JIDC | THE JOURNAL OF INFECTION IN DEVELOPING COUNTRIES

Emerging Problems in Infectious Diseases

Improving surveillance of hand, foot, and mouth disease in China using ARIMAX model incorporating Baidu search index

Qianqian Huang[1,2 #], Huan Ning[1,2 #], Tingyan Luo[1,2 #], Jing Yang[1,2], Jie Zhou[1,2], Zhuoxin Li[1,2], Huanzhuo Mai[1,2], Ping Cui[3], Hao Liang[2,3], Jiegang Huang[1,2,4]

[1] *School of Public Health, Guangxi Medical University, Nanning 530021, Guangxi, China*
[2] *Guangxi Key Laboratory of AIDS Prevention and Treatment, Nanning 530021, Guangxi, China*
[3] *Life Science Institute, Guangxi Medical University, Nanning 530021, Guangxi, China*
[4] *Guangxi Colleges and Universities Key Laboratory of Prevention and Control of Highly Prevalent Diseases, Guangxi Medical University, Nanning 530021, Guangxi, China*

# Authors contributed equally to this work, and should be considered as co-first authors.

Abstract

Introduction: We developed an optimized autoregressive integrated moving average with explanatory variable (ARIMAX) model to predict the incidence trend of hand, foot, and mouth disease (HFMD) using recently updated data available on the internet.

Methodology: We collected HFMD incidence and keyword search data (2011–September 2024) from the National Administration of Disease Prevention and Control and the Baidu Index website. We analyzed the data by selecting relevant keywords through Spearman rank correlation and cross-correlation analysis, constructed a comprehensive search index (CSI), and developed autoregressive integrated moving average (ARIMA) and ARIMAX models; and followed by evaluating their predictive performance.

Results: A total of 11 HFMD keywords constituted the CSI, which showed a strong correlation coefficient of 0.93 with monthly incidence. Both ARIMA(1, 0, 1)(0, 1, 1)(12) and ARIMA(1, 0, 1)(0, 1, 1)(12) + CSI (Lag = 0) passed the least squares test ($p < 0.05$) and the residual Ljung-Box test ($p > 0.05$). Their Akaike information criterion (AIC) values were 2333.52 and 2263.41; predicted mean absolute percentage error (MAPE) values were 5.92% and 2.05%; and predicted mean absolute error (MAE) values were 95,858.53 and 38,300.36, respectively.

Conclusions: The ARIMA(1 ,0, 1)(0, 1, 1)(12) + CSI (Lag = 0) model demonstrates outstanding predictive accuracy. When combining the CSI model with the ARIMAX model in China, the fluctuations in HFMD cases caused by public health emergencies (such as the coronavirus disease 2019 (COVID-19) pandemic) will be reflected in the predicted incidence data of this model.

Key words: HFMD; surveillance; ARIMAX; forecasting.

J Infect Dev Ctries 2026; 20(10):1420-1429. doi:10.3855/jidc.21697

(Received 09 April 2025 – Accepted 07 February 2026)

Copyright © 2026 Huang *et al*. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Hand, foot, and mouth disease (HFMD) is a highly contagious viral infection predominantly affecting children under 5 years of age. It spreads through respiratory droplets, fecal-oral routes, and direct contact; often causing epidemic surges in children [1,2]. Although usually self-limiting, severe cases may develop fatal neurological or pulmonary complications [3]. China classified HFMD as a Class C infectious disease in 2008. More than 1 million cases of HFMD are reported annually, and HFMD consistently ranks among the top 3 Class C infectious diseases. Despite routine case reporting, delays in data transmission and processing persist due to the complex reporting and validation procedures [4,5].

In January 2020, China implemented stringent non-pharmaceutical interventions, including city and school closures in response to the coronavirus disease 2019 (COVID-19) pandemic. These measures effectively contained COVID-19 transmission and simultaneously suppressed other infectious diseases. Studies reported reduced incidence of several infections, such as human immunodeficiency virus (HIV) and hepatitis B, during lockdowns [6–8]. HFMD is transmitted by respiratory droplets and direct contact, and showed a sharp decline during lockdowns. However, case numbers resurged after the restrictions were lifted.

Predicting infectious disease trends is essential for governments and healthcare institutions to develop proactive strategies and improve public health awareness. The autoregressive integrated moving average (ARIMA) model that is widely used in disease surveillance [9] relies only on historical data. This dependence limits its predictive power during unprecedented events. To address this issue, the autoregressive integrated moving average with explanatory variable (ARIMAX) model incorporates external variables [10]. We built an ARIMAX model

using the Baidu Search Index (BSI) as an external variable. BSI is a feature of China's widely used Baidu search engine and has provided keyword search data since 2011 [11]. These data capture public interest in HFMD dynamics and enhance ARIMAX prediction accuracy.

In this study, we developed ARIMA and ARIMAX models; the latter incorporating BSI, to forecast HFMD incidence. Our goal was to identify a more accurate and reliable prediction model that supports timely monitoring and efficient allocation of healthcare resources. The models covered 2020 to 2022 data, when HFMD incidence declined sharply due to COVID-19 prevention and control policies [12,13]. We applied both models to forecast HFMD incidence from 2020 to September 2024 to examine the epidemic's influence on predictions, and reported both overall and yearly projections.

Methodology

Data sources

We collected monthly HFMD case counts from January 2011 to September 2024 from the official website of the National Administration of Disease Prevention and Control [14]. This publicly available dataset contains no personal information.

We obtained daily search query volumes for HFMD-related keywords from the official Baidu Index website [15].

Establishing a keyword database

Keywords for HFMD were gathered or filtered through the following methods:

(1) ChinaZ.com [16] is a keyword aggregation platform where searches for "hand, foot, and mouth disease" can automatically retrieve related terms from various search engines;
(2) Baidu Index Demand Map [15];
(3) Guidelines for the diagnosis and treatment of hand, foot and mouth disease (2018 edition) [17];
(4) We excluded keywords with very low daily search volume.

The final selection of keywords was categorized into 5 categories: comprehensive, etiology, symptoms, prevention, and treatment for hand, foot, and mouth disease [18].

Keyword analysis

We used R 4.3.3 to calculate the monthly search volume for each keyword and obtained monthly averages. We then conducted Spearman rank correlation and cross-correlation analyses between the keyword averages and HFMD case counts using SPSS 23.0 [19]. Only keywords with both absolute rank correlation coefficients and maximum cross-correlation values greater than 0.7 were included, and the lag values corresponding to the maximum cross-correlation were recorded. All statistical tests were two-tailed with a significance level of $\alpha = 0.05$.

Construction of the composite search index (CSI)

After identifying the keywords, we grouped them according to time lag coefficients, and calculated the sum of the correlation coefficient and the weight for each keyword. Subsequently, the weight and CSI were computed as follows:

$$Weight_{ki} = \rho ki \sum_{i=1}^{n} \rho ki$$

$$CSI_k = \sum_{i=1}^{n} Weight_{ki} \text{keyword}_{ki}$$

Where "k" denotes the potential time lag of the keyword, "n" indicates the number of keywords with the same potential time lag, "ρki" represents the Spearman rank correlation coefficient of keyword (i), "$weight_{ki}$" represents the weight of keyword (i) in its lag period, and "$keyword_{ki}$" represents the BSI value for keyword (i).

Components of ARIMA and ARIMAX models

We used R 4.3.3 to construct the ARIMA and ARIMAX models. The ARIMA model, a statistical tool for time series analysis and forecasting, is expressed as ARIMA(p, d, q)(P, D, Q)s. The ARIMAX model extends ARIMA by incorporating external variables to improve forecasting [20]. In this study, we introduced BSI as an external factor, and expressed the model as ARIMAX(p, d, q)(P, D, Q)s.

Smoothing of time series

We assessed the smoothness of the monthly HFMD case counts and the monthly CSI of HFMD by plotting time series diagrams, performing the augmented Dickey-Fuller (ADF) test, and conducting the Ljung-box test. If both tests yielded $p < 0.05$, no further differencing was needed. Otherwise, we continued differencing until the time series became smooth.

Splitting the training set and the test set

The monthly HFMD incidence and monthly CSI formed two time series spanning 1 January 2011 to 30 September 2024. We divided the data into a training set (1 January 2011 to 31 December 2019; 108 months) and a test set (1 January 2020, to 30 September 2024; 57

months). The test set was further divided into 4 subgroups: 2020–2021, 2021–2022, 2022–2023, and January–September 2024. This subdivision allowed us to examine the influence of the COVID-19 pandemic on HFMD incidence predictions and to evaluate the contribution of the BSI to model accuracy.

Model building and evaluation

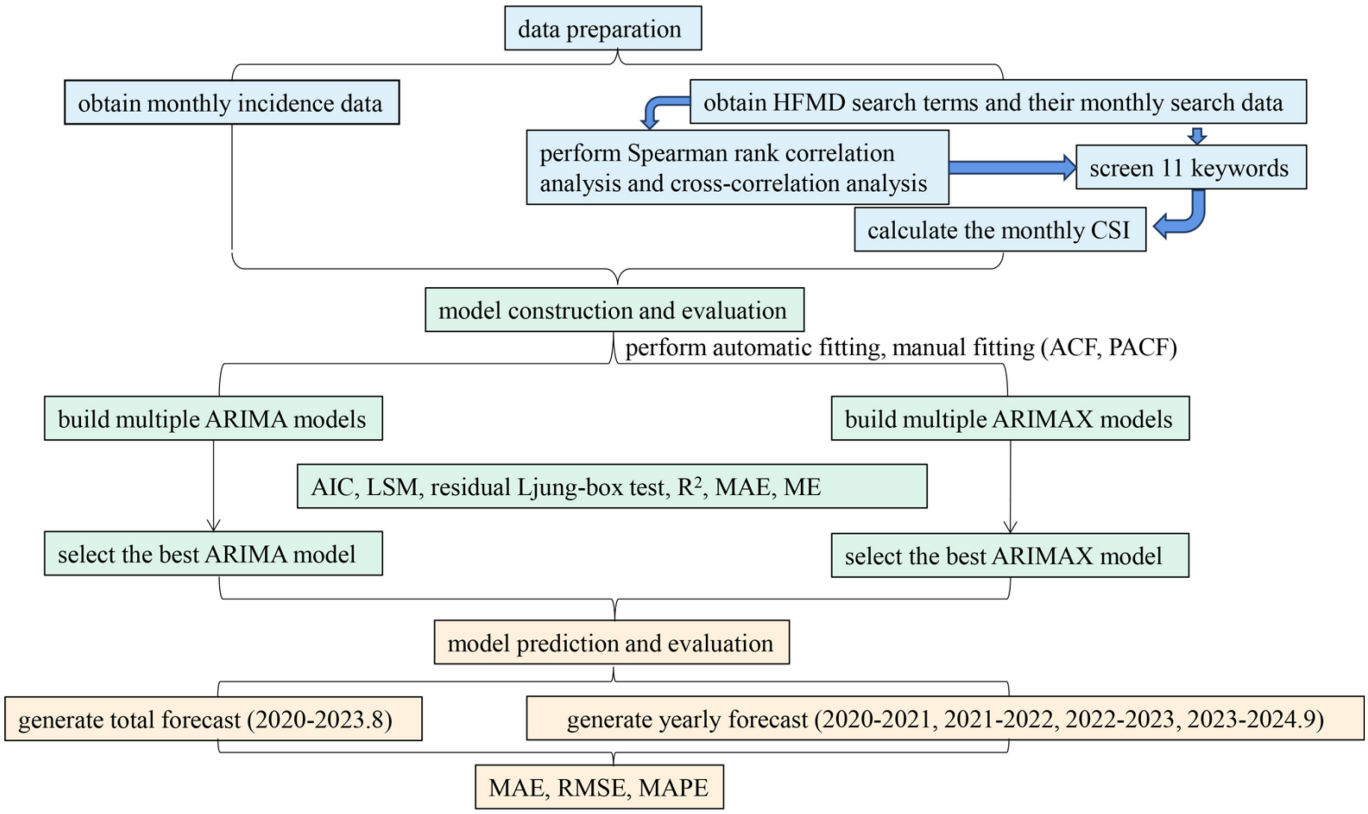
We conducted multiple model fittings using both automatic and manual parameter selection techniques. First, we performed automatic fitting with the auto.arima function to identify the optimal ARIMA and ARIMAX models. Next, we carried out manual fitting. Parameter selection was guided by plotting the autocorrelation function (ACF) and partial autocorrelation function (PACF) to determine the appropriate ranges for model orders (p and q) and the differencing values (d and D). The models were fitted sequentially, starting with the model with the lowest Akaike information criterion (AIC) and progressing through models with increasing AIC values.

We tested and diagnosed the validity of the models using the least squares method test, the residual Ljung-box test, and residual plots. Models that passed these diagnostic tests were considered valid, while those that failed were eliminated. In addition to prioritizing lower AIC values, we evaluated various performance metrics, including the coefficient of determination (R^2), mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE), to select the optimal model.

Model prediction and evaluation

We used the ARIMA and ARIMAX models to predict the overall HFMD cases from 1 January 2020 to 30 September 2024, and the annual case counts, respectively. Next, we plotted the predicted case trends against the actual case numbers. We evaluated prediction accuracy using metrics such as mean absolute error (MAE) and mean square error (MSE). The workflow for keyword selection and modelling is shown in the Figure 1.

Figure 1. Workflow for keyword selection and modelling.



HFMD: hand foot: and mouth disease; CSI: comprehensive search index; ACF: autocorrelation function; PACF: partial autocorrelation function; ARIMA: autoregressive integrated moving average; ARIMAX: autoregressive integrated moving average with explanatory variable; AIC: Akaike information criterion; LSM: least squares method; MAE: mean absolute error; ME: mean error; RMSE: root mean square error; MAPE: mean absolute percentage error.

Results

Overview of the incidence of HFMD in China from 2011 to 2024

The National Center for Disease Control reported a total of 24,661,454 cases of HFMD, with an average monthly incidence of 149,463 cases from 1 January 2011 to 30 September 2024. HFMD displayed a bimodal seasonal pattern, with peaks from May–July, and again from September–November each year. The annual number of reported HFMD cases remained relatively stable from 2011 to 2019, with monthly incidences ranging from 130,000 to 230,000 cases. However, the average monthly incidence dropped significantly to 50,000–110,000 cases during 2020–2022 due to the COVID-19 pandemic and related control measures. The incidence rebounded to pre-pandemic levels in 2023.

HFMD keyword screening and CSI construction

We identified 90 HFMD-related keywords, which we grouped into 5 categories: HFMD syndrome, etiology, symptoms, prevention, and treatment (Supplementary Table 1).

Spearman rank correlation analysis between keywords monthly average and the HFMD cases

We performed Spearman rank correlation analysis on the training set, which included the monthly search index of HFMD keywords and the monthly HFMD incidences from January 2011 to December 2019. Twelve keywords were initially screened according to the criteria of rs > 0.7 and $p < 0.05$, as shown in Table 1. Among them, the keywords "儿童手足口病 (hand, foot and mouth disease in children)" (Figure 2A) and "手足口 (hand, foot and mouth)" (Figure 2B) showed the strongest correlations with HFMD cases, reaching 0.928 and 0.914, respectively.

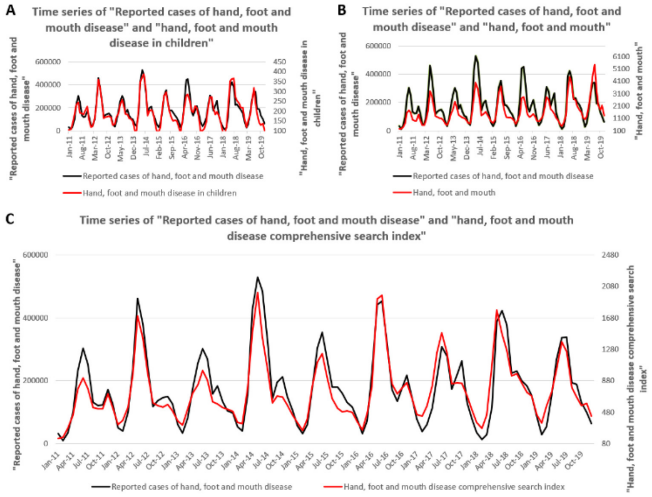
Cross-correlation analysis of keywords monthly average and the HFMD cases

The cross-correlation coefficients of the 11 key words exceeded 0.7 (Table 2). "儿童手足口病 (hand, foot and mouth disease in children)" remained the keyword that was most strongly associated with the monthly cases.

Spearman rank correlation analysis between CSI and HFMD cases

We weighted the correlation coefficients between the 11 keywords and the number of cases to construct the Baidu CSI. The correlation between the number of HFMD cases and the CSI in China was $R^2 = 0.93$ ($p < 0.001$), indicating a statistically significant association. Moreover, the trends of the two time series were

Figure 2. Sequence diagram of the number of HFMD cases and two keywords.



Time series between HFMD reported cases and (A) the keywords "hand: foot and mouth disease in children" (B) the keywords "hand: foot and mouth" or (C) the comprehensive search index of hand: food and mouth disease. HFMD: hand foot: and mouth disease.

Table 1. Spearman rank correlation analysis of keywords monthly search index and HFMD incidence.

Keywords [a]	r_s [b]	*p* value	Search volume, mean (SD)
手足口 (hand, foot and mouth)	0.914	< 0.001	1661.10 (1003.40)
小儿手足口病 (hand, foot and mouth disease in children)	0.803	< 0.001	259.38 (131.95)
儿童手足口病 (hand, foot and mouth disease in children)	0.928	< 0.001	195.01 (70.93)
什么是手足口病 (what is hand, foot and mouth disease?)	0.788	< 0.001	284.90 (152.61)
小孩手足口病症状 (hand, foot and mouth disease symptoms in children)	0.802	< 0.001	457.14 (254.84)
手足口并发症 (hand, foot and mouth complications)	0.858	< 0.001	33.14 (28.76)
手足口病的症状 (symptoms of hand, foot and mouth disease)	0.767	< 0.001	533.51 (463.11)
手足口病图片 (hand, foot and mouth disease pictures)	0.726	< 0.001	1944.28 (1846.56)
如何预防手足口病 (how to prevent hand, foot and mouth disease?)	0.733	< 0.001	188.40 (106.26)
手足口病预防 (hand, foot and mouth disease prevention)	0.797	< 0.001	181.05 (93.94)
手足口病用药 (hand, foot and mouth disease medications)	0.842	< 0.001	87.97 (42.12)
手足口病吃什么药 (what medicine to take for hand, foot and mouth disease?)	0.812	< 0.001	690.44 (449.40)

[a] Keywords are presented in a Chinese (English) format; [b] r_s : Spearman rank correlation; HFMD: hand foot and mouth disease.

Table 2. Cross-correlation analysis between monthly keyword search index and HFMD incidence.

Keywords [a]	Lag (months)	Maximum CCF [b] (SE)
手足口 (hand, foot, and mouth)	0	0.864 (0.096)
儿童手足口病 (hand, foot, and mouth disease in children)	0	0.800 (0.096)
小孩手足口病症状 (hand, foot, and mouth disease symptoms in children)	0	0.813 (0.096)
儿童手足口病 (hand, foot, and mouth disease in children)	0	0.939 (0.096)
手足口病用药 (hand, foot and mouth disease medications)	0	0.813 (0.096)
手足口病吃什么药 (what medicine to take for hand, foot, and mouth disease?)	0	0.801 (0.096)
手足口并发症 (hand, foot, and mouth complications)	0	0.836 (0.096)
什么是手足口病 (what is hand, foot and mouth disease?)	0	0.797 (0.096)
手足口病图片 (hand, foot, and mouth disease pictures)	0	0.702 (0.096)
如何预防手足口病 (how to prevent hand, foot, and mouth disease?)	0	0.789 (0.096)
手足口病预防 (hand, foot and mouth disease prevention)	– 1	0.864 (0.097)

[a] Keywords were presented in Chinese (English) format; [b] CCF: cross-correlation function; HFMD: hand foot and mouth disease.

essentially coincident (Figure 2C).

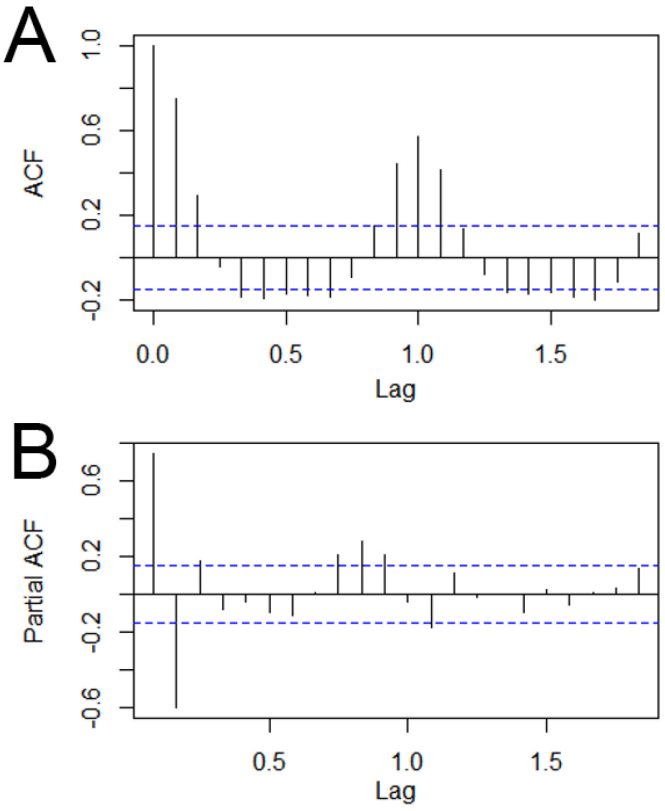
Time series flattening

The time series diagram of HFMD cases and the monthly CSI from January 2011 to September 2024 showed a stable overall trend, despite a decrease in cases from 2020 to 2024. The ADF test indicated p = 0.01 for both sequences, suggesting that no differencing was initially necessary. However, the HFMD incidence trend exhibited a seasonal cyclical pattern, with higher cases from May to July, and from September to November. Similarly, the CSI showed double peaks during these periods each year. We applied a first-order seasonal difference to render both series stationary (Supplementary Figure 1). After differencing, the ADF test for the HFMD incidence series resulted in p = 0.04, and for the CSI series, p = 0.03. Therefore, d was set to 0 and D to 1 when manually adjusting the model parameters. The seasonal cycle of HFMD is 12 months; therefore, s = 12.

Model ordering

According to the ACF and PACF plots of the monthly HFMD incidence (Figure 3), values of p and q can be considered as 1 or higher. Consequently, we fixed d = 0 and D = 1 during manual parameter tuning, and typically chose values of P and Q as 0, 1, or 2. We tested the values of P and Q within the range 0–2, adding them to the model one by one, starting from the lowest to the highest. Cross-correlation function (CCF) plots indicated that the CSI was most strongly correlated with HFMD incidence at a lag of 0 months (Supplementary Figure 2). Therefore, we constructed the ARIMAX model by incorporating the monthly incidence counts along with the CSI lagged by 0 months. After considering metrics such as Akaike information criterion (AIC), mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE), and manually adjusting the parameters, we ultimately selected the models ARIMA(1, 0, 1)(0, 1, 1)(12) and ARIMA(1, 0, 1)(0, 1, 1)(12) + CSI (lag = 0). The auto arima function automatically selected ARIMA(1, 0, 1)(1, 1, 0)(12) as the best model, incorporating seasonal differencing for the HFMD incidence series. Meanwhile, the best ARIMAX model, ARIMAX(1, 1, 2)(2, 0, 0)(12), applied time-series differencing to both the incidence series and the CSI series.

Figure 3A. ACF and **B:** PACF plots of HFMD.



ACF: autocorrelation function; PACF: partial autocorrelation function. HFMD: hand foot: and mouth disease.

Table 3. Evaluation of model fit and diagnosis.

Model	ADF[a], *p* value	Diagnosis			Fit				
		LSM[b], *P* value	Ljung-Box test		AIC[c]	R^2	RMSE[d]	MAE[e]	MAPE[f]
			Chi square	*p* value					
ARIMA[g](1, 0, 1)(1, 1, 0) $_{(12)}$	0.01	< 0.001	< 0.001	0.97	2330.43	0.89	38997.47	27115.94	18.66
ARIMAX[h](1, 1, 2)(2, 0, 0) $_{(12)}$	0.01	< 0.001	0.003	0.95	2528.39	0.93	30058.62	22212.50	19.23
ARIMA(1, 0, 1)(0, 1, 1) $_{(12)}$	0.01	< 0.001	0.059	0.81	2333.52	0.91	36039.66	24872.27	16.80
ARIMA(1, 0, 1)(0, 1, 1)$_{(12)}$ + CSI[i](lag = 0)	0.01	0.025	0.003	0.96	2263.41	0.96	25754.80	17045.40	12.23

[a] ADF: augmented Dickey-Fuller; [b] LSM: least squares method; [c] AIC: Akaike information criterion; [d] RMSE: root mean square error; [e] MAE: mean absolute error; [f] MAPE: mean absolute percentage error; [g] ARIMA: autoregressive integrated moving average; [h] ARIMAX: autoregressive integrated moving average with explanatory variable; [i] CSI: comprehensive search index.

Model validation

The manually selected ARIMA(1, 0, 1)(0, 1, 1)(12) model (model 1) had lower AIC, MAE, and MAPE values than the automatically selected ARIMA(1, 0, 1)(1, 1, 0)(12) model. Similarly, the manually selected ARIMA(1, 0, 1)(0, 1, 1)(12) + CSI (lag = 0) model (model 2) had lower parameter values than the automatically selected ARIMAX(1, 1, 2)(2, 0, 0)(12). These results indicate that the manually selected models provided a better fit and were therefore chosen for subsequent forecasting. The validity of all models was confirmed, with $p < 0.05$ for the least squares method test and $p > 0.05$ for the residuals Ljung box tests. Specific model parameters are presented in Table 3 and the Supplementary Table 2.

Models 1 (Figure 4) and 2 (Figure 5) effectively captured information from the original series and were suitable for forecasting. This was supported by Q-Q plots closely resembling a normal distribution (Figures 4B, 5B), residual plots centered around 0 (Figures 4C, 5C), and residuals ACF and PACF plots mostly within the 95% confidence interval (Figures 4D–E, 5D–E).

Total prediction

Figures 4A and 5A compare actual HFMD incidences from 2019 to September 2024 with model predictions, showing similar trends. The ARIMA model, relying solely on past data, predicted values close to historical levels. However, the unforeseen impact of COVID-19 drastically reduced HFMD incidence, causing significant deviations from the ARIMA predictions. In contrast, the ARIMAX model, which incorporated the BSI, aligned better with actual cases and displayed a superior curve fit. Evaluation metrics indicate that the ARIMAX model, with lower MAE, MAPE, and RMSE values, achieved higher prediction accuracy compared to the ARIMA model (Table 4, Supplementary Table 3).

Subgroup predictions

The original incidence sequence showed that HFMD cases declined significantly from 2020 to 2022, likely due to the prolonged COVID-19 pandemic. Therefore, both models were used to forecast annual HFMD incidence from 2020 to September 2024, and their yearly prediction performances were compared. The Supplementary Table 3 presents detailed model fitting and prediction results, showing that the prediction performance of both models decreased when actual incidence dropped sharply due to unforeseen events. However, the ARIMAX model, which incorporates online search data, consistently produced smaller prediction errors than the ARIMA model across all 4 subgroups.

Discussion

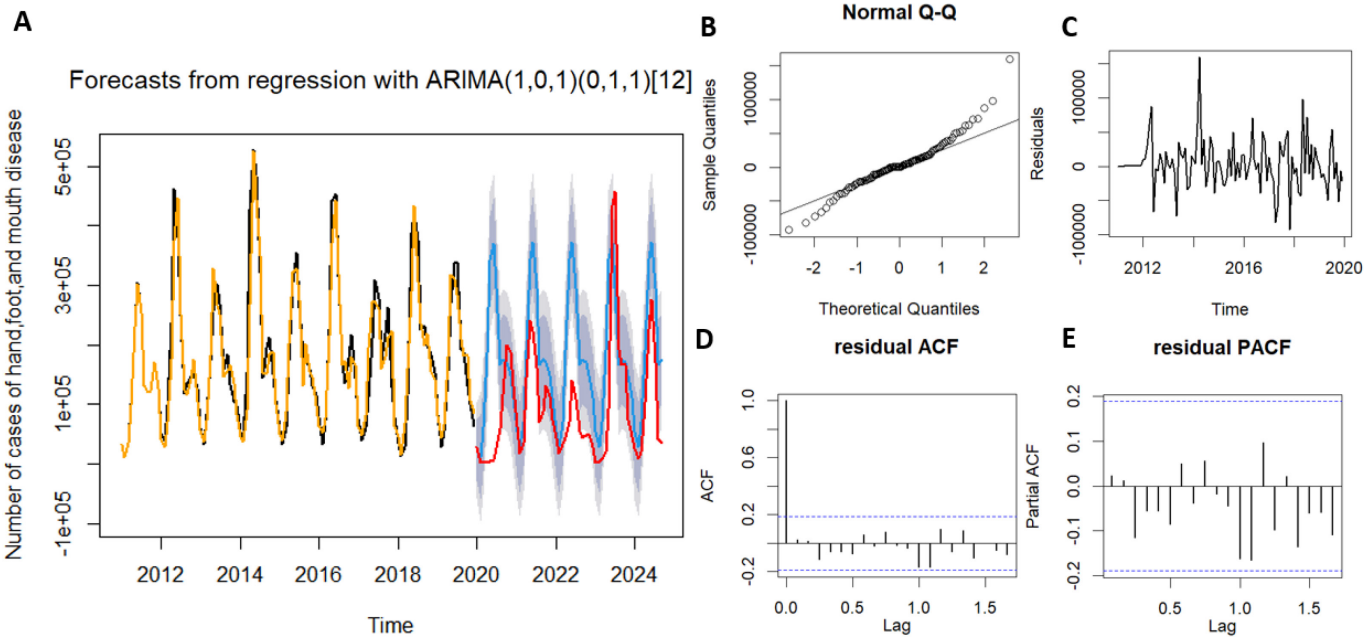
HFMD, a highly contagious enterovirus infection predominantly caused by EV-71 and CA-16, has caused global outbreaks in recent decades [21,22]. Although China's EV-71 vaccination program has reduced HFMD incidence, persistent disease burden and emerging atypical strains continue to pose pediatric health risks [23,24]. China reported over 23 million HFMD cases from 2011 to August 2023. The incidence of various infectious diseases declined in China during the COVID-19 pandemic (2020–2022), primarily due to rigorous government controls and heightened hygiene awareness [6,25]. HFMD incidence also decreased significantly during this period though resurgence occurred when the restrictions were eased [12,26]. Therefore, when conditions permit, we recommend the use of non-pharmacological intervention measures and emphasizing personal hygiene to carry out routine prevention of HFMD [24].

Table 4. Model prediction performance.

Model	MAE[a] (95% CI)	RMSE[b] (95% CI)	MAPE[c], % (95% CI)
ARIMA[d](1, 0, 1)(0, 1, 1) $_{(12)}$	95858.53 (62847.62, 128869.44)	127156.58 (94145.67, 160167.49)	5.92 (1.60, 10.25)
ARIMA(1, 0, 1)(0, 1, 1) $_{(12)}$ + CSI[e] (lag = 0)	38300.36 (24240.18, 52360.53)	54159.18 (40099.01, 68219.36)	2.05 (0.43, 3.67)

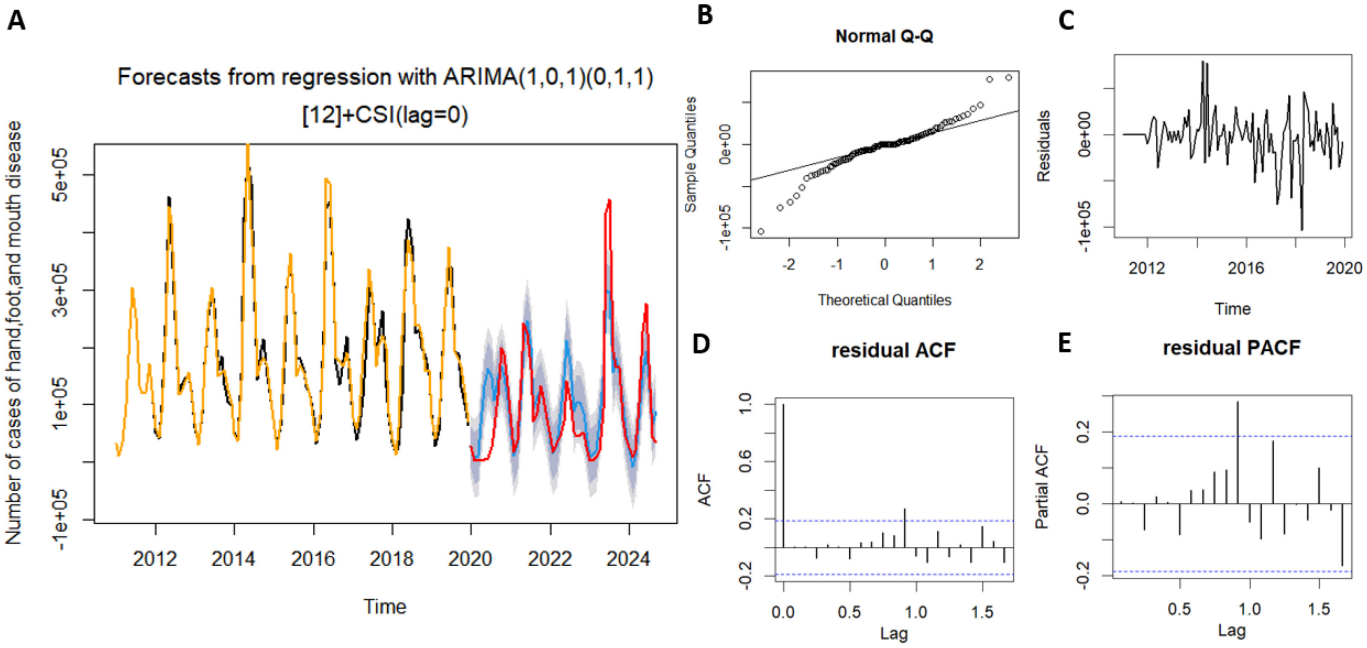
[a] MAE: mean absolute error; [b] RMSE: root mean square error; [c] MAPE: mean absolute percentage error; [d] ARIMA: autoregressive integrated moving average; [e] CSI: comprehensive search index.

Figure 4. ARIMA (1: 0: 1)(0: 1: 1)(12) model fitting: diagnosis: and prediction results.



A: Fitting and prediction results. Yellow and black are the fitted and true values of the training sets respectively. Red indicates the true values of the testing sets: and blue indicates the predicted values. The gray area represents the 95% confidence interval of the predicted value. **B:** Normal Q-Q plot. The solid black line represents the theoretical quantile line. **C:** Residual plot. **D:** ACF plot. **E:** PACF plot. The blue dashed lines in **D** and **E** indicate the confidence interval. ARIMA: autoregressive integrated moving average; ACF: autocorrelation function; PACF: partial autocorrelation function.

Figure 5. ARIMA (1: 0: 1)(0: 1: 1)(12) + CSI (lag = 0) model fitting: diagnosis: and prediction results.



A: Fitting and prediction results. Yellow and black are the fitted and true values of the training sets respectively. Red indicates the true values of the testing sets: and blue indicates the predicted values. The gray area represents the 95% confidence interval of the predicted value. **B:** Normal Q-Q plot. The solid black line represents the theoretical quantile line. **C:** Residual plot. **D:** ACF plot. **E:** PACF plot. The blue dashed lines in **D** and **E** indicate the confidence interval. ARIMA: autoregressive integrated moving average; CSI: comprehensive search index; ACF: autocorrelation function; PACF: partial autocorrelation function.

China's HFMD surveillance system, despite improvements, still experiences reporting delays and limited predictive capacity. The traditional ARIMA model can effectively analyze disease trends, but it is completely unable to predict the impact of external factors, such as ongoing epidemics (e.g. COVID-19) [27,28]; as 2020 predictions exceeded the 95% confidence intervals (Figure 4A). To address this, we constructed an ARIMAX model that incorporated BSI as an external variable. Nowadays, the internet is inseparably linked to our lives, and researchers increasingly use online data to predict diseases and health events [29–31]. Multiple internet-based monitoring systems have been developed to enhance the ability of infectious disease prediction. Google Flu Trends monitored and predicted influenza using internet search data. Although it initially performed well, later studies found that over-reliance on search engine data could introduce bias [32]. The method tended to overestimate case numbers and was sensitive to public search behavior fluctuations. As disease prediction models matured, researchers preferred combining traditional surveillance data with internet search data. This approach reduces the risk of network noise affecting predictions and improves robustness. Time series models using this combination have successfully predicted diseases such as acquired immunodeficiency syndrome (AIDS), influenza, and scarlet fever [33,34]. In the case of HFMD, several countries have used online data to assist in predicting outbreaks [35–37].

China had over 1.079 billion internet users by June 2023 [11]. Baidu, the leading search engine, maintained a 96.3% monthly active user rate [38]. BSI effectively reflects Chinese users' search trends and information interests, making it useful for disease prediction, as seen in prior studies [10,39]. In our study, we compiled a keyword database containing 11 terms closely linked to HFMD cases. Notably, "hand, foot and mouth disease in children" and "hand, foot and mouth" strongly correlated with incidence rates (rs = 0.928 and rs = 0.914, respectively). Similarly, the CSI correlated highly with incidence (rs = 0.932), indicating that BSI accurately tracked HFMD outbreaks. These findings reflect the growing reliance on search engines for information retrieval. Searches related to specific diseases mirror users' health needs and concerns, as evidenced by the increased search frequency for HFMD during outbreaks.

In general, ARIMA(1, 0, 1)(0, 1, 1)(12) and ARIMA(1, 0, 1)(0, 1, 1)(12) + CSI (lag = 0) exhibited satisfactory fitting. However, the ARIMAX model incorporating BSI achieved higher accuracy and predictive performance for HFMD incidence from 2020 to September 2023. The ARIMAX model's predictions closely matched the actual incidence curve and showed lower MAE and MAPE values. Although the predictive power of both models was affected by COVID-19 and policy changes in 2020 and 2023, the ARIMAX model outperformed the ARIMA model. Compared with ARIMA, MAPE decreased by 12.70% and 3.05%, respectively. The ARIMAX model captured public concern by incorporating BSI and produced more accurate HFMD forecasts. This higher fidelity can help policymakers identify risks sooner and respond with timely control measures. Similar results have been reported in other disease predictions. For example, when forecasting AIDS and brucellosis incidence in China, the ARIMAX model combined with BSI reduced MAPE by 1.46% and 0.95%, respectively, compared with the ARIMA model without BSI [25,33]. In our study, the MAPE decreased by 3.87%.

Our study has three main limitations. First, we selected keywords based on 2011–2019 data. This may reduce the generalizability of long-term predictions. Therefore, the vocabulary list should be updated regularly. Second, the model is not validated for subnational use; regional differences in internet access may have caused incomplete data collection in underserved areas. Future research should consider provincial sensitivity analysis or stratified modeling to improve resource allocation. Provincial case studies could also confirm effectiveness, provide spatial applicability, and serve as stress tests. Applying the proposed framework to multiple provinces would further verify its applicability in different epidemiological and policy environments. Third, relying on a single platform may restrict predictive power. Composite platform data could better support disease prediction. Future research should incorporate other platforms, such as Sogou Index or Sina Weibo, or combine multiple sources. This would allow testing the stability and generalizability of the HFMD prediction model.

Conclusions

Baidu data helps predict HFMD trends in China and alerts us to changes caused by unexpected events. It demonstrates clear value in public health.

Acknowledgements

The authors thank all the funders of this study and all researchers related to the article.

Funding
This work was supported by the Natural Science Foundation of Guangxi Zhuang Autonomous Region (2023GXNSFDA026036), National Natural Science Foundation of China (NSFC 82160385), One Thousand of Young and Middle-aged Key Teachers Training Program in Guangxi Colleges and Universities (DC2300017000).

Availability of data and materials
The data of the HFMD cases analyzed in this study were obtained from official website of the National Administration of Disease Prevention and Control (https://www.ndcpa.gov.cn/). The search engine data was obtained from Baidu Index (https://index.baidu.com/v2/index.html#/).

Authors' contributions
QH: data collection, primary data analysis, writing; HN, TL, JY: data collection, analysis, revisions; JZ, HM, ZL: conception and design; PC, HL: supervision; JH: supervision, final revisions and review. QH, HN, and TL contributed equally to this work and share first authorship.

Corresponding author
Jiegang Huang, MD.
School of Public Health, Guangxi Medical University, 22 Shuangyong Road,
Qingxiu District, Nanning, Guangxi 530021, China.
Tel: 86-07715334215
Email: jieganghuang@gxmu.edu.cn

Conflict of interest
No conflict of interest is declared.

References
1. Yang Z, Zhang Q, Cowling BJ, Lau EHY (2017) Estimating the incubation period of hand, foot and mouth disease for children in different age groups. Sci Rep 7: 16464. doi: 10.1038/s41598-017-16705-7.
2. Li F, Wang D, Su F, Zhao S, Guo J, Zhang F, Ran X, Wang J, Li S (2025) Epidemiological, etiological, and serological characteristics of hand, foot, and mouth disease in Guizhou Province, Southwest China, from 2008 to 2023. PLoS Negl Trop Dis 19: e0013394. doi: 10.1371/journal.pntd.0013394.
3. Chang LY, Lin TY, Hsu KH, Huang YC, Lin KL, Hsueh C, Shih SR, Ning HC, Hwang MS, Wang HS, Lee CY (1999) Clinical features and risk factors of pulmonary oedema after enterovirus-71-related hand, foot, and mouth disease. Lancet 354: 1682–1686. doi: 10.1016/S0140-6736(99)04434-7.
4. Hu H, Ge W, Yan J (2023) Analysis of the epidemiologic characteristics of children with hand, foot and mouth disease in China. Iran J Public Health 52: 773–779. doi: 10.18502/ijph.v52i4.12450.
5. Li J, Wang J, Xu C, Yin Q, Hu M, Sun Z, Shao D (2017) Hand, foot, and mouth disease in mainland China before it was listed as category C disease in May, 2008. Lancet Infect Dis 17: 1017–1018. doi: 10.1016/S1473-3099(17)30471-1.
6. Wu X, Zhou X, Chen Y, Zhai K, Sun R, Luo G, Lin Y-F, Li Y, Yang C, Zou H (2023) The impact of COVID-19 lockdown on cases of and deaths from AIDS, gonorrhea, syphilis, hepatitis B, and hepatitis C: interrupted time series analysis. JMIR Public Health Surveill 9: e40591. doi: 10.2196/40591.
7. Ding Z, Lu Q, Wu H, Wu C, Lin J, Wang X, Fu T, Yang K, Song Q (2024) Trend of hand, foot and mouth disease before, during, and after China's COVID control policies in Zhejiang, China. Front Public Health 12: 1472944. doi: 10.3389/fpubh.2024.1472944.
8. Ma W, Li X, Wang N, Wu J, Xiao Y, Hou S, Bi N, Gong L, Huang F (2024) Impact of non-pharmacological interventions on incidence of hand, foot and mouth disease during the COVID-19 pandemic: a large population-based observational study. BMC Infect Dis 24: 1353. doi: 10.1186/s12879-024-10252-z.
9. Alzahrani SI, Aljamaan IA, Al-Fakih EA (2020) Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. J Infect Public Health 13: 914–919. doi: 10.1016/j.jiph.2020.06.001.
10. Luo T, Zhou J, Yang J, Xie Y, Wei Y, Mai H, Lu D, Yang Y, Cui P, Ye L, Liang H, Huang J (2023) Early warning and prediction of scarlet fever in China using the Baidu search index and autoregressive integrated moving average with explanatory variable (ARIMAX) model: time series analysis. J Med Internet Res 25: e49400. doi: 10.2196/49400.
11. China Internet Network Information Center (2023) The China Internet Network Information Center has released the 52nd "Statistical Report on the Development of China's Internet". National Journal of Library Science 32: 13. [Article in Chinese].
12. Zhou K, Ding Z, Hu B, Zhan J, Cai K (2023) Circulating trends of hand, foot, and mouth disease in Hubei Province, China: impact from the COVID-19 pandemic. Heliyon 9: e22872. doi: 10.1016/j.heliyon.2023.e22872.
13. Chen B, Wang M, Huang X, Xie M, Pan L, Liu H, Liu Z, Zhou P (2021) Changes in incidence of notifiable infectious diseases in China under the prevention and control measures of COVID-19. Front Public Health 9: 728768. doi: 10.3389/fpubh.2021.728768.
14. National Administration of Disease Control and Prevention. Available: https://www.ndcpa.gov.cn/. Accessed 9 Apr 2025. [Article in Chinese].
15. Baidu search index. Available: https://index.baidu.com/v2/index.html#/. Accessed: 9 Apr 2025. [Article in Chinese].
16. ChinaZ.com. Available: https://data.chinaz.com/. Accessed 9 Apr 2025. [Article in Chinese].
17. China NHC of the PR (2018) Diagnosis and treatment guidelines for hand-foot-mouth disease (2018 Edition). Chinese Journal of Viral Diseases 8: 347–352. doi: 10.16505/j.2095-0136.2018.0063
18. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS (2013) Monitoring influenza epidemics in China with search query from Baidu. PLoS One 8: e64323. doi: 10.1371/journal.pone.0064323.
19. IBM Corp (2015). IBM SPSS Statistics for Windows (version 23.0) [Computer software]. IBM Corp.
20. Zheng Y, Wang K, Zhang L, Wang L (2021) Study on the relationship between the incidence of influenza and climate indicators and the prediction of influenza incidence. Environ Sci Pollut Res Int 28: 473-481. doi: 10.1007/s11356-020-10523-7.

21. Zhu P, Ji W, Li D, Li Z, Chen Y, Dai B, Han S, Chen S, Jin Y, Duan G (2023) Current status of hand-foot-and-mouth disease. J Biomed Sci 30: 15. doi: 10.1186/s12929-023-00908-4.
22. Chau NVV, Thuong TC, Hung NT, Hong NTT, Quy DT, Thien TB, Hiep CM, Minh NNQ, Khanh TH, Han DDK, Truc THC, Ny NTH, Thanh LK, Nguyet LA, Thuy CT, Nhu LNT, Van Quang P, Nguyen PNT, Qui PT, Rogier van Doorn H, Thwaites CL, Thanh TT, Dung NT, Thwaites G, Anh NT, Nhan LNT, Van Tan L, SEACOVARIANTS1 (2024) Emerging Enterovirus A71 subgenogroup B5 causing severe hand, foot, and mouth disease, Vietnam, 2023. Emerg Infect Dis 30: 363–367. doi: 10.3201/eid3002.231024.
23. Chen S, Chen Y, Ji W, Wang F, Zhang X, Jin Y, Liu Y (2023) Emerging concerns of atypical hand foot and mouth disease caused by recombinant Coxsackievirus A6 variants in Henan, China. J Med Virol 95: e29316. doi: 10.1002/jmv.29316.
24. Nie J, Huang T, Sun Y, Peng Z, Dong W, Chen J, Zheng D, Guo F, Shi W, Ling Y, Zhao W, Yang H, Shui T, Yan X (2024) Influence of the Enterovirus 71 vaccine and the COVID-19 pandemic on hand, foot, and mouth disease in China based on counterfactual models: observational study. JMIR Public Health Surveill 10: e63146. doi: 10.2196/63146.
25. Hu C-Y, Tang Y-W, Su Q-M, Lei Y, Cui W-S, Zhang Y-Y, Zhou Y, Li X-Y, Wang Z-F, Zhao Z-X (2021) Public health measures during the COVID-19 pandemic reduce the spread of other respiratory infectious diseases. Front Public Health 9: 771638. doi: 10.3389/fpubh.2021.771638.
26. Sun Y, Zhou J, Nie W, Tian D, Ye Q (2024) Study on the epidemiological characteristics of enterovirus among pediatric patients in Hangzhou, China: a comparison between the pre-COVID-19, COVID-19 pandemic, and post-COVID-19 periods. J Med Virol 96: e29412. doi: 10.1002/jmv.29412.
27. Yang J, Zhou J, Luo T, Xie Y, Wei Y, Mai H, Yang Y, Cui P, Ye L, Liang H, Huang J (2023) Predicting pulmonary tuberculosis incidence in China using Baidu search index: an ARIMAX model approach. Environ Health Prev Med 28: 68. doi: 10.1265/ehpm.23-00141.
28. Man OM, Fuller TL, Rosser JI, Nielsen-Saines K (2022) Re-emergence of arbovirus diseases in the State of Rio de Janeiro, Brazil: the role of simultaneous viral circulation between 2014 and 2019. One Health 15: 100427. doi: 10.1016/j.onehlt.2022.100427.
29. Yang Y, Tsao S-F, Basri MA, Chen HH, Butt ZA (2023) Digital disease surveillance for emerging infectious diseases: an early warning system using the internet and social media data for COVID-19 forecasting in Canada. Stud Health Technol Inform 302: 861–865. doi: 10.3233/SHTI230290.
30. Zhao C, Yang Y, Wu S, Wu W, Xue H, An K, Zhen Q (2020) Search trends and prediction of human brucellosis using Baidu index data from 2011 to 2018 in China. Sci Rep 10: 5896. doi: 10.1038/s41598-020-62517-7.
31. Desai R, Lopman BA, Shimshoni Y, Harris JP, Patel MM, Parashar UD (2012) Use of internet search data to monitor impact of rotavirus vaccination in the United States. Clin Infect Dis 54: e115–118. doi: 10.1093/cid/cis121.
32. Lazer D, Kennedy R, King G, Vespignani A (2014) Big data. The parable of Google Flu: traps in big data analysis. Science 343: 1203–1205. doi: 10.1126/science.1248506.
33. Li J, Yan X, Chu X, Zhang Y, Liu G, Li L, Li Y, Dong X, Mei Z, Liu Z, Yuan J, Sun X, Cao C (2025) A deep learning framework for using search engine data to predict influenza-like illness and distinguish epidemic and nonepidemic seasons: multifeature time series analysis. J Med Internet Res 27: e71786. doi: 10.2196/71786.
34. Samaras L, García-Barriocanal E, Sicilia M-A (2020) Comparing social media and Google to detect and predict severe epidemics. Sci Rep 10: 4747. doi: 10.1038/s41598-020-61686-9.
35. Niu Q, Liu J, Zhao Z, Onishi M, Kawaguchi A, Bandara A, Harada K, Aoyama T, Nagai-Tanima M (2022) Explanation of hand, foot, and mouth disease cases in Japan using Google Trends before and during the COVID-19: infodemiology study. BMC Infect Dis 22: 806. doi: 10.1186/s12879-022-07790-9.
36. Bednarska NG, Smith S, Bardsley M, Loveridge P, Byford R, Elson WH, Hughes HE, de Lusignan S, Todkill D, Elliot AJ (2025) Trends in general practitioner consultations for hand foot and mouth disease in England between 2017 and 2022. Epidemiol Infect 153: e22. doi: 10.1017/S095026882400181X.
37. Santangelo OE, Gianfredi V, Provenzano S, Cedrone F (2023) Digital epidemiology and infodemiology of hand-foot-mouth disease (HFMD) in Italy. Disease trend assessment via Google and Wikipedia. Acta Biomed 94: e2023107. doi: 10.23750/abm.v94i4.14184.
38. China Search Engine Industry (2023) China search engine industry report. https://www.ctrchina.cn/rich/report/558. Accessed 9 Apr 2025. [Article in Chinese].
39. Liu K, Wang T, Yang Z, Huang X, Milinovich GJ, Lu Y, Jing Q, Xia Y, Zhao Z, Yang Y, Tong S, Hu W, Lu J (2016) Using Baidu search index to predict dengue outbreak in China. Sci Rep 6: 38040. doi: 10.1038/srep38040.

Annex – Suplementary Items

Supplementary Table 1. Correlation between monthly Baidu search index keywords and reported cases of hand, foot and mouth disease from January 2011 to 2019.

Categories	Keywords[a]	r_s[b]	*p* value	Search volume mean (SD)
手足口病综合类 (Hand, foot and mouth disease comprehensive category)	手足 (hands and feet)	0.364	< 0.001	328.37 (51.18)
	手足口 (hand, foot and mouth)	0.914	< 0.001	1661.10 (1003.40)
	手足口症 (hand-foot-mouth syndrome)	0.413	< 0.001	69.02 (118.80)
	小儿手足口病 (hand, foot and mouth disease in children)	0.803	< 0.001	259.38 (131.95)
	儿童手足口病 (hand, foot and mouth disease in children)	0.928	< 0.001	195.01 (70.93)
	什么是手足口病 (What is hand, foot and mouth disease?)	0.788	< 0.001	284.90 (152.61)
手足口病病因类 (Hand, foot and mouth disease etiology category)	埃可病毒 (ECHO virus)	0.360	< 0.001	113.52 (107.99)
	肠道病毒 (enterovirus)	0.516	< 0.001	142.77 (33.63)
	轮状病毒 (Rotavirus)	-0.296	0.002	1608.95 (1156.42)
	诺如病毒 (Norovirus)	-0.158	0.102	1757.66 (2450.38)
	手足口病的传播途径 (transmission of hand, foot and mouth disease)	0.665		123.74 (56.23)
手足口病症状类 (Hand, foot and mouth disease symptoms category)	斑疹 (macules)	0.312	< 0.001	324.69 (324.39)
	斑丘疹 (maculopapular rash)	0.442	< 0.001	272.48 (85.66)
	斑疹 (macules)	0.312	0.001	324.69 (324.39)
	传染性红斑 (erythema infectiosum)	0.045	0.647	87.16 (22.65)
	带状疱疹 (herpes zoster)	0.265	0.006	6118.08 (2698.96)
	单纯疱疹 (herpes simplex)	0.192	0.047	777.86 (233.67)
	多动症 (attention deficit hyperactivity disorder)	0.437	< 0.001	1234.10 (209.98)
	鹅口疮 (thrush)	0.550	< 0.001	1892.08 (534.29)
	发烧后出疹子 (fever followed by a rash)	0.480	< 0.001	169.83 (81.92)
	风疹 (rubella)	0.623	< 0.001	1688.30 (519.55)
	汗疱疹 (sweat herpes)	0.677	< 0.001	1582.74 (1155.36)
	花柳病 (venereal disease)	0.078	0.425	748.25 (420.94)
	口疮 (aphtha)	0.411	< 0.001	582.44 (70.83)
	口蹄疫 (foot-and-mouth disease)	-0.291	0.002	512.33 (133.69)
	流行性感冒 (influenza)	-0.346	< 0.001	357.16 (160.35)
	麻疹 (measles)	0.360	< 0.001	1864.79 (897.47)
	麻疹的症状 (symptoms of measles)	0.047	0.629	367.44 (263.34)
	麻疹的症状和治疗 (symptoms and treatment of measles)	0.272	0.004	1032.82 (687.81)
	麻疹图片 (measles pictures)	0.307	0.001	669.85 (297.66)
	麻疹症状 (measles symptoms)	0.129	0.183	179.12 (70.29)
	泡疹 (herpes)	0.699	< 0.001	556.76 (147.27)
	疱疹 (herpes)	0.412	< 0.001	3898.24 (1286.24)
	疱疹性口腔炎 (herpetic stomatitis)	0.560	< 0.001	314.13 (93.38)
	疱疹性咽峡炎 (herpangina)	0.474	< 0.001	2590.76 (3449.55)
	皮肤红斑 (erythema of the skin)	0.473	< 0.001	216.07 (41.27)
	皮疹 (rash)	0.454	< 0.001	1229.19 (289.22)
	丘疹 (papule)	0.410	< 0.001	1006.92 (256.93)
	丘疹性荨麻疹 (papular urticaria)	0.647	< 0.001	1522.94 (902.52)
	什么是脚气病 (What is beriberi?)	0.323	0.001	90.35 (41.40)
	神经性皮炎 (neurodermatitis)	0.242	0.012	2509.95 (779.17)
	湿疹 (eczema)	0.570	< 0.001	8727.80 (3048.01)
	湿疹图片 (eczema pictures)	0.249	0.009	4411.99 (3490.28)
	手口足病的症状 (symptoms of hand, foot and mouth disease)	0.614	< 0.001	368.63 (240.22)
	小孩手足口病症状 (hand, foot and mouth disease symptoms in children)	0.802	< 0.001	457.14 (254.84)

	手癣 (tinea manis)	0.653	< 0.001	698.10 (187.69)
	手足口并发症 (hand, foot and mouth complications)	0.858	< 0.001	33.14 (28.76)
	手足口病并发症 (complications of hand, foot and mouth disease)	0.669	< 0.001	52.39 (40.30)
	手足口病初期症状 (initial symptoms of hand, foot and mouth disease)	0.483	< 0.001	3884.76 (4912.08)
	手足口病初期症状图 (ddiagram of the initial symptoms of hand, foot and mouth disease)	0.444	< 0.001	277.92 (390.20)
	手足口病的症状 (symptoms of hand, foot and mouth disease)	0.767	< 0.001	533.51 (463.11)
	手足口病潜伏期 (hand, foot and mouth disease incubation period)	0.592	< 0.001	241.03 (234.07)
	手足口病特征 (features of hand, foot and mouth disease)	0.590	< 0.001	81.54 (45.63)
	手足口病图片 (hand, foot and mouth disease pictures)	0.726	< 0.001	1944.28 (1846.56)
	手足口病严重吗 (Is hand, foot and mouth disease serious?)	0.572	< 0.001	508.65 (406.66)
	手足口病有什么症状 (What are the symptoms of hand, foot and mouth disease?)	0.613	< 0.001	173.43 (121.93)
	手足口病症状 (hand, foot and mouth disease symptoms)	0.464	< 0.001	3959.13 (4000.09)
	手足口病症状图片 (hand, foot and mouth disease symptom pictures)	0.395	< 0.001	1749.50 (2135.69)
	手足癣 (tinea of the hands and feet)	0.588	< 0.001	385.46 (110.65)
	水痘 (chickenpox)	0.231	0.016	3736.61 (1773.05)
	水痘的症状和图片 (symptoms and pictures of chickenpox)	0.120	0.217	371.95 (323.64)
	水痘的症状和治疗 (symptoms and treatment of chickenpox)	0.126	0.193	2357.11 (1624.89)
	水痘图片 (chickenpox pictures)	0.239	0.013	1501.65 (1112.59)
	水泡 (blister)	0.656	< 0.001	727.72 (223.32)
	水疱型脚气 (blistering athlete's foot)	0.406	< 0.001	242.02 (186.39)
	水疱型手足癣 (blistering tinea manis and foot)	0.382	< 0.001	160.39 (235.07)
	水疱疹 (blistering rash)	0.307	0.001	230.35 (204.66)
	天花 (smallpox)	0.252	0.009	1520.47 (387.34)
	猩红热 (scarlet fever)	0.351	< 0.001	1500.04 (679.82)
	婴儿鹅口疮 (thrush in infants)	0.181	0.061	262.44 (53.17)
	幼儿急疹 (rash in young children)	0.348	< 0.001	2909.61 (1567.95)
	珍珠疹 (pearly rash)	0.357	< 0.001	2493.05 (3865.72)
	珍珠疹图片 (pearly rash pictures)	0.346	< 0.001	467.84 (187.60)
	症状 (symptom)	0.240	0.012	356.45 (97.70)
	紫癜 (purpura)	0.289	0.002	2013.37 (850.30)
手足口病防治类 (Hand, foot and mouth disease prevention and treatment category)	免费咨询 (free consultation)	0.107	0.269	143.91 (103.17)
	检查 (examine)	0.091	0.351	1390.71 (407.28)
	预防手足口病 (prevention of hand, foot and mouth disease)	0.699	< 0.001	155.00 (74.38)
	如何预防手足口病 (How to prevent hand, foot and mouth disease?)	0.733	< 0.001	188.40 (106.26)
	手足口病预防 (hand, foot and mouth disease prevention)	0.797	< 0.001	181.05 (93.94)
	手足口病的预防 (prevention of hand, foot and mouth disease)	0.484	< 0.001	198.92 (170.73)
	手足口病的预防知识 (prevention of hand, foot and mouth disease)	0.501	< 0.001	28.03 (33.47)
	手足口病防治指南 (guidelines for the prevention and treatment of hand, foot and mouth disease)	0.602	< 0.001	70.19 (63.64)
	手足口病防治知识 (knowledge of hand, foot and mouth disease prevention and treatment)	0.231	0.016	315.38 (488.74)
	手足口病用药 (hand, foot and mouth disease medications)	0.842	< 0.001	87.97 (42.12)
	手足口病治疗 (hand, foot and mouth disease treatment)	0.595	< 0.001	410.97 (553.18)
	手足口病的治疗 (treatment of hand, foot and mouth disease)	0.661	< 0.001	188.13 (111.38)
	手足口病怎么治疗 (How to treat hand, foot and mouth disease)	0.543	< 0.001	1324.68 (1299.23)

	手足口病吃什么药 (What medicine to take for hand, foot and mouth disease?)	0.812	< 0.001	690.44 (449.40)
	水痘的治疗方法 (treatment for chickenpox)	0.053	0.583	156.07 (114.47)
	炉甘石洗剂 (calamine lotion)	0.417	< 0.001	2183.15 (1172.27)

[a] Categories and keywords were presented in a Chinese (English) format; [b]r_s: Spearman rank correlation.

Supplementary Table 2. Parameter estimation of candidate models.

Model	Variables	Estimate	Standard Error	Z value	Pr (>|z|)
ARIMA[a](1, 0, 1)(1, 1, 0)$_{(12)}$					
	ar1[d]	0.414	0.142	2.913	0.004
	ma1[e]	0.333	0.150	2.216	0.027
	sma1[f]	- 0.711	0.068	- 10.441	< 0.001
ARIMAX[b](1, 1, 2)(2, 0, 0)$_{(12)}$					
	ar1	0.362	0.169	2.141	0.032
	ma1	- 0.660	0.166	- 3.980	< 0.001
	ma2	- 0.302	0.145	- 2.084	0.037
	sar1[g]	0.211	0.118	1.786	0.074
	sar2	0.199	0.118	0.169	0.091
	CSI (lag = 0)	251.563	22.013	11.428	< 0.001
ARIMA(1, 0, 1)(0, 1, 1)$_{(12)}$					
	ar1	0.513	0.108	4.731	< 0.001
	ma1	0.423	0.108	3.911	< 0.001
	sma1	- 0.999	0.173	- 5.764	< 0.001
ARIMA(1, 0, 1)(0, 1, 1)$_{(12)}$ + CSI[c] (lag = 0)					
	ar1	0.466	0.137	3.409	< 0.001
	ma1	0.243	0.143	1.707	0.088
	sma1	- 0.916	0.321	- 2.855	0.004
	CSI (lag = 0)	199.045	18.685	10.652	< 0.001

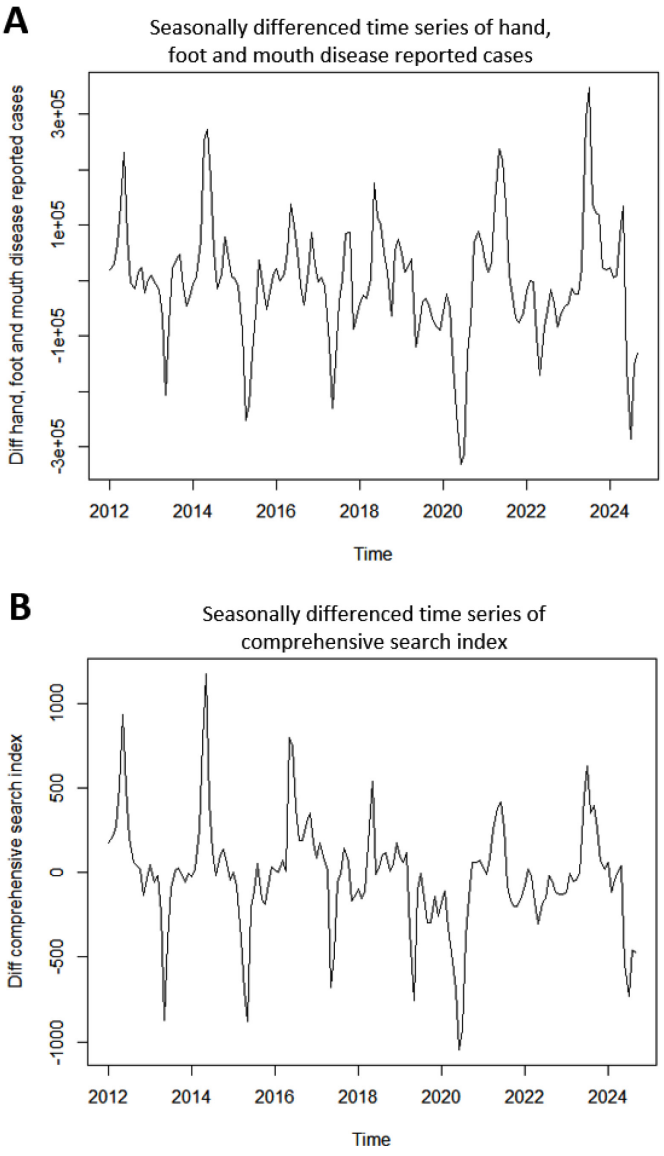
[a]ARIMA: autoregressive integrated moving average; [b]ARIMAX: autoregressive integrated moving average with explanatory variable; [c]CSI: comprehensive search index; [d]ar1: autoregressive coefficient of order 1; [e]ma1: moving average coefficient of order 1; [f]sma1: seasonal moving Average coefficient of order 1; [g]sar1: seasonal autoregressive coefficient of order 1.

Supplementary Table 3. Evaluation of the model fit, diagnosis, and prediction.

	Fit		Diagnosis			Prediction		
	AIC[a]	R^2	LSM[b] *p* value	Ljung-Box tes0 (white noise test)		MAE[c]	RMSE[d]	MAPE[e], %
				Chi square	*p* value			
ARIMA[f] (1, 0, 1)(0, 1, 1)$_{(12)}$ 2019.1–2024.9	2333.52	0.91	< 0.001	0.059	0.81	95858.53	127156.58	5.92
2020.1–2020.12						123724.94	176492.77	20.14
2021.1–2021.12						64734.39	78816.12	0.61
2022.1–2022.12						120359.79	143620.85	2.08
2023.1–2023.12						80483.09	108053.44	4.02
2024.1–2024.9						88034.43	97554.14	1.71
ARIMA(1, 0, 1)(0, 1, 1)$_{(12)}$+ CSI[g] (lag = 0) 2019.1–2024.9	2263.41	0.95	0.025	0.003	0.96	38300.36	54159.18	2.05
2020.1–2020.12						54787.96	74202.85	7.44
2021.1–2021.12						17190.49	20836.75	0.18
2022.1–2022.12						39879.93	46794.07	0.70
2023.1–2023.12						44252.73	66187.20	0.97
2024.1–2024.9						34420.78	43544.87	0.63

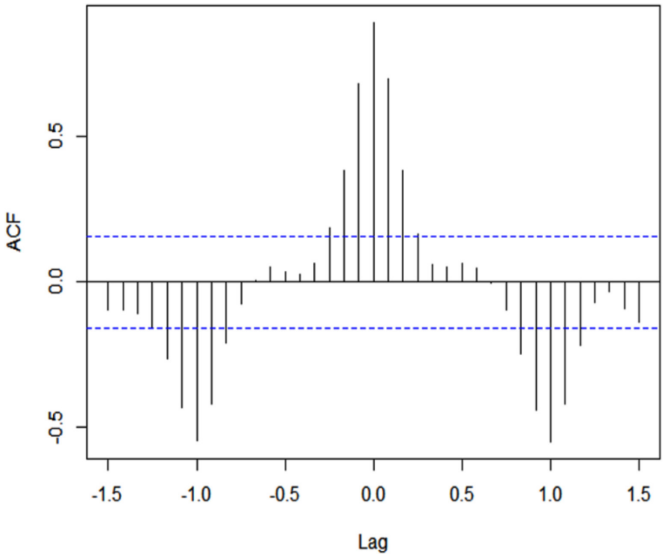
[a]AIC: Akaike information criterion; [b]LSM: least squares method; [c]MAE: mean absolute error; [d]RMSE: root mean square error; [e]MAPE: mean absolute percentage error; [f]ARIMA: autoregressive integrated moving average; [g]CSI: comprehensive search index.

Supplementary Figure 1. Seasonally differenced time series of hand: foot and mouth disease reported cases and comprehensive search index.



Diff: Differencing.

Supplementary Figure 2. The CCF between hand: foot and mouth disease and the comprehensive search index. Blue dashed lines indicate the 95% CI.



ACF: autocorrelation function; CCF: cross-correlation function.