

Original Paper

# Forecasting the Incidence of Mumps Based on the Baidu Index and Environmental Data in Yunnan, China: Deep Learning Model Study

Xin Xiong<sup>1\*</sup>, BS; Linghui Xiang<sup>1\*</sup>, BS; Litao Chang<sup>2</sup>, BS; Irene XY Wu<sup>1,3</sup>, PhD; Shuzhen Deng<sup>2</sup>, MS

<sup>1</sup>Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China

<sup>2</sup>Department of School Health, Yunnan Provincial Center for Disease Control and Prevention, Kunming, China

<sup>3</sup>Hunan Provincial Key Laboratory of Clinical Epidemiology, Central South University, Changsha, China

\*these authors contributed equally

**Corresponding Author:**

Shuzhen Deng, MS

Department of School Health

Yunnan Provincial Center for Disease Control and Prevention

No.1177 Xianghe Street, Luolong Street, Chenggong District

Kunming, 650500

China

Phone: 86 15096624164

Email: [461447164@qq.com](mailto:461447164@qq.com)

## Abstract

**Background:** Mumps is a viral respiratory disease characterized by facial swelling and transmitted through respiratory secretions. Despite the availability of an effective vaccine, mumps outbreaks have reemerged globally, including in China, where it remains a significant public health issue. In Yunnan province, China, the incidence of mumps has fluctuated markedly and is higher than that in mainland China, underscoring the need for improved outbreak prediction methods. Traditional surveillance methods, however, may not be sufficient for timely and accurate outbreak prediction.

**Objective:** Our study aims to leverage the Baidu search index, representing search volumes from China's most popular search engine, along with environmental data to develop a predictive model for mumps incidence in Yunnan province.

**Methods:** We analyzed mumps incidence in Yunnan Province from 2014 to 2023, and used time series data, including mumps incidence, Baidu search index, and environmental factors, from 2016 to 2023, to develop predictive models based on long short-term memory networks. Feature selection was conducted using Pearson correlation analysis, and lag correlations were explored through a distributed nonlinear lag model (DNLM). We constructed four models with different combinations of predictors: (1) model BE, combining the Baidu index and environmental factors data; (2) model IB, combining mumps incidence and Baidu index data; (3) model IE, combining mumps incidence and environmental factors; and (4) model IBE, integrating all 3 data sources.

**Results:** The incidence of mumps in Yunnan showed significant variability, peaking at 37.5 per 100,000 population in 2019. From 2014 to 2023, the proportion of female patients ranged from 41.3% in 2015 to 45.7% in 2020, consistently lower than that of male patients. After excluding variables with a Pearson correlation coefficient of  $<0.10$  or  $P$  values of  $<0.05$ , we included 3 Baidu index search term groups (disease name, symptoms, and treatment) and 6 environmental factors (maximum temperature, minimum temperature, sulfur dioxide, carbon monoxide, particulate matter with a diameter of 2.5  $\mu\text{m}$  or less, and particulate matter with a diameter of 10  $\mu\text{m}$  or less) for model development. DNLM analysis revealed that the relative risks consistently increased with rising Baidu index values, while nonlinear associations between temperature and mumps incidence were observed. Among the 4 models, model IBE exhibited the best performance, achieving the coefficient of determination of 0.72, with mean absolute error, mean absolute percentage error, and root-mean-square error values of 0.33, 15.9%, and 0.43, respectively, in the test set.

**Conclusions:** Our study developed model IBE to predict the incidence of mumps in Yunnan province, offering a potential tool for early detection of mumps outbreaks. The performance of model IBE underscores the potential of integrating search engine data and environmental factors to enhance mumps incidence forecasting. This approach offers a promising tool for improving public health surveillance and enabling rapid responses to mumps outbreaks.

**KEYWORDS**

mumps; deep learning; baidu index; forecasting; incidence prediction; time series analysis; Yunnan; China

## Introduction

Mumps is an acute respiratory infectious disease caused by the mumps virus, characterized by unilateral or bilateral facial swelling [1]. It is transmitted through contact with the respiratory secretions of infected individuals, with an incubation period typically ranging from 16 to 18 days [2]. Nonspecific symptoms such as fever, headache, and malaise could occur before the inflammation of the parotid and other salivary glands [1,2]. While mild symptoms are more common, severe complications such as aseptic meningitis, encephalitis, sensorineural deafness, and orchitis can also develop [3,4]. Out of 2 doses of the mumps vaccine have shown nearly 90% effectiveness in protecting against mumps [5], reducing the global incidence of mumps by 97% [6]. Until 2019, a total of 122 countries have introduced mumps vaccination [7], and the number reached to 124 Member States by the end of 2023 [8]. However, in recent years, mumps outbreaks have reemerged in many countries, even those with high vaccine coverage, such as the United States [9], the United Kingdom [10], and Ireland [11], which may have many potential reasons, but the waning immunity may contribute majorly [12].

In China, mumps is classified as a Class C notifiable infectious disease. Despite being vaccine-preventable, mumps remained the most prevalent disease among the 44 notifiable infectious diseases from 2008 to 2017 in persons aged 6 to 22 years, with incidences of 83.207 per 100,000 population in 2008 and 69.771 per 100,000 population in 2017 [13]. Similar to the global condition, even after mumps was included in the expanded immunization program in mainland China in 2008, an abnormal increase in incidence was observed after the program [13,14]. Although no exact reason, waning immunity, changing environment, high population mobility, along with misinformation about the vaccination may play roles [7,15,16]. After the onset of the COVID-19 pandemic, the incidence of mumps in mainland China was 7.37 and 6.48 per 100,000 population in 2022 and 2023, ranking fourth and fifth among Class C notifiable infectious diseases, respectively [17,18]. These trends highlight the continued public health challenge posed by mumps in China.

Yunnan, a province located in southwestern China, has experienced marked variability in mumps incidence. The average annual percentage change in incidence was 13.97% from 2004 to 2018, ranking third, just slightly lower than that of Hunan (15.15%) and Hainan (15.10%) provinces [19]. Furthermore, the mumps incidence in 2022 and 2023 was 14.2 and 12.8 per 100,000 population, respectively, both of which were notably higher than the national averages. These sharp fluctuations in the incidence of mumps, along with relatively higher incidence presented challenges in predicting mumps outbreaks, particularly given the lag of traditional surveillance system, which rely on case reports from medical and health institutions.

Several studies have developed prediction models for the incidence of mumps in China [20,21]. However, due to regional variations of mumps incidence, these models might not be suitable for Yunnan province. In addition, these studies only used mumps incidence, environmental factors data or both as predictors, which may not be able to explain the nonseasonal fluctuations. Incorporating alternative predictors, such as search engine query data, may improve the timeliness and accuracy of outbreak predictions. Previous studies have identified search engine data, such as that from Twitter [rename as X], Google, and Baidu, along with environmental factors, as effective predictors of certain infectious diseases such as influenza and chicken pox, due to their lower cost and higher timeliness [22-24]. Baidu, the most popular search engine in China, provides a rich source of real-time data that can be leveraged for predictive purposes. Long short-term memory (LSTM) networks, introduced by Hochreiter and Schmidhuber [25] in 1997, are a specialized type of recurrent neural network designed to capture long-term dependencies in sequential data by mitigating the vanishing and exploding gradient problems [25]. These networks have shown exceptional performance in capturing temporal dependencies in larger horizons [26]. Therefore, our study aims to use multisource data to develop a prediction model of mumps incidence in Yunnan province based on LSTM.

## Methods

### Data Sources

Our study used a total of 3 time series data about Yunnan province, including the incidence of mumps, the Baidu index [27], and environmental factors.

We obtained the daily number of mumps cases with their age, sex, and careers, and the annual population data from the Yunnan Center for Disease Control and Prevention, which collected accurate case characteristics from the Chinese information system for infectious diseases control and prevention, for the period between January 1, 2014, and December 31, 2023. From the system, we collected a total of 20 career types in our study, reporting the number of patients of the top 6 most common ones while grouping the remaining careers under the category “other careers.” In addition, since more than 80% of patients were younger than 20 years, we organized the age distribution into 5-year groups for those younger than 20 years, while others were presented as a separate category. The Baidu index [27] is the weighted sum of the search frequencies of various search terms by scientifically analyzing and calculating based on users’ search volumes on Baidu, which is the most popular search engine in China. By extensive search, we initially collected daily Baidu index for 30 search terms in Yunnan from January 1, 2016, and December 31, 2023. After group discussion, we excluded 3 search terms due to irrelevance and another 3 because they were introduced after 2016, resulting

in incomplete data (Table S1 in [Multimedia Appendix 1](#)). Ultimately, we included 24 search terms in our study.

We collected information about environmental factors during January 1, 2016, and December 31, 2023, from China Meteorological Administration [28], including maximum temperature (°C), minimum temperature (°C), air quality index, particulate matter (PM) with an aerodynamic diameter less than or equal to 2.5  $\mu\text{m}$  (PM<sub>2.5</sub>; measured as  $\mu\text{g}/\text{m}^3$ ), PM with an aerodynamic diameter less than or equal to 10  $\mu\text{m}$  (PM<sub>10</sub>; measured as  $\mu\text{g}/\text{m}^3$ ), sulfur dioxide (SO<sub>2</sub>; measured as  $\mu\text{g}/\text{m}^3$ ), nitrogen dioxide (NO<sub>2</sub>; measured as  $\mu\text{g}/\text{m}^3$ ), carbon monoxide (CO; measured as  $\text{mg}/\text{m}^3$ ), and ozone (O<sub>3</sub>; measured as  $\mu\text{g}/\text{m}^3$ ).

## Statistical Analysis

### Feature Selection and Descriptive Analysis

To predict the weekly incidence of mumps, we calculated the weekly values for each variable. The weekly mumps incidence and Baidu index [27] were obtained by summing their daily values, while the weekly environmental factors were calculated as the mean values for each week. We performed Pearson correlation tests to determine the correlation between each variable and the incidence of mumps. After excluding variables with a Pearson correlation coefficient (PCC) of <0.10 or a *P* value of <.05, we consequently included 16 search terms and 6 environmental factors (Table S2 in [Multimedia Appendix 1](#)). These search terms were classified into 3 groups based on their content: disease name, symptoms, and treatment. We then calculated the weekly sum of the Baidu index for these 3 groups as the predictors.

Our study used the Kolmogorov-Smirnov test to inspect the normality of continuous variables. Normally distributed continuous variables were described as mean and SD, while nonnormally distributed continuous variables are described as median and IQR.

### Log Correlation Analysis

Previous studies have demonstrated that environmental factors and the Baidu index [27] had a lag impact on the incidence of infectious disease, which means exposure affects health for a period of days [22,29]. To capture these complex relationships, we used the distributed lag nonlinear model (DLNM). DLNM is a modeling framework that simultaneously represents nonlinear exposure-response relationships and lagged effects using cross-basis functions [30]. We used natural spline functions for both the exposure-response and lag-response components, allowing for flexible modeling of nonlinear associations. We selected DLNM over other nonlinear methods, such as generalized additive models, because DLNM is specifically tailored to handle both nonlinear and lagged effects, which is crucial for understanding the temporal dynamics of environmental influences on mumps incidence. Our study used R package DLNM (version 2.4.7) to make DLNM between the

Baidu index [27] and environmental factors and mumps incidence [31].

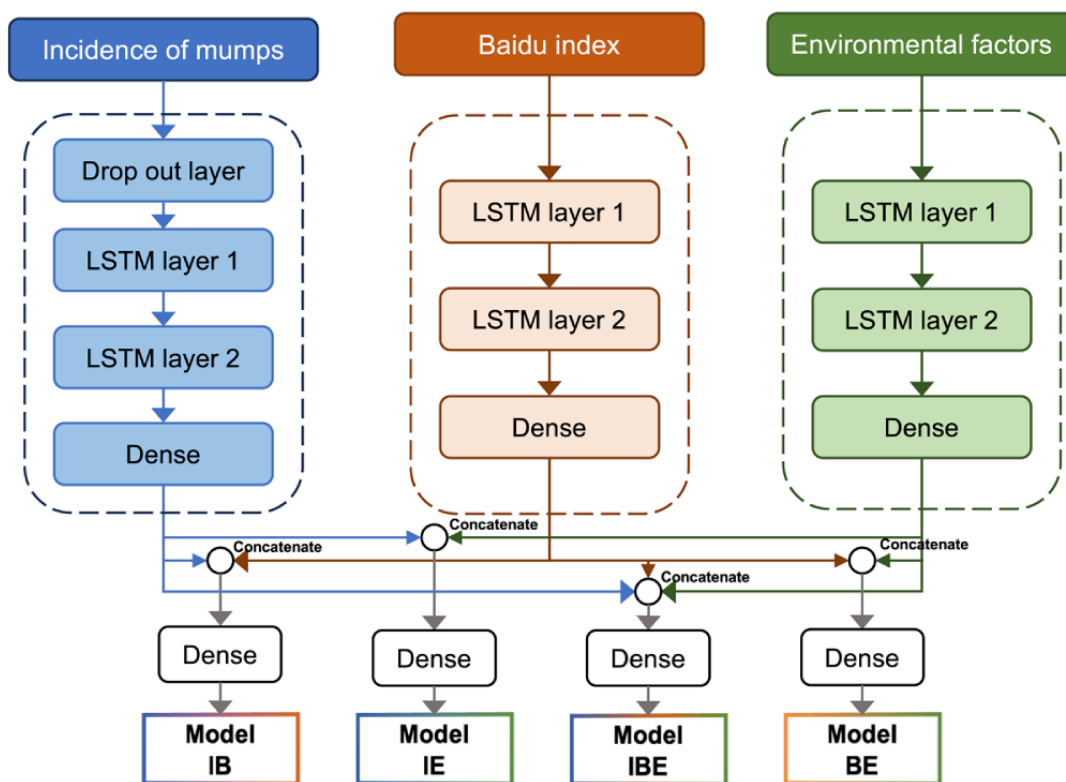
$$\log(E[Y_t]) = \left[ \alpha \right] + \sum_{i=1}^p \beta_i \text{cb}(X_{t-i}, \text{argvar}, \text{arglag})$$

$Y_t$  represents the outcome value (mumps incidence) at week  $t$  while  $E[Y_t]$  is its expected value.  $\alpha$  refers to the interception term.  $p$  refers to the maximum lag time, which was 3 weeks in our study considering the incubation and previous study [32].  $\beta_i$  the regression coefficient for the  $i$ -th lag time, representing the effect size of the exposure variable at lag  $i$  on the outcome variable.  $\text{cb}(X_{t-i}, \text{argvar}, \text{arglag})$  is a cross-basis function used to capture the nonlinear relationship between the exposure variable  $X_{t-i}$  and the lag time. This function depends on 3 parameters:  $X_{t-i}$  represents the value of the exposure variable at time  $t-i$ ,  $\text{argvar}$  were parameters defining the effect of the exposure variable, and  $\text{arglag}$  were parameters defining the lag effect. Relative risks (RR) were used to assess the effect of a specific exposure level at one lag time compared with the reference exposure value.

### Model Construction and Performance Evaluation

Our study developed 4 models based on LSTM to predict mumps incidence 1 week in advance ([Figure 1](#)). First, to better capture the characteristics of the 3 time series data, we constructed 3 primary LSTM models for each data set. The primary model for mumps incidence included a dropout layer to prevent overfitting, followed by 2 LSTM layers and a dense layer. The Baidu index [27] and environmental factors models each consisted of 2 LSTM layers followed by a dense layer. Among the 3 models, we used data from the previous 2 weeks to predict the incidence of mumps one week in advance. Subsequently, we concatenated the outputs of these primary models to construct composite models. Model IB combined the mumps incidence and Baidu index outputs, model IE combined the mumps incidence and environmental factors outputs, model BE combined the Baidu index [27] and environmental factors outputs, and model IBE combined the outputs from all 3 primary models. Each composite model then passed the concatenated outputs through an additional dense layer to produce the final prediction. This approach ensured that primary models effectively integrated information from different data sources for accurate predictions. To simulate real-world scenarios where future data are unknown and must be predicted based on past observations, we divided our data sets into train, validation and test set based on time. The train set interval was from January 1, 2016, to December 31, 2021. Data from January 1, 2022, to December 31, 2022, and those from January 1, 2023, to December 31, 2023, were used as the validation and test sets, respectively. These models were constructed by keras3 package [ONEIROS] (version 1.0.0), tensorflow package [Google Brain Team] (version 2.16.0) in R (version 4.4.1; R Core Team).

**Figure 1.** Model constructions. LSTM: Long short-term memory. Model BE: model including Baidu index and environmental factors; Model IB: model including mumps incidence and Baidu index; Model IE: model including mumps incidence and environmental factors; Model IBE: model including mumps incidence and Baidu index and environmental factors.



Out of 4 performance metrics were applied to evaluate these 4 models: coefficient of determination ( $R^2$ ), mean absolute error (MAE), mean absolute percentage error (MAPE), and root-mean-square error (RMSE).  $R^2$  refers to the proportion of the variance for a dependent variable that is explained by the model, with values less than 1, where larger values indicate better performance. MAE, MAPE, and RMSE measure the errors between predictions and actual values, with MAE and RMSE ranging greater than 0, and MAPE ranging from 0 to 100%, where lower values indicate better model accuracy. Compared with the MAE, MAPE represents the average of absolute percentage errors expressed as a percentage, making it more suitable for comparison across different data sets and models, while RMSE is more sensitive to outliers and larger errors.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - [\hat{y}_i])^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100\%$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - [\hat{y}_i])^2}$$

In these equations, n was the number of observations, while  $y_i$  and  $\hat{y}_i$  refer to the actual incidence of mumps in the  $i_{th}$  week and its prediction, respectively.

### Ethical Considerations

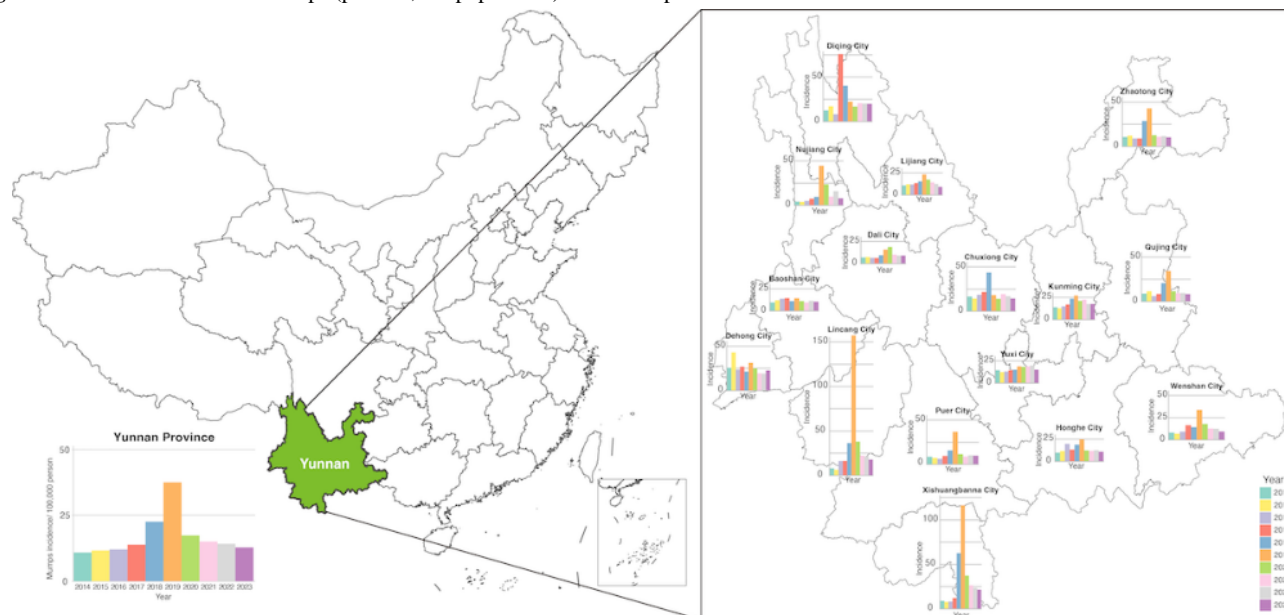
Our research was approved by the Medical Ethics Committee, Xiangya School of Public Health, Central South University, Changsha, China (XYGW-2024-65). Informed consent was waived by our institutional review board because all data were fully anonymized before being accessed by the researchers, ensuring that no identifiable information about individuals was available. Therefore, compensation is not applicable in our study.

### Results

#### The Temporal Distribution of Mumps Incidence and Predictors

In Yunnan Province, the incidence of mumps increased from 10.9 per 100,000 population in 2014, peaked at 37.5 per 100,000 population in 2019, and then steadily decreased to 12.8 per 100,000 population by 2023 (Figure 2). The outbreaks of mumps occurred from 2017 to 2020 in different cities in Yunnan. From 2014 to 2023, the proportion of female patients remained fairly stable, ranging from 41.3% in 2015 to 45.7% in 2020. (Table 1). The median age of patients showed some fluctuations over the years, starting at 7 years (IQR 4-12) in 2014, peaking at 11 years (IQR 7-14) in 2019, and then decreasing to 6 years (IQR 4-10) by 2023. The proportion of cases in children between 5 and 10 years old remained substantial, with the highest proportion recorded in 2023 (43.7%). Students consistently represented the largest affected group, comprising 44.3% to 67.5% of cases in 2014 and 2019, respectively.

**Figure 2.** Annual incidence of mumps (per 100,000 population) in Yunnan province from 2014 to 2023.



**Table 1.** Characteristics of reported mumps patients in Yunnan from 2014 to 2023.

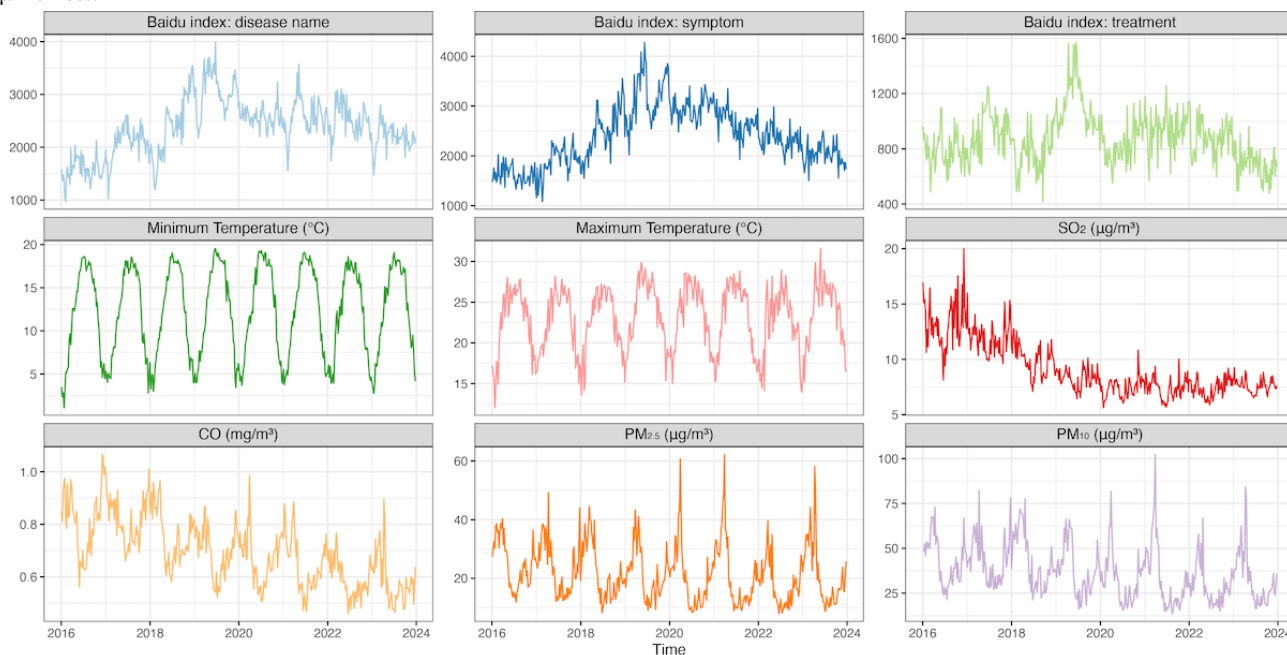
	2014 (N=5343)	2015 (N=5611)	2016 (N=5850)	2017 (N=6843)	2018 (N=11,249)	2019 (N=18,703)	2020 (N=8592)	2021 (N=7296)	2022 (N=6788)	2023 (N=6085)
Female	2290 (42.9)	2320 (41.3)	2484 (42.5)	2927 (42.8)	4961 (44.1)	8214 (43.9)	3929 (45.7)	3250 (44.5)	2929 (43.1)	2599 (42.7)
Age (years), median (IQR)	7 (4-12)	8 (5-11)	8 (5-12)	9 (5-13)	10 (6-14)	11 (7-14)	8 (5-14)	6 (4-11)	6 (4-11)	6 (4-10)
<b>Age groups (years), n (%)</b>										
<5	1389 (26)	1387 (24.7)	1456 (24.9)	1329 (19.4)	1679 (14.9)	2266 (12.1)	2051 (23.9)	2183 (29.9)	1940 (28.6)	1700 (27.9)
5 to <10	2082 (39)	2224 (39.6)	2300 (39.3)	2484 (36.3)	3492 (31)	5236 (28.0)	2801 (32.6)	2828 (38.8)	2873 (42.3)	2661 (43.7)
10 to <15	1007 (18.8)	1105 (19.7)	1074 (18.4)	1807 (26.4)	3786 (33.7)	7019 (37.5)	1618 (18.8)	944 (12.9)	850 (12.5)	715 (11.8)
15 to <20	162 (3)	200 (3.6)	210 (3.6)	308 (4.5)	1098 (9.8)	2272 (12.1)	679 (7.9)	289 (4.0)	201 (3.0)	152 (2.5)
≥20	703 (13.2)	695 (12.4)	810 (13.8)	915 (13.4)	1194 (10.6)	1910 (10.2)	1443 (16.8)	1052 (14.4)	924 (13.6)	857 (14.1)
<b>Careers, n (%)</b>										
Student	2368 (44.3)	2608 (46.5)	2654 (45.4)	3627 (53)	7027 (62.5)	12618 (67.5)	3912 (45.5)	2794 (38.3)	2571 (37.9)	2269 (37.3)
Diaspora children	1246 (23.3)	1229 (21.9)	1073 (18.3)	1001 (14.6)	1169 (10.4)	1660 (8.9)	1259 (14.7)	1059 (14.5)	972 (14.3)	796 (13.1)
Nursery children	992 (18.6)	1041 (18.6)	1271 (21.7)	1267 (18.5)	1798 (16.0)	2302 (12.3)	1961 (22.8)	2398 (32.9)	2326 (34.3)	2163 (35.5)
Peasant	519 (9.7)	539 (9.6)	614 (10.5)	670 (9.8)	926 (8.2)	1546 (8.3)	1100 (12.8)	761 (10.4)	666 (9.8)	616 (10.1)
Unemployment	33 (0.6)	33 (0.6)	45 (0.8)	48 (0.7)	57 (0.5)	101 (0.5)	62 (0.7)	55 (0.8)	61 (0.9)	68 (1.1)
Teacher	17 (0.3)	20 (0.4)	24 (0.4)	29 (0.4)	56 (0.5)	100 (0.5)	62 (0.7)	32 (0.4)	30 (0.4)	19 (0.3)
Other careers	168 (3.1)	141 (2.5)	169 (2.9)	201 (2.9)	216 (1.9)	376 (2)	236 (2.7)	197 (2.7)	162 (2.4)	154 (2.5)

The temporal distribution from 2016 to 2023 revealed distinct trends in both the Baidu search index and environmental factors (Figure 3). The distribution of the 3 Baidu index mirrored the

pattern observed in mumps incidence, with a peak in 2019 and followed by a decline. Most environmental factors except SO<sub>2</sub> displayed obvious seasonal cycles. Maximum and minimum

temperatures reached peaks in summer, while air pollutants, including CO, PM<sub>2.5</sub>, and PM<sub>10</sub>, had higher concentrations in winter. In addition, SO<sub>2</sub> and CO showed a long-term decreasing trend over the 10-year period.

**Figure 3.** The temporal distribution of weekly Baidu index and environmental factors in Yunnan Province from 2016 to 2023. CO: carbon monoxide; SO<sub>2</sub>: sulfur dioxide; PM: particulate matter; PM<sub>2.5</sub>: particulate matter with a diameter of 2.5 μm or less; PM<sub>10</sub>: particulate matter with a diameter of 10 μm or less.



### Correlation Between Predictors and Incidence of Mumps

The Pearson correlation analysis reveals that 3 Baidu search terms, disease name ( $r=0.67$ ), symptom ( $r=0.66$ ), and treatment ( $r=0.53$ ), show strong positive correlations with mumps

incidence in Yunnan Province from 2016 to 2023 (Table 2). Among environmental factors, both minimum ( $r=0.17$ ) and maximum temperature ( $r=0.16$ ) have weak positive correlations, while pollutants including SO<sub>2</sub> ( $r=-0.16$ ), CO ( $r=-0.12$ ), PM<sub>2.5</sub> ( $r=-0.20$ ) and PM<sub>10</sub> ( $r=-0.10$ ) show weak negative correlations with mumps incidence.

**Table 2.** Distribution of predictors and Pearson correlation coefficient between weekly predictors and mumps incidence in Yunnan province from 2016 to 2023.

Predictors	Baidu search terms in Chinese	Median (IQR)	PCC <sup>a</sup>	P value
<b>Baidu search terms</b>				
Disease name	—	2394 (2053-2697)	0.67	<.001
Mumps	流行性腮腺炎	560 (457-655)	0.42	<.001
Parotitis	腮腺炎	1397 (1225-1564)	0.76	<.001
Zha sai (a colloquial name for mumps in Chinese)	炸腮	301 (174-418)	0.36	<.001
San xian yan (a term phonetically like “parotitis” in Chinese)	三线炎	114 (57-172)	0.39	<.001
Symptom	—	2360 (1928-2821)	0.66	<.001
Jaw pain	下巴疼	114 (57-174)	0.38	<.001
Swelling on one side of the face	半边脸肿	0 (0-57)	0.23	<.001
Pain at the base of the ear	耳根痛	0 (0-0)	0.16	<.001
Facial swelling	脸肿	440 (348-524)	0.39	<.001
Complications of parotitis	腮腺炎并发症	0 (0-57)	0.24	<.001
Early symptoms of parotitis	腮腺炎的早期症状	115 (57-294)	0.26	<.001
Symptoms of parotitis + parotitis symptoms	腮腺炎的症状+腮腺炎症状	1315 (876-1560)	0.33	<.001
Parotitis with orchitis	腮腺炎睾丸炎	0 (0-0)	0.13	.01
Is parotitis contagious?	腮腺炎传染吗	436 (172-508)	0.48	<.001
Treatment	—	860 (729-1000)	0.53	<.001
What medicine to take for parotitis	腮腺炎吃什么药	464 (441-506)	0.30	<.001
How to treat parotitis	腮腺炎怎么治疗	300 (233-405)	0.54	<.001
Treatment of parotitis	腮腺炎的治疗	57 (0-115)	0.13	.007
<b>Environmental factors</b>				
Maximum temperature, °C	—	24 (20.2-26.3)	0.16	<.001
Minimum temperature, °C	—	12.8 (7.6-17.2)	0.17	<.001
SO <sub>2</sub> <sup>b</sup> (µg/m <sup>3</sup> )	—	8.2 (7.4-10.6)	-0.20	<.001
CO <sup>c</sup> (mg/m <sup>3</sup> )	—	0.7 (0.6-0.8)	-0.12	.02
PM <sub>2.5</sub> <sup>d</sup> (µg/m <sup>3</sup> )	—	19.5 (13.8-27.4)	-0.20	<.001
PM <sub>10</sub> (µg/m <sup>3</sup> )	—	34.3 (25.5-46.0)	-0.10	.04

<sup>a</sup>PCC: Pearson correlation coefficient.

<sup>b</sup>SO: sulfur dioxide.

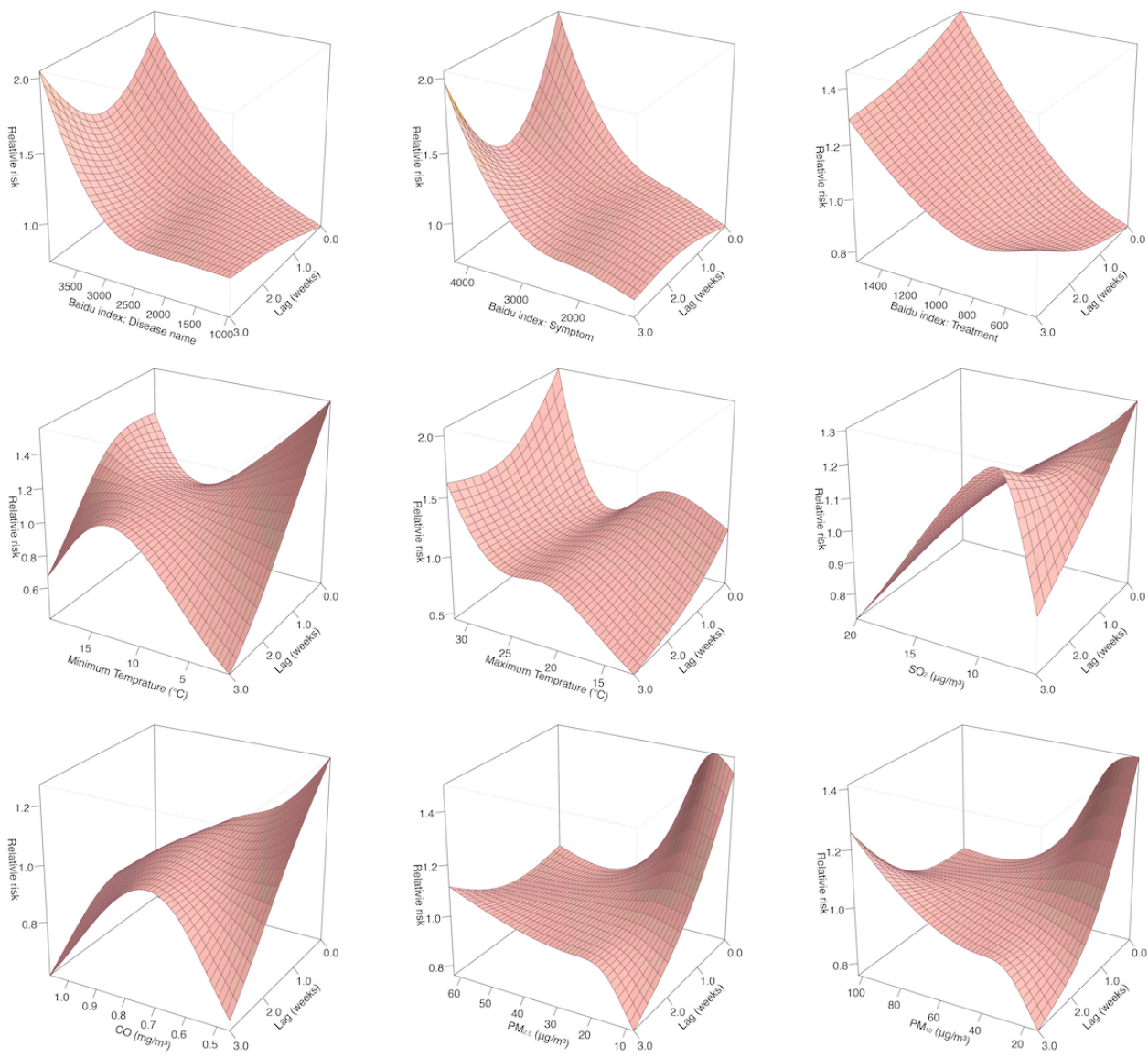
<sup>c</sup>CO: carbon monoxide.

<sup>d</sup>PM: particulate matter.

Considering the potential lag impact of these predictors on mumps incidence, we leveraged DLNM to explore their short-term impacts (Figure 4). The impact of various predictors on the incidence of mumps varies across different lag weeks. For 3 Baidu search terms, the RR consistently increases with the rising Baidu index. The analysis also revealed nonlinear

associations between temperatures and incidence, with minimum temperature showing a U-shaped curve and maximum temperature exhibiting a complex pattern. In addition, elevated levels of SO<sub>2</sub>, CO, PM<sub>2.5</sub>, and PM<sub>10</sub> generally corresponded with decreases in RR at a 1-week lag.

**Figure 4.** Relative risks for mumps incidence by predictors and lags using distributed lag nonlinear models in Yunnan province from 2016 to 2023, with references at median of each predictor.



**Model Performances and Comparisons**

Table 3 shows the comparisons of the performance metrics of 4 models with different combinations of predictors. Model IBE, which includes all 3 predictors, consistently outperformed the other models in both the train and validation set, as well as the test set. Specifically, model IBE achieved the highest  $R^2$  values (0.89 in the train and validation set, 0.72 in the test set) and the lowest error metrics, with an MAE of 0.51, MAPE of 15.2%, and RMSE of 0.74 in the train and validation set, and an MAE

of 0.33, MAPE of 15.9%, and RMSE of 0.43 in the test set. In comparison, the IE model ( $R^2=0.54$ , MAE=0.45, MAPE=28.2%, and RMSE=0.55) and the IB model ( $R^2=0.49$ , MAE=0.45, MAPE=29.2%, and RMSE=0.58) showed similar levels of prediction accuracy in the test set, both outperforming model BE, which had the worst performance ( $R^2=0.33$ , MAE=0.54, MAPE=32.0%, and RMSE=0.66). The actual values and model predictions of the weekly incidence of mumps are presented in Figure 5.



**Table 3.** Performance metrics of different models.

	Model BE <sup>a</sup>	Model IE <sup>b</sup>	Model IB <sup>c</sup>	Model IBE <sup>d</sup>
<b>Train and validation set</b>				
$R^2$	0.80	0.84	0.84	0.89
MAE <sup>e</sup>	0.69	0.59	0.65	0.51
MAPE <sup>f</sup> (%)	21.2	17.1	21.6	15.2
RMSE <sup>g</sup>	0.97	0.87	0.88	0.74
<b>Test set</b>				
$R^2$	0.33	0.54	0.49	0.72
MAE	0.54	0.45	0.45	0.33
MAPE (%)	32.0	28.2	29.2	15.9
RMSE	0.66	0.55	0.58	0.43

<sup>a</sup>Model BE: the model included Baidu index and environmental factors.

<sup>b</sup>Model IB: the model included mumps incidence and Baidu index.

<sup>c</sup>Model IE: the model included mumps incidence and environmental factors.

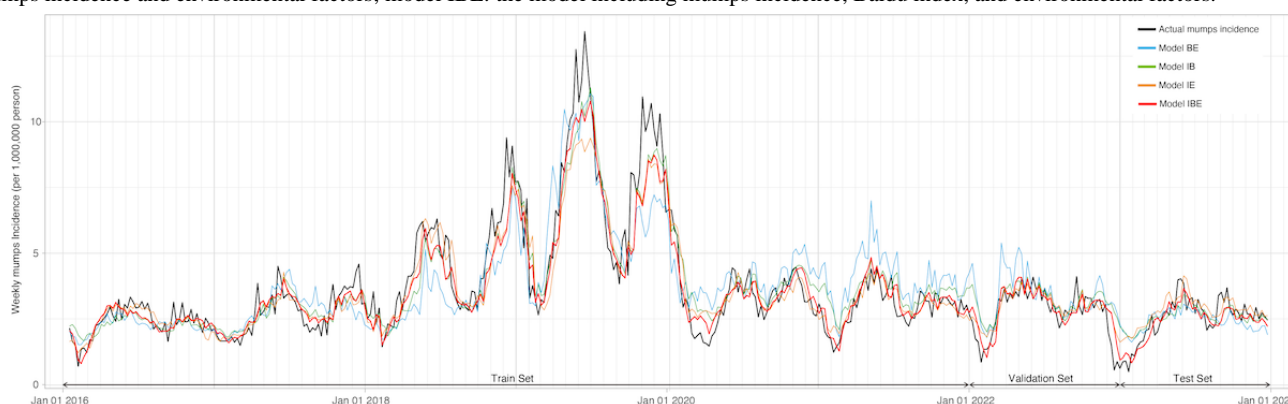
<sup>d</sup>Model IBE: the model included mumps incidence, Baidu index, and environmental factors.

<sup>e</sup>MAE: mean absolute error.

<sup>f</sup>MAPE: mean absolute percentage error.

<sup>g</sup>RMSE: root-mean-square error.

**Figure 5.** Predictions of weekly mumps incidence 1 week in advance by different models in Yunnan province from 2016 to 2023. Model BE: a model including Baidu index and environmental factors; model IB: the model including mumps incidence and Baidu index; model IE: a model including mumps incidence and environmental factors; model IBE: the model including mumps incidence, Baidu index, and environmental factors.



## Discussion

### Principal Findings

Our study presented the incidence of mumps from 2014 to 2023 in Yunnan, a representative province of irregular mumps activity, and found continuous mumps outbreaks in different cities from 2017 to 2020. Using multisource data, we constructed models for predicting the incidence of mumps one week in advance based on LSTM. Model IBE (ie, the model including the incidence of mumps, Baidu index, and environmental factors) showed the best performance among 4 models.

### Comparison to Previous Work

The mean incidence of mumps in Yunnan from 2014 to 2023 was 16.9 per 100,000 persons, which is lower than that of the period between 2004 to 2012 (26.2/100,000) in Yunnan and

comparable to the incidence in mainland China from 2013 to 2018 (16.9/100,000) [19]. The 10-year incidence pattern of mumps in Yunnan shows an increase followed by a decrease starting in 2020. The decline may be related to the COVID-19 pandemic, which extensively reduced opportunities for interpersonal contact, thereby blocking the spread of infectious diseases. However, we do not have a definite explanation for the peak in 2019. Given the rapidly increasing number of cases among students in 2019, factors such as vaccine coverage, waning immunity, and the close gathering of adolescents on campuses may have contributed to the outbreak [3,33].

Previous studies have found nonlinear associations between meteorological factors and the incidence of several infectious diseases including mumps [34,35], influenza [23,24,36], and chickenpox [37]. These findings suggest meteorological factors are common predictors for the field of infectious disease

prediction. Our study found 6 environmental factors with a PCC of  $>0.10$  and presented their lag correlations with the incidence of mumps. The most relevant factors were  $PM_{2.5}$  and  $SO_2$ , both showing weak negative correlations ( $PCC=-0.2$ ) with mumps incidence in Yunnan. Notably, the PCC is obviously lower than that of other infectious disease. For example, a study in Singapore found a stronger correlation ( $PCC=0.32$ ) between minimum temperature and the COVID-19 pandemic incidence [38]. This variability suggests that while environmental factors can be valuable for predicting some infectious diseases, they may not be sufficient for accurate prediction across all diseases, highlighting the need for additional predictors.

In recent years, using search engine data, such as Google Trends [39], and the Baidu index [24,40], to predict the incidence of infectious disease has become popular and demonstrated excellent predictive performance. The advantages of search engine data included timeliness, low cost, and sometimes stronger correlation with disease incidence than other data sources. Given the different disease outcomes, fairly comparing our study with other research using search engine data is challenging, as this is, to our knowledge, the first study to construct a mumps prediction model using search engine data. However, we noted that Yang et al [24] and Choi and Ahn [39], who predicted influenza incidence in mainland China and South Korea respectively, achieved similar or even better performance using Baidu index and Google Trends alone, compared with our study's combination of the Baidu index and environmental factors. It suggests that the contribution of search engine data to incidence prediction may vary depending on types of disease, regions, and data sources. In our study, the PCCs of the Baidu index of 3 search term groups were higher than those of environmental factors. This may be because the Baidu index, derived from individuals' active search activities, better reflected the incidence peak between 2018 to 2020. In contrast, the relatively stable environmental factors might contribute more to the seasonal trends in incidence but failed to capture the sharp fluctuations observed during outbreaks.

Our study deployed multisource data to construct and compare 4 models, with the model IBE demonstrating superior performance. Notably, when Baidu index data was incorporated, model IBE exhibited even greater predictive accuracy compared with the IE model. This demonstrates the capability of the Baidu index to predict mumps incidence and the potential of integrating multiple data sources to complement the surveillance systems in predicting mumps more accurately, as these data sets can complement one another and provide a more comprehensive understanding of the influencing factors and early signs of disease spread. Further research could build on the proposed framework in our study or explore the integration of additional data sources, such as social media activity and movement behaviors, to further enhance the predictive models.

### Strengths and Limitations

To our knowledge, our study is the first to develop a prediction model for the incidence of mumps that combines the Baidu

index and environmental factors as predictors, and it is also the first to create such a model specifically for Yunnan province, China. In addition, by leveraging data such as the Baidu index, our model offers a novel and practical approach to monitor mumps incidence 1 week in advance.

However, there are several limitations to our study. First, at the city level rather than the provincial level, we observed different outbreak times among cities from 2017 to 2020, which may indicate the spread trajectory of mumps and be important for its control and prevention. Unfortunately, the time-series forecast model in our study cannot leverage this geographical information. Future studies could explore integrating temporal and spatial data, which may be crucial for preventing widespread transmission. Second, some Baidu search terms were introduced after 2016 (eg, the search term "Symptoms and treatment of mump" was introduced on April 2, 2018), thereby decreasing the number of available predictors of the Baidu index, which may compromise the performance of Baidu index. Nonetheless, after extensive research, our study identified and included a sufficient number of predictors. Third, we did not have access to some meteorological factors such as sunshine duration and wind speed, which have been identified as effective predictors in other studies, potentially hindering better performance of our models. Future research should incorporate a broader range of meteorological variables to improve prediction accuracy further. Fourth, applying the model to practical public health interventions still faced several challenges. Future work should focus on developing a user-friendly tool that can automatically collect predictors and output results, making the model more accessible and actionable for public health authorities. Finally, due to economic and geographic variations in China [19], various epidemic patterns of infectious diseases, and the different nature of data sets from other search platforms, this model may not generalize well across different regions, diseases, or data sources without careful adaptation and validation. While the model in this study was applied to mumps, similar approaches could potentially be applied to other infectious diseases with seasonal or outbreak-driven incidence patterns. Diseases such as influenza, chickenpox, and hand, foot, and mouth disease, where public awareness, mobility, and environmental factors play a role in transmission, might benefit from similar predictive frameworks. Future studies should focus on tailoring models to specific diseases and data sets to evaluate their broader applicability.

### Conclusions

In conclusion, our study showed the incidence of mumps in Yunnan province, China, from 2014 to 2023, highlighting a sharp peak around 2019. We used data from 2016 to 2023, including mumps incidence, the Baidu index, and environmental factors, to construct 4 prediction models based on LSTM networks. Among these models, the model IBE demonstrates the best performance. This indicates the feasibility of using search engine data and environmental factors to forecast the incidence of mumps, offering a valuable tool for public health surveillance and rapid outbreak assessment.

## Acknowledgments

XX, LX, and IXW designed the study. XX extracted data. LX conducted data analysis. XX and LX wrote the draft. IX W, SD, and LC critically revised the manuscript.

Special thanks would be extended to IXW and SD for eququally involvling in the correspondence with collaborators and the review process. While not listed as a correspoding author, IXW made significant contributions throughout the process.

We would like to thank Baidu and the China Center for Disease Control and Prevention.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Additional tables.

[\[DOCX File , 25 KB-Multimedia Appendix 1\]](#)

## References

1. Grennan D. Mumps. *JAMA*. 2019;322(10):1022. [doi: [10.1001/jama.2019.10982](https://doi.org/10.1001/jama.2019.10982)] [Medline: [31503310](https://pubmed.ncbi.nlm.nih.gov/31503310/)]
2. Lam E, Rosen JB, Zucker JR. Mumps: an update on outbreaks, vaccine efficacy, and genomic diversity. *Clin Microbiol Rev*. 2020;33(2):e00151-e00119. [FREE Full text] [doi: [10.1128/CMR.00151-19](https://doi.org/10.1128/CMR.00151-19)] [Medline: [32102901](https://pubmed.ncbi.nlm.nih.gov/32102901/)]
3. Principi N, Esposito S. Mumps outbreaks: a problem in need of solutions. *J Infect*. 2018;76(6):503-506. [doi: [10.1016/j.jinf.2018.03.002](https://doi.org/10.1016/j.jinf.2018.03.002)] [Medline: [29678496](https://pubmed.ncbi.nlm.nih.gov/29678496/)]
4. Dayan GH, Quinlisk MP, Parker AA, Barskey AE, Harris ML, Schwartz JMH, et al. Recent resurgence of mumps in the United States. *N Engl J Med*. 2008;358(15):1580-1589. [doi: [10.1056/NEJMoa0706589](https://doi.org/10.1056/NEJMoa0706589)] [Medline: [18403766](https://pubmed.ncbi.nlm.nih.gov/18403766/)]
5. Di Pietrantonj C, Rivetti A, Marchione P, Debalini MG, Demicheli V. Vaccines for measles, mumps, rubella, and varicella in children. *Cochrane Database Syst Rev*. 2021;11(11):CD004407. [FREE Full text] [doi: [10.1002/14651858.CD004407.pub5](https://doi.org/10.1002/14651858.CD004407.pub5)] [Medline: [34806766](https://pubmed.ncbi.nlm.nih.gov/34806766/)]
6. Galazka AM, Robertson SE, Kraigher A. Mumps and mumps vaccine: a global review. *Bull World Health Organ*. 1999;77(1):3-14. [FREE Full text] [Medline: [10063655](https://pubmed.ncbi.nlm.nih.gov/10063655/)]
7. Kauffmann F, Heffernan C, Meurice F, Ota MOC, Vetter V, Casabona G. Measles, mumps, rubella prevention: how can we do better? *Expert Rev Vaccines*. 2021;20(7):811-826. [FREE Full text] [doi: [10.1080/14760584.2021.1927722](https://doi.org/10.1080/14760584.2021.1927722)] [Medline: [34096442](https://pubmed.ncbi.nlm.nih.gov/34096442/)]
8. World Health Organization. Immunization coverage. URL: <https://www.who.int/en/news-room/fact-sheets/detail/immunization-coverage> [accessed 2024-11-25]
9. Zamir CS, Schroeder H, Shoob H, Abramson N, Zentner G. Characteristics of a large mumps outbreak: clinical severity, complications and association with vaccination status of mumps outbreak cases. *Hum Vaccin Immunother*. 2015;11(6):1413-1417. [FREE Full text] [doi: [10.1080/21645515.2015.1021522](https://doi.org/10.1080/21645515.2015.1021522)] [Medline: [25874726](https://pubmed.ncbi.nlm.nih.gov/25874726/)]
10. Aasheim ET, Inns T, Trindall A, Emmett L, Brown KE, Williams CJ, et al. Outbreak of mumps in a school setting, United Kingdom, 2013. *Hum Vaccin Immunother*. 2014;10(8):2446-2449. [FREE Full text] [doi: [10.4161/hv.29484](https://doi.org/10.4161/hv.29484)] [Medline: [25424953](https://pubmed.ncbi.nlm.nih.gov/25424953/)]
11. Ferenczi A, Gee S, Cotter S, Kelleher K, Mumps Outbreak Control Team. Ongoing mumps outbreak among adolescents and young adults, Ireland, August 2018 to January 2020. *Euro Surveill*. 2020;25(4):2000047. [FREE Full text] [doi: [10.2807/1560-7917.ES.2020.25.4.2000047](https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000047)] [Medline: [32019666](https://pubmed.ncbi.nlm.nih.gov/32019666/)]
12. Marshall HS, Plotkin S. The changing epidemiology of mumps in a high vaccination era. *Lancet Infect Dis*. 2019;19(2):118-119. [doi: [10.1016/S1473-3099\(18\)30541-3](https://doi.org/10.1016/S1473-3099(18)30541-3)] [Medline: [30558996](https://pubmed.ncbi.nlm.nih.gov/30558996/)]
13. Dong Y, Wang L, Burgner DP, Miller JE, Song Y, Ren X, et al. Infectious diseases in children and adolescents in China: analysis of national surveillance data from 2008 to 2017. *BMJ*. 2020;369:m1043. [FREE Full text] [doi: [10.1136/bmj.m1043](https://doi.org/10.1136/bmj.m1043)] [Medline: [32241761](https://pubmed.ncbi.nlm.nih.gov/32241761/)]
14. Cui A, Zhu Z, Hu Y, Deng X, Sun Z, Zhang Y, et al. Mumps epidemiology and mumps virus genotypes circulating in mainland China during 2013-2015. *PLoS One*. 2017;12(1):e0169561. [FREE Full text] [doi: [10.1371/journal.pone.0169561](https://doi.org/10.1371/journal.pone.0169561)] [Medline: [28085897](https://pubmed.ncbi.nlm.nih.gov/28085897/)]
15. Paradise Vit A, Magid A. Differences in fear and negativity levels between formal and informal health-related websites: analysis of sentiments and emotions. *J Med Internet Res*. 2024;26:e55151. [FREE Full text] [doi: [10.2196/55151](https://doi.org/10.2196/55151)] [Medline: [39120928](https://pubmed.ncbi.nlm.nih.gov/39120928/)]
16. Westphal DW, Eastwood A, Levy A, Davies J, Huppertz C, Gilles M, et al. A protracted mumps outbreak in Western Australia despite high vaccine coverage: a population-based surveillance study. *Lancet Infect Dis*. 2019;19(2):177-184. [doi: [10.1016/S1473-3099\(18\)30498-5](https://doi.org/10.1016/S1473-3099(18)30498-5)] [Medline: [30558994](https://pubmed.ncbi.nlm.nih.gov/30558994/)]

17. Department of Planning, Development, and Information Technology, National Health Commission of the People's Republic of China. Statistical bulletin of China's health development in 2023. *Chin J Viral Dis.* 2024;14(05):416-424. [FREE Full text] [doi: [10.16505/j.2095-0136.2024.0083](https://doi.org/10.16505/j.2095-0136.2024.0083)]
18. Department of Planning, Development and Information Technology, National Health Commission of the People's Republic of China. Statistical bulletin of China's health development in 2022. *Chin J Viral Dis.* 2024;14(04):317-325. [FREE Full text] [doi: [10.16505/j.2095-0136.2024.0066](https://doi.org/10.16505/j.2095-0136.2024.0066)]
19. Fu X, Ge M, Xu W, Yu M, Ju J, Zhong Y, et al. Epidemiological features and sociodemographic factors associated with mumps in mainland China from 2004 to 2018. *J Med Virol.* 2022;94(10):4850-4859. [FREE Full text] [doi: [10.1002/jmv.27955](https://doi.org/10.1002/jmv.27955)] [Medline: [35739613](https://pubmed.ncbi.nlm.nih.gov/35739613/)]
20. Zhang H, Su K, Zhong X. Association between meteorological factors and mumps and models for prediction in Chongqing, China. *Int J Environ Res Public Health.* 2022;19(11):6625. [FREE Full text] [doi: [10.3390/ijerph19116625](https://doi.org/10.3390/ijerph19116625)] [Medline: [35682208](https://pubmed.ncbi.nlm.nih.gov/35682208/)]
21. Jia J, Liu M, Xue Z, Wang Z, Pan Y. Prediction of mumps incidence trend in China based on difference grey model and artificial neural network learning. *Iran J Public Health.* 2021;50(7):1405-1415. [FREE Full text] [doi: [10.18502/ijph.v50i7.6630](https://doi.org/10.18502/ijph.v50i7.6630)] [Medline: [34568179](https://pubmed.ncbi.nlm.nih.gov/34568179/)]
22. Wang Z, He J, Jin B, Zhang L, Han C, Wang M, et al. Using baidu index data to improve chickenpox surveillance in Yunnan, China: infodemiology study. *J Med Internet Res.* 2023;25:e44186. [FREE Full text] [doi: [10.2196/44186](https://doi.org/10.2196/44186)] [Medline: [37191983](https://pubmed.ncbi.nlm.nih.gov/37191983/)]
23. Su K, Xu L, Li G, Ruan X, Li X, Deng P, et al. Forecasting influenza activity using self-adaptive AI model and multi-source data in Chongqing, China. *EBioMedicine.* 2019;47:284-292. [FREE Full text] [doi: [10.1016/j.ebiom.2019.08.024](https://doi.org/10.1016/j.ebiom.2019.08.024)] [Medline: [31477561](https://pubmed.ncbi.nlm.nih.gov/31477561/)]
24. Yang L, Zhang T, Han X, Yang J, Sun Y, Ma L, et al. Influenza epidemic trend surveillance and prediction based on search engine data: deep learning model study. *J Med Internet Res.* 2023;25:e45085. [FREE Full text] [doi: [10.2196/45085](https://doi.org/10.2196/45085)] [Medline: [37847532](https://pubmed.ncbi.nlm.nih.gov/37847532/)]
25. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997;9(8):1735-1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)] [Medline: [9377276](https://pubmed.ncbi.nlm.nih.gov/9377276/)]
26. Lara-Benítez P, Carranza-García M, Riquelme JC. An experimental review on deep learning architectures for time series forecasting. *Int J Neural Syst.* 2021;31(3):2130001. [doi: [10.1142/S0129065721300011](https://doi.org/10.1142/S0129065721300011)] [Medline: [33588711](https://pubmed.ncbi.nlm.nih.gov/33588711/)]
27. Baidu index. URL: <https://index.baidu.com> [accessed 2024-08-16]
28. Tianqihoubao. URL: <http://www.tianqihoubao.com> [accessed 2024-09-20]
29. Zhang T, Qin W, Nie T, Zhang D, Wu X. Effects of meteorological factors on the incidence of varicella in Lu'an, Eastern China, 2015-2020. *Environ Sci Pollut Res Int.* 2023;30(4):10052-10062. [doi: [10.1007/s11356-022-22878-0](https://doi.org/10.1007/s11356-022-22878-0)] [Medline: [36066801](https://pubmed.ncbi.nlm.nih.gov/36066801/)]
30. Gasparini A, Armstrong B, Kenward MG. Distributed lag non-linear models. *Stat Med.* 2010;29(21):2224-2234. [FREE Full text] [doi: [10.1002/sim.3940](https://doi.org/10.1002/sim.3940)] [Medline: [20812303](https://pubmed.ncbi.nlm.nih.gov/20812303/)]
31. Gasparini A. Distributed lag linear and non-linear models in R: the package dlnm. *J Stat Softw.* 2011;43(8):1-20. [FREE Full text] [Medline: [22003319](https://pubmed.ncbi.nlm.nih.gov/22003319/)]
32. Li R, Lin H, Liang Y, Zhang T, Luo C, Jiang Z, et al. The short-term association between meteorological factors and mumps in Jining, China. *Sci Total Environ.* 2016;568:1069-1075. [doi: [10.1016/j.scitotenv.2016.06.158](https://doi.org/10.1016/j.scitotenv.2016.06.158)] [Medline: [27353959](https://pubmed.ncbi.nlm.nih.gov/27353959/)]
33. Vygen S, Fischer A, Meurice L, Mouchetrou Njoya I, Gregoris M, Ndiaye B, et al. Waning immunity against mumps in vaccinated young adults, France 2013. *Euro Surveill.* 2016;21(10):30156. [FREE Full text] [doi: [10.2807/1560-7917.ES.2016.21.10.30156](https://doi.org/10.2807/1560-7917.ES.2016.21.10.30156)] [Medline: [26987576](https://pubmed.ncbi.nlm.nih.gov/26987576/)]
34. Yu G, Yang R, Yu D, Cai J, Tang J, Zhai W, et al. Impact of meteorological factors on mumps and potential effect modifiers: an analysis of 10 cities in Guangxi, Southern China. *Environ Res.* 2018;166:577-587. [doi: [10.1016/j.envres.2018.06.047](https://doi.org/10.1016/j.envres.2018.06.047)] [Medline: [29966878](https://pubmed.ncbi.nlm.nih.gov/29966878/)]
35. Hu W, Li Y, Han W, Xue L, Zhang W, Ma W, et al. Meteorological factors and the incidence of mumps in Fujian Province, China, 2005-2013: Non-linear effects. *Sci Total Environ.* 2018;619-620:1286-1298. [FREE Full text] [doi: [10.1016/j.scitotenv.2017.11.108](https://doi.org/10.1016/j.scitotenv.2017.11.108)] [Medline: [29734606](https://pubmed.ncbi.nlm.nih.gov/29734606/)]
36. Athanasiou M, Fragkozidis G, Zarkogianni K, Nikita KS. Long short-term memory-based prediction of the spread of influenza-like illness leveraging surveillance, weather, and twitter data: model development and validation. *J Med Internet Res.* 2023;25:e42519. [FREE Full text] [doi: [10.2196/42519](https://doi.org/10.2196/42519)] [Medline: [36745490](https://pubmed.ncbi.nlm.nih.gov/36745490/)]
37. Chen B, Sumi A, Wang L, Zhou W, Kobayashi N. Role of meteorological conditions in reported chickenpox cases in Wuhan and Hong Kong, China. *BMC Infect Dis.* 2017;17(1):538. [FREE Full text] [doi: [10.1186/s12879-017-2640-1](https://doi.org/10.1186/s12879-017-2640-1)] [Medline: [28774264](https://pubmed.ncbi.nlm.nih.gov/28774264/)]
38. Pani SK, Lin N, RavindraBabu S. Association of COVID-19 pandemic with meteorological parameters over Singapore. *Sci Total Environ.* 2020;740:140112. [FREE Full text] [doi: [10.1016/j.scitotenv.2020.140112](https://doi.org/10.1016/j.scitotenv.2020.140112)] [Medline: [32544735](https://pubmed.ncbi.nlm.nih.gov/32544735/)]
39. Choi SB, Ahn I. Forecasting seasonal influenza-like illness in South Korea after 2 and 30 weeks using Google Trends and influenza data from Argentina. *PLoS One.* 2020;15(7):e0233855. [FREE Full text] [doi: [10.1371/journal.pone.0233855](https://doi.org/10.1371/journal.pone.0233855)] [Medline: [32673312](https://pubmed.ncbi.nlm.nih.gov/32673312/)]

40. Chen S, Liu X, Wu Y, Xu G, Zhang X, Mei S, et al. The application of meteorological data and search index data in improving the prediction of HFMD: a study of two cities in Guangdong Province, China. *Sci Total Environ.* 2019;652:1013-1021. [doi: [10.1016/j.scitotenv.2018.10.304](https://doi.org/10.1016/j.scitotenv.2018.10.304)] [Medline: [30380469](https://pubmed.ncbi.nlm.nih.gov/30380469/)]

## Abbreviations

**CO:** carbon monoxide  
**DLNM:** distributed nonlinear lag model  
**LSTM:** long short-term memory  
**MAE:** mean absolute error  
**MAPE:** mean absolute percentage error  
**PCC:** Pearson correlation coefficient  
**PM:** particulate matter  
**PM2.5:** particulate matter with a diameter of 2.5 µm or less  
**PM10:** particulate matter with a diameter of 10 µm or less  
**RMSE:** root-mean-square error  
**RR:** relative risk  
**SO2:** sulfur dioxide

*Edited by A Schwartz; submitted 03.09.24; peer-reviewed by R Guo, T Zhang; comments to author 27.09.24; revised version received 04.10.24; accepted 04.12.24; published 06.02.25*

*Please cite as:*

*Xiong X, Xiang L, Chang L, Wu IXY, Deng S*

*Forecasting the Incidence of Mumps Based on the Baidu Index and Environmental Data in Yunnan, China: Deep Learning Model Study*

*J Med Internet Res 2025;27:e66072*

*URL: <https://www.jmir.org/2025/1/e66072>*

*doi: [10.2196/66072](https://doi.org/10.2196/66072)*

*PMID:*

©Xin Xiong, Linghui Xiang, Litao Chang, Irene XY Wu, Shuzhen Deng. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 06.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.