots - \Theta_Q B^{Qs} \tag{5}$$

where p, q, P, and Q denote the order of auto-regression, the order of moving average, seasonal auto regression lag, and seasonal moving average, respectively. D and d denote the degree of seasonal and degree of trend differences, respectively, and s denotes the length of the seasonal period. Where B is the backward shift operator,  $Y_t$  is the HFMD incidence time-series at time t, and  $\varepsilon_t$  is the residual of the HFMD incidence time-series. Where  $\phi_p$  is the p-order auto-regressive coefficient polynomial,  $\theta_q$  is the q-order moving average coefficient polynomial,  $\Phi P(B^S)$  is the seasonal polynomial function of order P, and  $\Theta Q(B^S)$  is the seasonal polynomial function of order Q.

Four major steps are involved in the construction of the SARIMA model [19, 28]. The first step was to determine whether the time series were stationary. In general,

the stationary of a time series is determined by plotting the original time series chart or using methods such as Augmented Dickey-Fuller (ADF) tests. If the time series is non-stationary, it must be converted into a stationary time series using a difference or logarithmic transformation. The second step is to identify the parameters of the SARIMA model. The possible parameters of p, q, P, and Q were initially determined by plotting auto-correlation function (ACF) and partial auto-correlation function (PACF) charts. We then initially fitted the candidate SARIMA models based on the possible parameters p, q, P, and Q. The third step was to conduct model diagnosis. Residual tests were performed using the Ljung-Box Q test. Statistical significance of the model parameters was assessed using t-test and p-value. The fourth step is to identify the parameters of the SARIMA model and select the optimal model. The optimal model was selected based on the white noise residuals and the lowest AIC and BIC values.

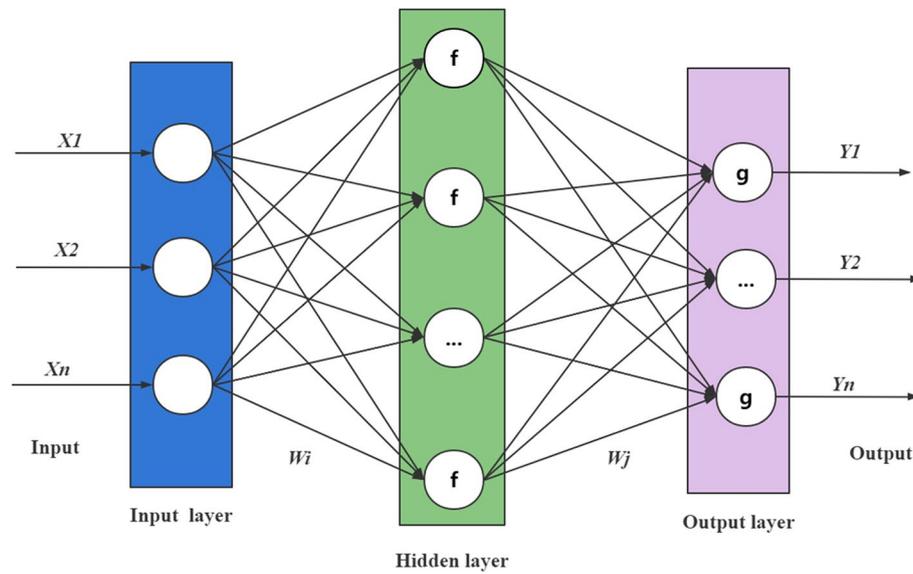
**BPNN model**

A backward propagation neural network (BPNN) is a multilayer feedforward neural network with output results using forward propagation and errors using backward propagation [29]. The main working principle of the BPNN is to use machine learning to continuously iterate the training model, calculate the error between the actual and expected output values based on the minimum mean squared error criterion, and adjust the weights and thresholds of each layer of the network using the gradient descent strategy to minimize the error [30]. A classical BPNN is a 3-layer neural network consisting of an input layer, hidden layer, and output layer with fully interconnected neurons between adjacent layers and unconnected neurons within the same layer [21].

There are three main steps in BPNN modeling [21, 31]: (1) initialization of the network and setting of network parameters, (2) normalization of the original data, dividing the training and test sets of the data, and back-propagation of the associated error calculation and adjustment of thresholds and weights, and (3) inverse normalization of the data to obtain the predicted values. The basic structure of a BPNN is shown in Fig. 1.

As shown in Fig. 1, we set up a 3-layer neural network with an input layer, hidden layer, and output layer. Assume that the input vector  $A = [X_1, X_2, X_3, \dots, X_i]$ , hidden layer input vector  $F = [f_1, f_2, f_3, \dots, f_n]$ , output layer input vector  $G = [g_1, g_2, g_3, \dots, g_j]$ , and actual output vector  $Y = [Y_1, Y_2, Y_3, \dots, Y_k]$ .

The hidden layer output ( $H_i$ ) is expressed as [32, 33]:



**Fig. 1** The basic structure of a BP neural network

$$H_i = f\left(\sum_{i=1}^n W_{ij} - \theta_j\right) \quad j = 1, 2, \dots, n \quad (6)$$

where  $i$  is the number of hidden nodes,  $W_{ij}$  is the connection weight of the input layer unit  $i$  to hidden layer unit  $j$ ,  $\theta_j$  is the threshold from the input layer to the hidden layer, and  $f$  is the excitation function.

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (7)$$

The prediction output ( $O_k$ ) is expressed as:

$$O_k = \sum_{j=1}^l H_j W_{jk} - \theta_k \quad (8)$$

where  $W_{jk}$  is the connection weight of the hidden layer unit  $j$  to the output layer unit  $k$  and  $\theta_k$  is the threshold from the hidden layer to the output layer.

And then the prediction error  $e$  is expressed as:

$$e_k = Y_k - O_k \quad (9)$$

where  $O_k$  is the prediction output, and  $Y_k$  is the actual output.

The number of hidden layer nodes was calculated using Eq. (10).

$$h = \sqrt{m + n} + a \quad (10)$$

where  $h$  is the number of hidden layer nodes and  $m$  and  $n$  are the numbers of input layer nodes and output layer

nodes, respectively. Where  $a$  is the adjustment constant between 1 and 10.

#### SARIMA-BPNN hybrid model

The SARIMA-BPNN hybrid model was developed similarly to that of the BPNN model. The modeling steps of the SARIMA-BPNN hybrid model are as follows: (1) The optimal SARIMA model was constructed using the HFMD incidence data from January 2008 to December 2018 in mainland China. (2) A 3-layer BPNN model was constructed using the predicted values from the optimal SARIMA model as the input variables and the observed value of the HFMD incidence data as the output variable. (3) According to the modeling steps of the BPNN model, the data were divided into a training set (70% of the data) and a test set (30% of the data) and then normalized before constructing the model. (4) The model with the trained BPNN was then simulated and the data obtained from the simulation were back-normalized to obtain the predicted values. The optimal SARIMA-BPNN hybrid model was determined by the largest R-squared value and lowest MAE and RMSE values.

#### SARIMA-PSO-BPNN hybrid model

Particle Swarm Optimization (PSO) is a heuristic search technique with simple implementation, high global search capability, and superior performance [34]. PSO simulates the foraging behavior of birds and is used to solve optimization problems [35]. PSO was introduced into the

BPNN model to accelerate the convergence of the traditional BPNN algorithm. The main modeling steps of the SARIMA-PSO-BPNN hybrid model are as follows [36]: (1) The parameters of the PSO algorithm were set based on the established SARIMA-BPNN hybrid model. The size of the population, variable range, inertia weight, learning factor, and maximum number of iterations were determined by many attempts. (2) Based on Eq. (11), the fitness value of each particle is calculated as follows [36]:

$$fitness = \frac{1}{1 + E} \tag{11}$$

where E is the training error precision.

And then, according to Eq. 12, the velocity and position of the particle are updated.

$$\begin{aligned} v_{i+1}(t + 1) &= \omega v_i(t) + c_1 r_1 (pbest_i(t) - x_i(t)) \\ &\quad + c_2 r_2 (gbest_i(t) - x_i(t)), x_{i+1}(t + 1) \\ &= x_i(t) + v_{i+1}(t + 1) \end{aligned} \tag{12}$$

where pbest and gbest are the best particle and swarm positions, respectively.  $x_i$  is the position vector;  $v_i$  is the velocity vector;  $c_1$  and  $c_2$  are the learning factors; and  $r_1$  and  $r_2$  are random values between 0 and 1. (3) The PSO-optimized weights and thresholds were substituted into the BPNN. The neural network optimized using PSO was trained using training samples until the error requirement was satisfied. Finally, an optimal SARIMA-PSO-BP hybrid model was constructed.

**Evaluation of prediction performance**

The mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), and correlation analyses were applied to comprehensively evaluate the prediction performance of the SARIMA, SARIMA-BPNN, and SARIMA-PSO-BPNN hybrid models. The smaller the values of MAE, MSE, RMSE, and MAPE, the better is the prediction performance of the model [37]. These indicators are expressed as follows:

$$MAE = \frac{\sum_{t=1}^n |X_t - \hat{X}_t|}{n} \tag{13}$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t)^2 \tag{14}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (X_t - \hat{X}_t)^2}{n}} \tag{15}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \hat{X}_t}{X_t} \right| \tag{16}$$

where  $\hat{X}_t$  is the predicted value,  $X_t$  is the observed value of the monthly HFMD incidence, and n is the sequence sample size.

Pearson’s and Spearman’s correlation coefficients were used to test the correlation between the predicted values of each model and observed values. A correlation coefficient with an absolute value closer to 1 indicates a stronger correlation between two variables [38, 39]. The strength of the correlation was evaluated [38, 39] as shown in Table 1.

**Data analysis**

R software (version 4.1.1) was used to construct the SARIMA model, and MATLAB software (Version R2020b, MathWorks, Natick, MA, USA) was used to construct the SARIMA-BPNN and SARIMA-PSO-BPNN hybrid models. The level of significance was set at  $p < 0.05$ .

**Results**

**General description**

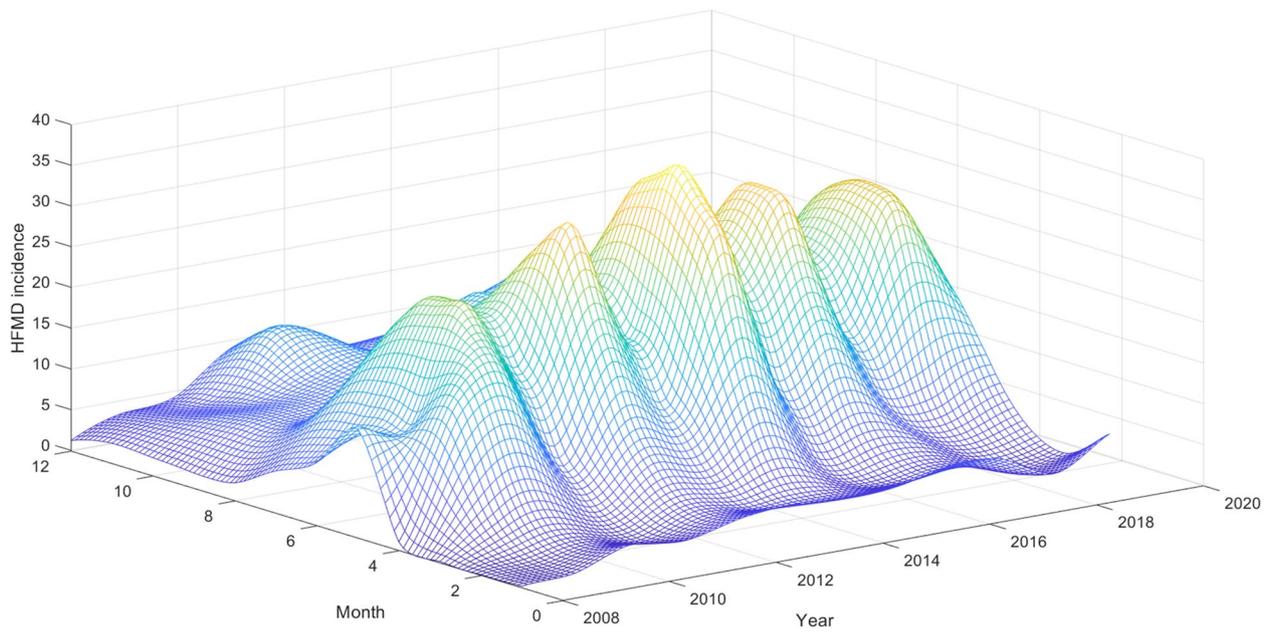
As shown in Fig. 2, the incidence of HFMD time series presented clear seasonality and periodicity patterns. The HFMD incidence rate increased at an average monthly rate of 2.56% from January 2008 to December 2019, with a peak incidence between May and June each year and a low incidence between January and February each year.

**SARIMA model**

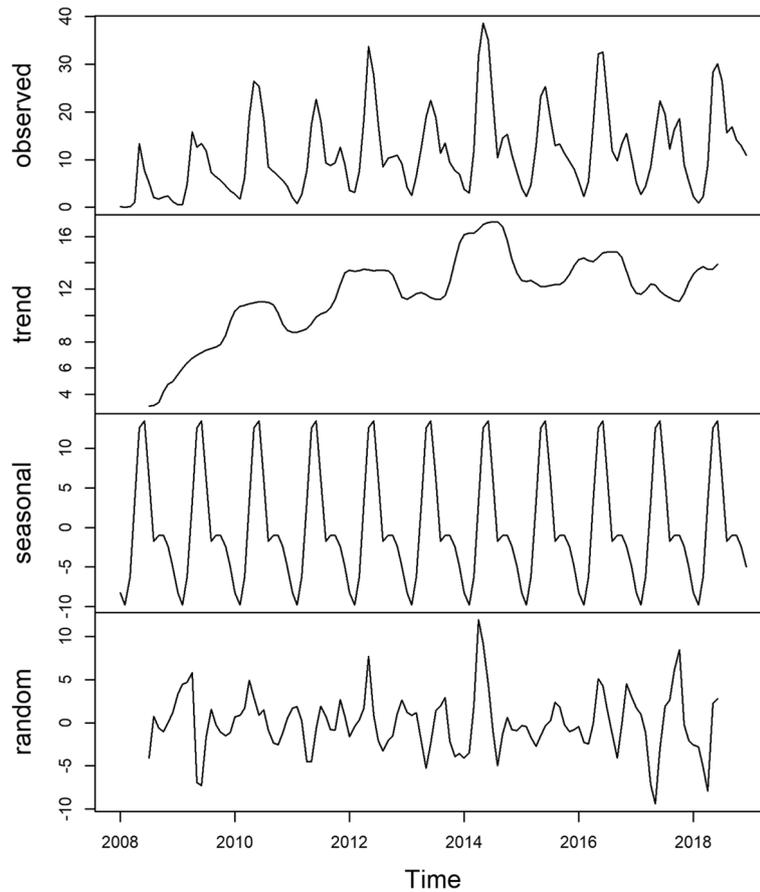
The decomposition () function in the R software was used to decompose the time series of HFMD incidence data from January 2008 to December 2019 in mainland China into seasonal, trend, and random components (Fig. 3). The results reconfirmed that the time series of

**Table 1** Correlation strength judgment

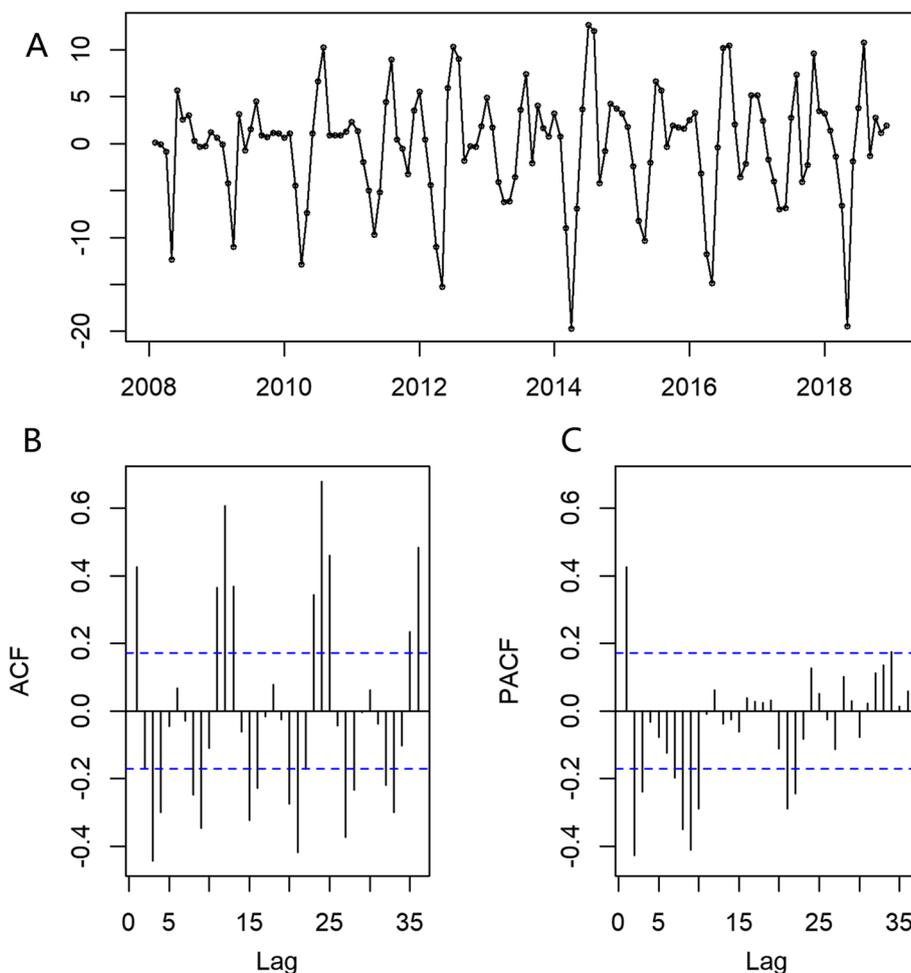
Pearson		Spearman	
Coefficients	Interpretation	Coefficients	Interpretation
0.00–0.10	Negligible correlation	0.00–0.30	Negligible correlation
0.10–0.39	Weak correlation	0.30–0.50	Weak correlation
0.40–0.69	Moderate correlation	0.50–0.70	Moderate correlation
0.70–0.89	Strong Correlation	0.70–0.90	Strong Correlation
0.90–1.00	Very Strong Correlation	0.90–1.00	Very Strong Correlation



**Fig. 2** Three-dimensional time series of HFMD incidence in China from January 2008 to December 2021



**Fig. 3** Decomposition of additive HFMD time series from January 2008 to December 2021 in mainland China



**Fig. 4** Differenced HFMD time series, ACF, and PACF plots (A) Differenced HFMD time series. B ACF plot of Differenced HFMD time series. C PACF plot of Differenced HFMD time series

HFMD incidence data from January 2008 to December 2019 in China had obvious seasonality, trends, periodicity, and randomness.

After a difference in the original HFMD data, the transformation time series became stationary. The Augmented Dickey-Fuller (ADF) test also confirmed this result ( $t = -6.264, p = 0.01$ ); thus,  $d$  was 0 or 1 and  $D$  was 1 (Fig. 4A). The ACF diagram shows an obvious peak at lag 1; thus, the value of the non-seasonal  $p$  may be 1. In addition, a spike at 12 in the first cycle indicated that the value of seasonal  $Q$  may be 1 or 2 (Fig. 4B). Similarly, the PACF plot showed a trailing trend at lag 2; thus, the value of seasonal  $P$  may be 2, and a spike at 1 or 2 in the first cycle indicates that non-seasonal  $q$  may be 1 or 2 (Fig. 4C). After the initial determination of the model parameters using ACF and PACF diagrams, the candidate SARIMA models were constructed (Table 2). The optimal SARIMA model was SARIMA(1,0,1)(2,1,2)<sub>[12]</sub>, which had the lowest AIC and

BIC values. All four candidate SARIMA models passed the Ljung-Box Q test, indicating that the residual series of these models was white-noise (Table 2). We then used the SARIMA(1,0,1)(2,1,2)<sub>[12]</sub> model to predict the HFMD incidence data from January to December 2019 in mainland China.

**Table 2** The candidate SARIMA models and Ljung-Box Q test

Candidate Models	AIC	BIC	Ljung-Box Q Statistics	p-value
SARIMA(1,1,1)(2,1,1) <sub>[12]</sub>	654.46	671.13	0.056	0.811
SARIMA(1,1,1)(2,1,2) <sub>[12]</sub>	652.94	672.39	0.060	0.805
SARIMA(1,0,1)(2,1,2) <sub>[12]</sub>	638.72	661.02	0.0004	0.984
SARIMA(1,0,2)(2,1,1) <sub>[12]</sub>	645.82	665.33	0.187	0.665
SARIMA(1,0,2)(2,1,2) <sub>[12]</sub>	644.78	667.07	0.170	0.680

**SARIMA-BPNN hybrid model**

The number of neural nodes in the BPNN hidden layer was calculated using Eq. (10) to yield a range of values of  $h$  between 2 and 11. Through repeated experiments, we set the number of nodes in the input, hidden, and output layers to 1, 10, and 1, respectively, and the epochs, learning rate, and minimum error of the training target to 1000, 0.01, and 0.00001, respectively, to obtain an optimal BPNN model. The R-squared, MAE, and RMSE values of the model were 0.78, 3.30, and 4.15, respectively.

**SARIMA-PSO-BPNN hybrid model**

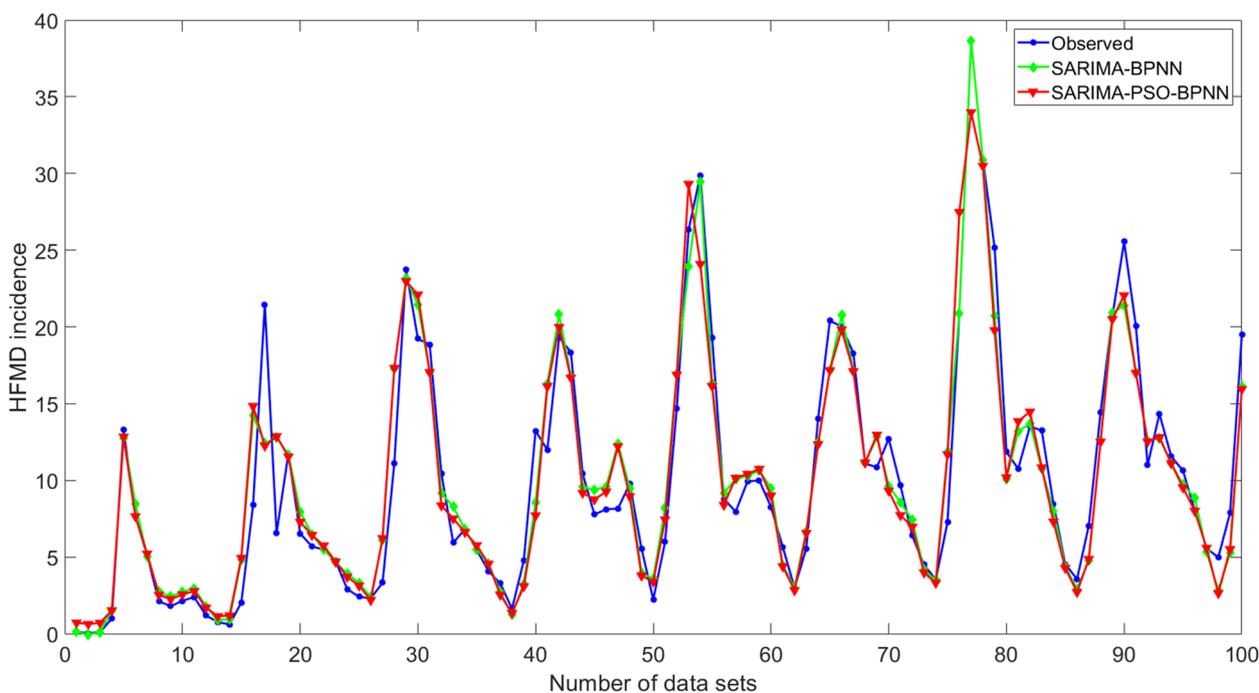
Based on the SARIMA-BPNN hybrid model, the SARIMA-PSO-BPNN hybrid model was constructed using the PSO algorithm. Therefore, the network structure of the SARIMA-PSO-BPNN hybrid model was consistent with that of the SARIMA-BPNN hybrid model. The number of population updates was set to 50, the population size to 10, the acceleration coefficient  $c1 = c2 = 1$ , the maximum velocity  $V_{max}$  to 2, the minimum velocity  $V_{min}$  to -1,  $popmax = [100, 10]$ ,  $popmin = [0.1, 0.6]$ , inertia weight to 1, and a probability of variation of 0.95, the optimal SARIMA-PSO-BPNN hybrid model was obtained. The R-squared, MAE, and RMSE values of the model were 0.86, 2.89, and 3.57, respectively. The predictions from the SARIMA-BPNN and SARIMA-PSO-BPNN hybrid models for the training and test sets are shown in Figs. 5 and 6.

**Prediction performance**

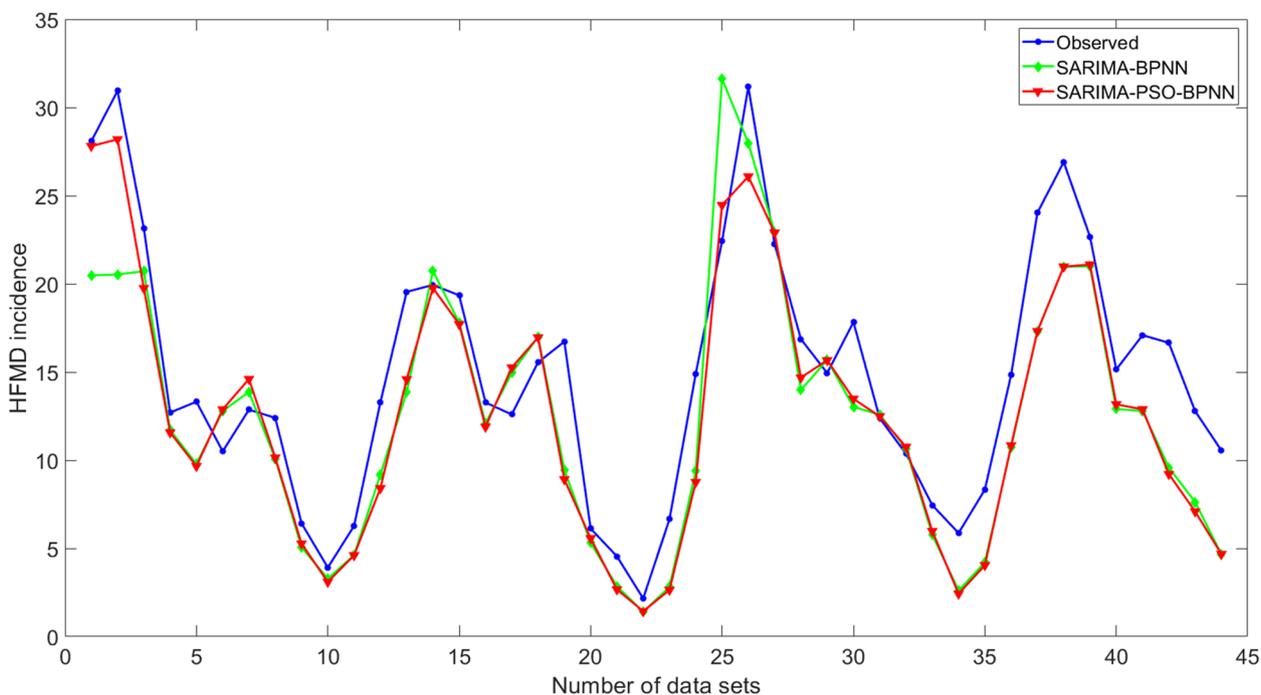
In the training set, except for the MAPE value, the MAE, MSE, and RMSE values of the SARIMA-PSO-BPNN hybrid model are smaller than those of the SARIMA and SARIMA-BPNN hybrid models. In the test set, the MAE, MSE, RMSE, and MAPE values of the SARIMA-PSO-BPNN hybrid model were all smaller than those of the SARIMA and SARIMA-BPNN hybrid models (Table 3). For both the training and test sets, the predicted values from the SARIMA, SARIMA-BPNN, and SARIMA-PSO-BPNN hybrid models are strongly correlated with the observed values (Table 4). The performance of the three models in predicting the incidence of HFMD from January to December 2019 is shown in Fig. 7.

**Discussion**

Our findings suggest that the incidence of HFMD in mainland China from January 2008 to December 2019 exhibited obvious seasonality, with a peak from May to June and a low peak from January to February, which is consistent with previous studies [2, 40, 41]. This is mainly because of two reasons. The first and most important reason for this is climatic factors. Numerous studies have indicated that the occurrence of HFMD is associated with meteorological factors, including temperature, humidity, sunshine, and wind speed [13, 42–44]. For example, the virulence and spread of enteroviruses under ambient conditions are mainly influenced by temperature and



**Fig. 5** Comparison of observed and predicted values from SARIMA-BPNN and SARIMA-PSO-BPNN hybrid models in the training set



**Fig. 6** Comparison of observed and predicted values from SARIMA-BPNN and SARIMA-PSO-BPNN hybrid models in the test set

**Table 3** Predictive performance of the three models in the training and test sets

Models	Training set				Test set			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
SARIMA	1.98	16.55	23.41	0.23	3.93	7.52	10.63	0.62
SARIMA-BPNN	1.11	12.91	18.25	0.13	0.96	2.41	3.41	0.08
SARIMA-PSO-BPNN	0.88	8.58	12.14	0.33	0.81	2.31	3.27	0.06

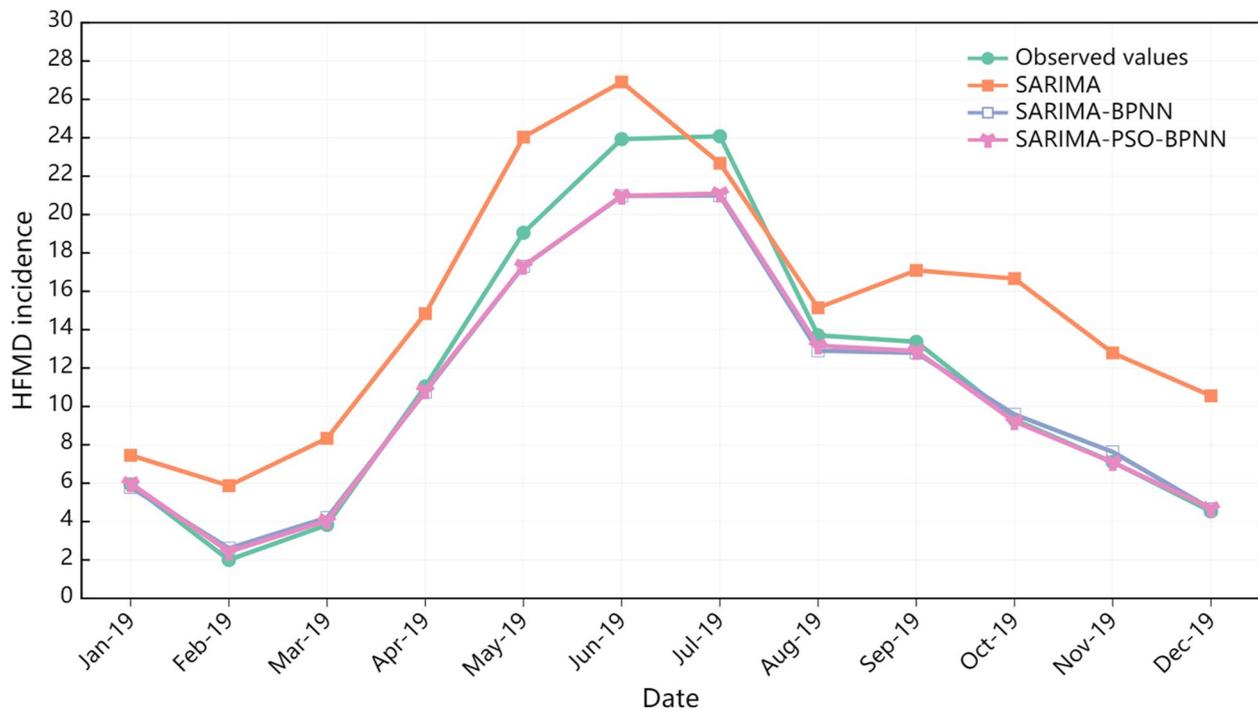
**Table 4** The correlations between the predicted and observed values of the three models

Models	Training set		Test set	
	Pearson	Spearman	Pearson	Spearman
SARIMA	0.94	0.93	0.95	0.92
SARIMA-BPNN	0.98	0.99	0.99	1.00
SARIMA-PSO-BPNN	0.99	1.00	0.99	1.00

relative humidity, with a higher survival rate at 20 °C, and more likely to survive at 80% humidity [41]. A previous study found that an increase of 200 mm in rainfall was associated with a 19% increase in the risk of developing HFMD in Vietnam [45]. Wang et al. [42] found a positive correlation between wind speed, sunshine, and hospitalization due to HFMD. The main reason for this meteorological factor is that the vast territory of China spans a

wide range of latitudes, most of which are located in the temperate zone, and some southern regions are located in subtropical and tropical regions, featuring a significant monsoon climate characterized by cold winters, hot summers, dry winters, and rainy summers [46].

Another significant factor influencing the development of HFMD is the holiday effect [47]. Previous studies have shown that holidays lasting more than a month may have a positive impact on reducing the transmission of HFMD from schools, whereas short holidays may have a limited impact on the transmission of HFMD [47, 48]. For primary and kindergarten schools in China, long public holidays occur twice a year, with annual winter and summer holidays. In general, winter holidays fall from January to February each year and summer holidays fall from July to August each year. These two public holidays last for approximately one-two months, offering relatively long holidays. Moreover, owing to seasonal changes in mainland China, meteorological factors such as temperature



**Fig. 7** Predictive performance of the three models

and humidity are beneficial for the transmission of HFMD between May and June each year [41]. Although primary and kindergarten schoolchildren have short holidays, such as May Day and the Dragon Boat Festival, which reduce the likelihood of their gathering, this short holiday has a limited impact on the transmission of HFMD in schools; instead, the incidence of HFMD was higher during this period.

Time-series data are collected at different times, which describe the changes in the state of something over time [49]. There are several time-series models, such as ARIMA, exponential smoothing, GARCH, VAR, and prophet models. However, ARIMA is one of the most classic time-series models and has been widely used to predict infectious diseases, including COVID-19[50], hepatitis B [28], tuberculosis [19], human brucellosis [51], HFMD [52], and pertussis [53]. SARIMA is a powerful forecasting tool in public health informatics [50] that provides an important reference for surveillance and early warning of infectious diseases. Several studies have confirmed that SARIMA achieves better predictive performance in forecasting the incidence of HFMD [16, 53–56]. This might be because of the capability of SARIMA models to effectively capture the nature of the dependency between current and past observations based on historical data while also considering the dynamic nature of infectious diseases [57]. Although the SARIMA model

has many advantages in the prediction of infectious disease incidence, it still has some shortcomings. For example, it is incapable of dealing with the nonlinear part of the information in an infectious disease time series [24].

To overcome the shortcomings of the SARIMA model, machine-learning forecasting models can effectively extract nonlinear relationships from data [21]. Previous studies have confirmed that the BPNN approach achieves a superior performance in predicting the incidence of HFMD. Liu et al. [21] found that the multivariate BPNN model could effectively forecast the HFMD incidence series from 2009 to 2016 in Jiangsu Province, China, because of its robustness, fault tolerance, and adaptive learning ability. Li et al. [58] confirmed that the BPNN model had higher forecasting accuracy and more accurate predictions than the ARIMA model. Furthermore, the BPNN is not required to satisfy strict assumptions, and can satisfy both linear and nonlinear mappings to handle complicated and multivariate time-series issues [59]. Therefore, in this study, the BPNN model was used to construct a hybrid model for comparison with SARIMA as the baseline model.

However, the BPNN also has some disadvantages; for instance, it is prone to fall into local minimum values, slow convergence speed, and poor training efficiency of the network [60–63]. To overcome the shortcomings of

the BPNN model, the PSO algorithm is introduced to construct the PSO-BPNN model. The PSO algorithm provides good global optimization capability by learning from population intelligence. The PSO algorithm was developed by optimizing the BPNN by replacing the gradient descent method to adjust the network weights and thresholds and to achieve an optimal BPNN model [60]. Therefore, the PSO-BPNN model combines the advantages of both the PSO algorithm and the BPNN model, and it can improve the accuracy of predictions, which has been confirmed in previous studies [60–62].

Therefore, we developed a basic SARIMA-BPNN model based on the SARIMA model. Then, we optimize the SARIMA-BPNN model with the PSO algorithm to obtain the optimal SARIMA-PSO-BPNN model. Compared to the results of similar studies, in the study, the MAPE value of the SARIMA-PSO-BPNN hybrid model in the test set (6%) was slightly lower than reported in the previous study (8.82%) [53]. Our results revealed that the prediction performance of the SARIMA-PSO-BPNN hybrid model (Table 4 and Fig. 7) outperformed the SARIMA and SARIMA-BPNN hybrid models, indicating that the SARIMA-PSO-BPNN hybrid model has a stronger generalization ability and provides excellent predictability. This may be because, in practice, the HFMD time series contains both linear and nonlinear complex time-series relationships [23]. The SARIMA model provides superior handling of linear information in the HFMD time series. The PSO algorithm optimizes the parameters of the SARIMA-BPNN hybrid model and improves the capability of the SARIMA-PSO-BPNN hybrid model to handle the complex time series relationships of HFMD; therefore, the SARIMA-PSO-BPNN hybrid model has a better generalization capability. These results are consistent with those of previous studies. Yu et al. [16] used SARIMA, NNAR, SARIMA-NNAR, and a wavelet-based SARIMA-NNAR hybrid model to predict the number of HFMD cases using data from 2009 to 2016 in Zhengzhou, China. Their study showed that the SARIMA-NNAR hybrid model demonstrated excellent prediction performance and was able to effectively forecast the incidence of HFMD. Zou et al. [23] indicated that the SARIMA-SVR hybrid model can accurately predict the incidence of HFMD and provide an effective decision-making tool for the prevention and control of HFMD in Wuhan, China.

However, our findings are inconsistent with those of previous studies. Yoshida et al. [64] used the LSTM model to forecast the incidence of HFMD in Japan, and their results showed that the LSTM approach could accurately estimate future epidemic patterns of HFMD in Japan. Meng et al. [65] used the XGBoost model and Random Forest model to predict the incidence of HFMD

from January 2009 to December 2017 in mainland China and found that the XGBoost model was more suitable for predicting the incidence of HFMD in mainland China. Zhang et al. [52] used ARIMA and LSTM to forecast the incidence of HFMD in Ningbo, China, and their study indicated that LSTM forecasting performance was superior to that of the ARIMA model. These differing findings may be due to the different choices of study areas and time period of the research.

This study had several limitations. First, data for our study were obtained from the China Public Health Science Data Center. Although the data are authoritative, there may be under-reporting and misreporting of HFMD cases in mainland China. Second, overall data on the incidence of HFMD in 31 provinces and municipalities in mainland China were collected, and we were unable to independently collect data on HFMD in southern and northern mainland China. However, there was a difference between the northern and southern incidences of HFMD [66]; therefore, there was some bias in the prediction of HFMD incidence in our study. Third, HFMD occurs because of several factors, including meteorological [11, 13, 43] and holiday effect factors [47]. However, in this study, the meteorological and holiday effect factors were excluded from the prediction model. In future studies, we will attempt to individually collect HFMD incidence data from 31 provinces and municipalities in mainland China, while considering meteorological and holiday effect factors are taken into account in the prediction model to obtain more accurate prediction results for HFMD surveillance and early warning in mainland China.

## Conclusions

The present study found that the SARIMA-PSO-BPNN hybrid model overcomes the problem of insufficient optimization of the parameters of the traditional hybrid model, which improves the accuracy of prediction and provides an information reference for early warning and surveillance of HFMD in mainland China.

## Abbreviations

HFMD	Hand, foot, and mouth disease
ARIMA	Auto-regressive integrated moving average
SARIMA	Seasonal auto-regressive integrated moving average
BPNN	Backward propagation neural network
AIC	Akaike information criterion
BIC	Bayesian Schwarz information criterion
AICc	Akaike Information Criterion, corrected
ADF	Augmented Dickey-Fuller
MAE	Mean absolute error
MSE	Mean square error
RMSE	Root mean square error
ACF	Auto-correlation function
PACF	Partial auto-correlation function

PSO	Particle swarm optimization
GARCH	Generalized autoregressive conditional heteroskedasticity
VAR	Vector auto-regression
NNAR	Neural network auto-regression
SVR	Support vector regression
XGBoost	Extreme gradient boosting
LSTM	Long short-term memory network

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12889-023-15543-9>.

### Additional file 1.

## Acknowledgements

We thank the Sichuan Provincial Primary Health Service Development Research Center (Grant no. SWFZ21-Q-59), and Sichuan Provincial Orthopedics Hospital (Grant No. 2021GL01) for funding this study.

## Authors' contributions

Daren Zhao, Huiwu Zhang, Ruihua Zhang conceived and designed the study. Daren Zhao, Sizhang He extracted and analyzed the data. Daren Zhao, Sizhang He wrote the manuscript. Daren Zhao, Huiwu Zhang, Ruihua Zhang revised the manuscript. All authors have contributed to the manuscript and approved the submitted version.

## Funding

This study was supported by the Sichuan Provincial Primary Health Service Development Research Center (Grant no. SWFZ21-Q-59), and Sichuan Provincial Orthopedics Hospital (Grant No. 2021GL01).

## Availability of data and materials

Data supporting the findings of this study are available from the China Public Health Science Data Center website (<https://www.phsciencedata.cn/Share/index.jsp>), the National Health Commission of the People's Republic of China website ([http://www.nhc.gov.cn/jkj/pgzdt/new\\_list.shtml](http://www.nhc.gov.cn/jkj/pgzdt/new_list.shtml)), and the Chinese Statistical Yearbook website (<http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm>) without restrictions.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Department of Medical Administration, Sichuan Provincial Orthopedics Hospital, Chengdu, Sichuan, People's Republic of China. <sup>2</sup>School of Management, Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, People's Republic of China. <sup>3</sup>General Practitioners Training Center of Sichuan Province, Chengdu, Sichuan, People's Republic of China. <sup>4</sup>Department of Information and Statistics, The Affiliated Hospital of Southwest Medical University, Luzhou 64600, Sichuan, China.

Received: 7 February 2023 Accepted: 28 March 2023

Published online: 31 March 2023

## References

- Xu L, Shi Y, Rainey JJ, Zhang Z, Zhang H, Zhao J, Li Y, Rao H, Li Y, Liao Q, Ma Y. Epidemiological features and spatial clusters of hand, foot, and

- mouth disease in Qinghai Province, China, 2009–2015. *BMC Infect Dis.* 2018;18(1):624. <https://doi.org/10.1186/s12879-018-3509-7>.
- Zhang X, Xu C, Xiao G. Space-time heterogeneity of hand, foot and mouth disease in children and its potential driving factors in Henan, China. *BMC Infect Dis.* 2018;18(1):638. <https://doi.org/10.1186/s12879-018-3546-2>.
- Zhang X, Xie R, Liu Z, Pan Y, Liu R, Chen P. Identifying pre-outbreak signals of hand, foot and mouth disease based on landscape dynamic network marker. *BMC Infect Dis.* 2021;21(Suppl 1):6. <https://doi.org/10.1186/s12879-020-05709-w>.
- Zhong X, Wang P, Zhang H. Spatial homogeneity pursuit of regression coefficients for hand, foot and mouth disease in Xinjiang Uygur Autonomous Region in 2018. *Sci Rep.* 2022;12(1):21439. <https://doi.org/10.1038/s41598-022-26003-6>.
- Wong SS, Yip CC, Lau SK, Yuen KY. Human enterovirus 71 and hand, foot and mouth disease. *Epidemiol Infect.* 2010;138(8):1071–89. <https://doi.org/10.1017/S0950268809991555>.
- Qi L, Tang W, Zhao H, Ling H, Su K, Zhao H, Li Q, Shen T. Epidemiological characteristics and spatial-temporal distribution of hand, foot, and mouth disease in Chongqing, China, 2009–2016. *Int J Environ Res Public Health.* 2018;15(2):270. <https://doi.org/10.3390/ijerph15020270>.
- Wang J, Li S. Nonlinear effect of temperature on hand, foot, and mouth disease in Lanzhou, China. *Medicine (Baltimore).* 2020;99(45):e23007. <https://doi.org/10.1097/MD.00000000000023007>.
- Lai FF, Yan Q, Ge SX, Tang X, Chen RJ, Xu HM. Epidemiologic and etiologic characteristics of hand, foot, and mouth disease in Chongqing, China between 2010 and 2013. *J Med Virol.* 2016;88(3):408–16. <https://doi.org/10.1002/jmv.24349>.
- Yu C, Xu C, Li Y, Yao S, Bai Y, Li J, Wang L, Wu W, Wang Y. Time series analysis and forecasting of the hand-foot-mouth disease morbidity in china using an advanced exponential smoothing state space TBATS model. *Infect Drug Resist.* 2021;14:2809–21. <https://doi.org/10.2147/IDR.S304652>.
- Koh WM, Badaruddin H, La H, Chen MI, Cook AR. Severity and burden of hand, foot and mouth disease in Asia: a modelling study. *BMJ Glob Health.* 2018;3(1):e000442. <https://doi.org/10.1136/bmjgh-2017-000442>.
- Ren FR, Cui Z, Zhang M, Wang YY, Zhang XR, Lu YQ. Effects of meteorological factors and atmospheric pollution on hand, foot, and mouth disease in Urumqi Region. *Front Public Health.* 2022;10:913169. <https://doi.org/10.3389/fpubh.2022.913169>.
- Wang B, Liu F, Liao Q, Wu P, Chang Z, Huang J, Long L, Luo L, Li Y, Leung GM, Cowling BJ, Yu H. Epidemiology of hand, foot and mouth disease in China, 2008 to 2015 prior to the introduction of EV-A71 vaccine. *Euro Surveill.* 2017;22(50):16–00824. <https://doi.org/10.2807/1560-7917.ES.2017.22.50.16-00824>.
- Xu J, Yang M, Zhao Z, Wang M, Guo Z, Zhu Y, Rui J, Wang Y, Liu X, Lin S, Luo L, Su Y, Zhao B, Zhou Y, Frutos R, Chen T. Meteorological factors and the transmissibility of hand, foot, and mouth disease in Xiamen City, China. *Front Med (Lausanne).* 2021;7:597375. <https://doi.org/10.3389/fmed.2020.597375>.
- Wang XL, Wei HX, Jia L, Huo D, Wang HQ, Wang QY. Summary of research in economic burden of hand, foot, and mouth disease in China. *Zhonghua Liu Xing Bing Xue Za Zhi.* 2020;41(2):273–9. <https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.023>.
- Han Y, Ji H, Shen W, Duan C, Cui T, Chen L, Hang H, Zhang Z, Sun H, Zhang X, Jin H. Disease burden in patients with severe hand, foot, and mouth disease in Jiangsu Province: a cross-sectional study. *Hum Vaccin Immunother.* 2022;18(5):2049168. <https://doi.org/10.1080/21645515.2022.2049168>.
- Yu G, Feng H, Feng S, Zhao J, Xu J. Forecasting hand-foot-and-mouth disease cases using wavelet-based SARIMA-NNAR hybrid model. *PLoS One.* 2021;16(2):e0246673. <https://doi.org/10.1371/journal.pone.0246673>.
- Liu L, Luan RS, Yin F, Zhu XP, Lü Q. Predicting the incidence of hand, foot and mouth disease in Sichuan province, China using the ARIMA model. *Epidemiol Infect.* 2016;144(1):144–51. <https://doi.org/10.1017/S0950268815001144>.
- Peng Y, Yu B, Wang P, Kong DG, Chen BH, Yang XB. Application of seasonal auto-regressive integrated moving average model in forecasting the incidence of hand-foot-mouth disease in Wuhan, China. *J Huazhong Univ Sci Technol Med Sci.* 2017;37(6):842–8. <https://doi.org/10.1007/s11596-017-1815-8>.

19. Zhao D, Zhang H, Cao Q, Wang Z, He S, Zhou M, Zhang R. The research of ARIMA, GM(1,1), and LSTM models for prediction of TB cases in China. *PLoS One*. 2022;17(2):e0262734. <https://doi.org/10.1371/journal.pone.0262734>.
20. Meng D, Xu J, Zhao J. Analysis and prediction of hand, foot and mouth disease incidence in China using Random Forest and XGBoost. *PLoS One*. 2021;16(12):e0261629. <https://doi.org/10.1371/journal.pone.0261629>.
21. Liu W, Bao C, Zhou Y, Ji H, Wu Y, Shi Y, Shen W, Bao J, Li J, Hu J, Huo X. Forecasting incidence of hand, foot and mouth disease using BP neural networks in Jiangsu province, China. *BMC Infect Dis*. 2019;19(1):828. <https://doi.org/10.1186/s12879-019-4457-6>.
22. Lin X, Wang X, Wang Y, Du X, Jin L, Wan M, Ge H, Yang X. Optimized Neural Network Based on Genetic Algorithm to Construct Hand-Foot-and-Mouth Disease Prediction and Early-Warning Model. *Int J Environ Res Public Health*. 2021;18(6):2959. <https://doi.org/10.3390/ijerph18062959>.
23. Zou JJ, Jiang GF, Xie XX, Huang J, Yang XB. Application of a combined model with seasonal autoregressive integrated moving average and support vector regression in forecasting hand-foot-mouth disease incidence in Wuhan, China. *Medicine (Baltimore)*. 2019;98(6):e14195. <https://doi.org/10.1097/MD.00000000000014195>.
24. Yu L, Zhou L, Tan L, Jiang H, Wang Y, Wei S, Nie S. Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in Shenzhen, China. *PLoS One*. 2014;9(6):e98241. <https://doi.org/10.1371/journal.pone.0098241>.
25. Wu WW, Li Q, Tian DC, Zhao H, Xia Y, Xiong Y, Su K, Tang WG, Chen X, Wang J, Qi L. Forecasting the monthly incidence of scarlet fever in Chongqing, China using the SARIMA model. *Epidemiol Infect*. 2022;150:e90. <https://doi.org/10.1017/S0950268822000693>.
26. Qiu H, Zhao H, Xiang H, Ou R, Yi J, Hu L, Zhu H, Ye M. Forecasting the incidence of mumps in Chongqing based on a SARIMA model. *BMC Public Health*. 2021;21(1):373. <https://doi.org/10.1186/s12889-021-10383-x>.
27. Zhu Z, Zhu X, Zhan Y, Gu L, Chen L, Li X. Development and comparison of predictive models for sexually transmitted diseases-AIDS, gonorrhea, and syphilis in China, 2011–2021. *Front Public Health*. 2022;10:966813. <https://doi.org/10.3389/fpubh.2022.966813>.
28. Zhao D, Zhang H, Cao Q, Wang Z, Zhang R. The research of SARIMA model for prediction of hepatitis B in mainland China. *Medicine (Baltimore)*. 2022;101(23):e29317. <https://doi.org/10.1097/MD.00000000000029317>.
29. Liu T, Zou G. Evaluation of mechanical properties of materials based on genetic algorithm optimizing BP neural network. *Comput Intell Neurosci*. 2021;2021:2115653. <https://doi.org/10.1155/2021/2115653>.
30. Zhang Z. Research on learning evaluation of online general education course based on BP neural network. *Comput Intell Neurosci*. 2021;2021:3570273. <https://doi.org/10.1155/2021/3570273>.
31. Hua L, Liu G. Development of basketball tactics basic cooperation teaching system based on CNN and BP neural network. *Comput Intell Neurosci*. 2021;2021:9497388. <https://doi.org/10.1155/2021/9497388>.
32. Fu W, Zhang H, Huang F. Internet-based supply chain financing-oriented risk assessment using BP neural network and SVM. *PLoS One*. 2022;17(1):e0262222. <https://doi.org/10.1371/journal.pone.0262222>.
33. He Z, Zhou S. BPNN-Based Behavioral Modeling of the S-Parameter Variation Characteristics of PAs with Frequency at Different Temperatures. *Micromachines (Basel)*. 2022;13(11):1831. <https://doi.org/10.3390/mi13111831>.
34. Zhang X, Tang Z. Construction of computer model for enterprise green innovation by PSO-BPNN algorithm and its impact on economic performance. *PLoS One*. 2022;17(1):e0262963. <https://doi.org/10.1371/journal.pone.0262963>.
35. Freitas D, Lopes LG, Morgado-Dias F. Particle swarm optimisation: a historical review up to the current developments. *Entropy (Basel)*. 2020;22(3):362. <https://doi.org/10.3390/e22030362>.
36. Wang L. Performance evaluation of knowledge sharing in an industry-university-research alliance based on PSO-BPNN. *Comput Intell Neurosci*. 2022;2022:1283588. <https://doi.org/10.1155/2022/1283588>.
37. Wang Y, Xu C, Zhang S, Wang Z, Yang L, Zhu Y, Yuan J. Temporal trends analysis of tuberculosis morbidity in mainland China from 1997 to 2025 using a new SARIMA-NARNNX hybrid model. *BMJ Open*. 2019;9(7):e024409. <https://doi.org/10.1136/bmjopen-2018-024409>.
38. Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg*. 2018;126(5):1763–8. <https://doi.org/10.1213/ANE.0000000000002864>.
39. Kumar N, Kumar P, Badagabettu SN, Lewis MG, Adiga M, Padur AA. Determination of spearman correlation coefficient (r) to evaluate the linear association of dermal collagen and elastic fibers in the perspectives of skin injury. *Dermatol Res Pract*. 2018;2018:4512840. <https://doi.org/10.1155/2018/4512840>.
40. Yang C, An S, Qiao B, Guan P, Huang D, Wu W. Exploring the influence of COVID-19 on the spread of hand, foot, and mouth disease with an automatic machine learning prediction model. *Environ Sci Pollut Res Int*. 2022;1–17. <https://doi.org/10.1007/s11356-022-23643-z>.
41. Chen C, Jiang Q, Song Z, Li Y, Wang H, Lu Y, Wang D, Li M, Li T. Influence of temperature and humidity on hand, foot, and mouth disease in Guangzhou, 2013–2017. *J Int Med Res*. 2020;48(6):300060520929895. <https://doi.org/10.1177/0300060520929895>.
42. Wang P, Goggins WB, Chan EY. Hand, foot and mouth disease in Hong Kong: a time-series analysis on its relationship with weather. *PLoS One*. 2016;11(8):e0161006. <https://doi.org/10.1371/journal.pone.0161006>.
43. Xie L, Huang R, Wang H, Liu S. Spatial-temporal heterogeneity and meteorological factors of hand-foot-and-mouth disease in Xinjiang, China from 2008 to 2016. *PLoS One*. 2021;16(8):e0255222. <https://doi.org/10.1371/journal.pone.0255222>.
44. Chen Y, Sun W, Ling F, Sun J, Cao Y, Chen Z, Miao Z. Seasonality and meteorological factors associated with different hand, foot, and mouth disease: serotype-specific analysis from 2010 to 2018 in Zhejiang Province. *China Front Microbiol*. 2022;13:901508. <https://doi.org/10.3389/fmicb.2022.901508>.
45. Van Pham H, Phan UTN, Pham ANQ. Meteorological factors associated with hand, foot and mouth disease in a Central Highlands province in Viet Nam: an ecological study. *Western Pac Surveill Response J*. 2019;10(4):18–23. <https://doi.org/10.5365/wpsar.2017.8.1.003>.
46. Zheng J, Yin Y, Li B. A new scheme for climate regionalization in China. *Acta Geogr Sin*. 2010;65(1):3–12.
47. Xie C, Wen H, Yang W, Cai J, Zhang P, Wu R, Li M, Huang S. Trend analysis and forecast of daily reported incidence of hand, foot and mouth disease in Hubei, China by Prophet model. *Sci Rep*. 2021;11(1):1445. <https://doi.org/10.1038/s41598-021-81100-2>.
48. Chen Y, Badaruddin H, Lee VJ, Cutter J, Cook AR. The effect of school closure on hand, foot, and mouth disease transmission in Singapore: a modeling approach. *Am J Trop Med Hyg*. 2018;99(6):1625–32. <https://doi.org/10.4269/ajtmh.18-0099>.
49. Lin Z, Sun X, Ji Y. Landslide displacement prediction based on time series analysis and double-BiLSTM model. *Int J Environ Res Public Health*. 2022;19(4):2077. <https://doi.org/10.3390/ijerph19042077>.
50. Duangchaemkarn K, Boonchieng W, Wiwatanadate P, Chouvatut V. SARIMA model forecasting performance of the COVID-19 daily statistics in Thailand during the Omicron Variant Epidemic. *Healthcare (Basel)*. 2022;10(7):1310. <https://doi.org/10.3390/healthcare10071310>.
51. Zhai M, Li W, Tie P, Wang X, Xie T, Ren H, Zhang Z, Song W, Quan D, Li M, Chen L, Qiu L. Research on the predictive effect of a combined model of ARIMA and neural networks on human brucellosis in Shanxi Province, China: a time series predictive analysis. *BMC Infect Dis*. 2021;21(1):280. <https://doi.org/10.1186/s12879-021-05973-4>.
52. Zhang R, Guo Z, Meng Y, Wang S, Li S, Niu R, Wang Y, Guo Q, Li Y. Comparison of ARIMA and LSTM in forecasting the incidence of HFMD combined and uncombined with exogenous meteorological variables in Ningbo, China. *Int J Environ Res Public Health*. 2021;18(11):6174. <https://doi.org/10.3390/ijerph18116174>.
53. Wang M, Pan J, Li X, Li M, Liu Z, Zhao Q, Luo L, Chen H, Chen S, Jiang F, Zhang L, Wang W, Wang Y. ARIMA and ARIMA-ERNN models for prediction of pertussis incidence in mainland China from 2004 to 2021. *BMC Public Health*. 2022;22(1):1447. <https://doi.org/10.1186/s12889-022-13872-9>.
54. Zhao D, Wang L, Cheng J, Xu J, Xu Z, Xie M, Yang H, Li K, Wen L, Wang X, Zhang H, Wang S, Su H. Impact of weather factors on hand, foot and mouth disease, and its role in short-term incidence trend forecast in Huainan City, Anhui Province. *Int J Biometeorol*. 2017;61(3):453–61. <https://doi.org/10.1007/s00484-016-1225-9>.
55. Tian CW, Wang H, Luo XM. Time-series modelling and forecasting of hand, foot and mouth disease cases in China from 2008 to 2018. *Epidemiol Infect*. 2019;147:e82. <https://doi.org/10.1017/S095026881800362X>.

56. Liu S, Chen J, Wang J, Wu Z, Wu W, Xu Z, Hu W, Xu F, Tong S, Shen H. Predicting the outbreak of hand, foot, and mouth disease in Nanjing, China: a time-series model based on weather variability. *Int J Biometeorol*. 2018;62(4):565–74. <https://doi.org/10.1007/s00484-017-1465-3>.
57. Wang Y, Xu C, Wang Z, Yuan J. Seasonality and trend prediction of scarlet fever incidence in mainland China from 2004 to 2018 using a hybrid SARIMA-NARX model. *PeerJ*. 2019;7:e6165. <https://doi.org/10.7717/peerj.6165>.
58. Li HY, Wang LS. A prediction model for hand, foot and mouth disease based on GA-BP neural network. *J Binzhou Univ*. 2018;34(6):47–51. <https://doi.org/10.13486/j.cnki.1673-2618.2018.06.009>.
59. Jia D, Xue R. Research on earnings management of growth enterprise market in China stock market: comparative analysis based on the BPNN, GBDT, and MLR Models. *Comput Intell Neurosci*. 2022;2022:6064536. <https://doi.org/10.1155/2022/6064536>.
60. Fu H. An internet of things-oriented adaptive mutation PSO-BPNN Algorithm to assist the construction of entrepreneurship evaluation models for college students. *Comput Intell Neurosci*. 2021;2021:3371383. <https://doi.org/10.1155/2021/3371383>.
61. Guo T. Evaluation of the emotion model in electronic music based on PSO-BP. *Comput Intell Neurosci*. 2022;2022:5601689. <https://doi.org/10.1155/2022/5601689>.
62. Liu X, Liu Z, Liang Z, Zhu SP, Correia JAFO, De Jesus AMP. PSO-BP neural network-based strain prediction of wind turbine blades. *Materials (Basel)*. 2019;12(12):1889. <https://doi.org/10.3390/ma12121889>.
63. Xi C. Music emotion analysis based on PSO-BP neural network and big data analysis. *Comput Intell Neurosci*. 2021;2021:6592938. <https://doi.org/10.1155/2021/6592938>.
64. Yoshida K, Fujimoto T, Muramatsu M, Shimizu H. Prediction of hand, foot, and mouth disease epidemics in Japan using a long short-term memory approach. *PLoS One*. 2022;17(7):e0271820. <https://doi.org/10.1371/journal.pone.0271820>.
65. Meng D, Xu J, Zhao J. Analysis and prediction of hand, foot and mouth disease incidence in China using Random Forest and XGBoost. *PLoS One*. 2021;16(12):e0261629. <https://doi.org/10.1371/journal.pone.0261629>.
66. Yang C, An S, Qiao B, Guan P, Huang D, Wu W. Exploring the influence of COVID-19 on the spread of hand, foot, and mouth disease with an automatic machine learning prediction model. *Environ Sci Pollut Res Int*. 2023;30(8):20369–85. <https://doi.org/10.1007/s11356-022-23643-z>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

