The Effect of COVID-19 on Infectious Disease to Outpatient of Children: A Machine Learning Study

Dan Li¹, Dan Xiang¹, Shun-Xian Zhang²,³,⁴ and Jin-Xin Zheng²,³*

¹ Department of NICU, the Children’s Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, 310003, People’s Republic of China.

² School of Global Health, Chinese Centre for Tropical Diseases Research, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, People’s Republic of China.

³ National Institute of Parasitic Diseases at Chinese Centre for Disease Control and Prevention (Chinese Centre for Tropical Diseases Research), National Health Commission of the People’s Republic of China (NHC) Key Laboratory of Parasite and Vector Biology, World Health Organization (WHO) Collaborating Centre for Tropical Diseases, National Centre for International Research On Tropical Diseases of the Chinese Ministry of Science and Technology, Shanghai, 200025, People’s Republic of China.

⁴ Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, 200032, People’s Republic of China.

*Corresponding Author: Jin-Xin Zheng, National Institute of Parasitic Diseases at Chinese Centre for Disease Control and Prevention (Chinese Centre for Tropical Diseases Research), National Health Commission of the People’s Republic of China (NHC) Key Laboratory of Parasite and Vector Biology, World Health Organization (WHO) Collaborating Centre for Tropical Diseases, National Centre for International Research On Tropical Diseases of the Chinese Ministry of Science and Technology, Shanghai, 200025, People’s Republic of China.

DOI: https://doi.org/10.58624/SVOAPD.2023.02.042

Received: July 18, 2023  Published: August 07, 2023

Abstract

The outbreak of coronavirus disease 2019 (COVID-19) has had a significant impact worldwide. However, the effects of COVID-19 on human enteroviruses such as Herpangina and Hand, Foot, and Mouth Disease (HFMD) in children are still not fully understood. This study aimed to analyze the reported cases of Herpangina and HFMD among child outpatients in Hangzhou, China, from January 2018 to June 2021, before, during, and after the COVID-19 outbreak. Machine learning approaches were employed to analyze the time series data of Herpangina and HFMD from 2018 to 2019 and predict the trends and seasonal patterns from 2020 to 2021. The seasonal patterns at weekly and monthly levels were compared across the three stages of the COVID-19 outbreak. Prior to 2020, the reported incidence of Herpangina and HFMD exhibited a certain degree of seasonal periodicity. However, following the COVID-19 outbreak in 2020, there was a sharp decrease in the cases of Herpangina and HFMD, followed by a rebound in the first half of 2021. The observed cases reached their highest peak in July 2020, showing a decrease of 93.6% for Herpangina and 96.5% for HFMD compared to the predicted values. These findings indicate that the COVID-19 outbreak had a significant impact on the transmission of Herpangina and HFMD. Therefore, integrated strategies should be implemented to control and prevent the resurgence of Herpangina and HFMD, alongside the effective management of COVID-19.

Keywords: Herpangina; HFMD; Machine Learning; COVID-19

Background

Human enteroviruses are a common infectious agent primarily affecting children and leading to the development of Herpangina and Hand, Foot, and Mouth Disease (HFMD) [1]. These viral infections are prevalent worldwide and primarily affect children under the age of five [2]. Herpangina is characterized by the formation of ulcers at the back of the mouth [3], while HFMD causes the appearance of blisters on the hands, feet, and inside the mouth [4].
Both diseases exhibit rapid and wide transmission and are primarily spread through respiratory droplets, as well as through nasal, oral, and other bodily secretions and excretions [5]. Despite the potential severity of these infections, including complications involving the nervous system and the risk of reduced cognitive function or mortality, they are often overlooked. Consequently, Herpangina and HFMD impose a significant disease burden, surpassing that of other pediatric infectious diseases. In response, national and local health authorities must implement stringent strategies for controlling and preventing outbreaks, as these diseases pose a substantial threat to public health [6].

Since the emergence of the coronavirus disease 2019 (COVID-19) at the end of 2019, it has rapidly spread worldwide[7]. In response to this global health crisis, the Chinese government implemented a comprehensive and proactive approach to public health emergencies, which included effective intervention measures like physical distancing, handwashing, and mask usage, resulting in the containment of COVID-19 transmission[8][9]. Interestingly, these control measures not only suppressed COVID-19 but also led to a decrease in the incidence of other infectious diseases that are primarily transmitted through contact or respiratory droplets[10]. Recent studies have reported significant reductions in diseases such as chickenpox, influenza, acute otitis media (AOM), and pneumonia, particularly among children[11][12].

The field of Machine Learning has witnessed remarkable advancements and has become instrumental in addressing real-world problems, including forecasting time-series data[13]. Machine Learning techniques have been successfully employed to predict trends related to COVID-19[14][15], as well as to analyze and predict the outbreaks of other infectious diseases[16]. However, to the best of our knowledge, there is a paucity of studies that have utilized combined machine learning algorithms to forecast the incidence of Herpangina and HFMD in time series analysis.

This retrospective study aimed to assess the impact of the COVID-19 pandemic on the incidence of human enterovirus diseases, specifically Herpangina and Hand, Foot, and Mouth Disease (HFMD), among children in Hangzhou, China. Our hypothesis was that in the absence of the COVID-19 outbreak, Herpangina and HFMD would exhibit seasonal fluctuations with distinct peaks. Therefore, we analyzed the incidence of these two diseases from 2018 to 2021 and employed a machine learning model to simulate the patterns and changes in Herpangina and HFMD before and after the COVID-19 outbreak.

Methods

Study setting

This retrospective study was conducted at The Children's Hospital of Zhejiang University School of Medicine, which serves as the largest tertiary hospital and a leading center for comprehensive pediatric healthcare in Hangzhou, Zhejiang Province, China. The study aimed to analyze the reported cases of two human enteroviruses, Herpangina and Hand, Foot, and Mouth Disease (HFMD), within the hospital’s electronic medical record system. All outpatient cases of Herpangina and HFMD admitted to the hospital between January 1, 2018, and June 30, 2021, were included in the study. The data collected encompassed various demographics, including the patient’s name, age, gender, current address, onset date of symptoms, and date of diagnosis. Additionally, data on reported COVID-19 cases were collected from January 1, 2020, to June 30, 2021. Specifically, the focus was on locally reported COVID-19 cases in Hangzhou, which could be obtained from the Hangzhou Municipal Health Commission’s records.

Case definitions

A probable case of Hand, Foot, and Mouth Disease (HFMD) was defined as a patient exhibiting a rash on their hands, feet, limbs, or buttocks, along with ulcers or vesicles in the mouth, with or without fever [4]. Herpangina is characterized by a sudden onset of fever accompanied by symptoms such as sore throat, headache, anorexia, and frequently neck pain [17]. The diagnosis of both diseases is based on the presence of specific symptoms and characteristic oral lesions. A laboratory-confirmed case was defined as a probable case with laboratory evidence of coxsackievirus and enterovirus infection. Both HFMD and herpangina were assigned ICD-10 (International Statistical Classification of Diseases and Related Health Problems) codes in the hospital’s electronic medical record system.

Machine model constructions

The time series signature comprises a collection of valuable features that describe the time series index of a time-based dataset. It encompasses numerous features that aid in predicting future trends and patterns. One popular statistical method for time series analysis is the autoregressive integrated moving average (ARIMA) model [18]. The statistical properties of the ARIMA model have made it widely used in scientific papers [19]. However, the ARIMA model is based on the assumption of linearity in the underlying time series, which may be inadequate when dealing with real-world medical datasets that contain non-linear time series data [20].
To address the limitations of linear statistical time series models, researchers have proposed incorporating both linear and non-linear relationships by utilizing machine learning models such as Random Forest and XGBOOST [21][22][23][24]. These machine learning models have been extensively discussed in the literature and offer several advantages. Unlike statistical models, they do not assume specific statistical distributions for the underlying time series, and they naturally handle non-linear modeling [20].

In this study, we developed the ARIMA, Random Forest, and XGBOOST models to analyze the daily incidence data of Herpangina and Hand, Foot, and Mouth Disease (HFMD). The model simulation and prediction were conducted based on the following steps:

1. Collection of daily incidence data from January 1, 2018, to December 31, 2019, followed by splitting the data into training set (70%) and testing set (30%).

2. Determination of the best parameters for each model by simulating the fitted data using the three models on the training set and evaluating their performance on the testing set.


By following these steps, we aimed to analyze and forecast the incidence patterns of Herpangina and HFMD for the specified period.

**Machine model performance**

We assessed the predictive performance of our model using four commonly used forecasting parameters: Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) [25]. MAE represents the arithmetic average of the absolute errors between the predicted and true values, providing a measure of the mean absolute forecasting error. MSE serves as a loss function to quantify the error between the predicted and true values. RMSE is a widely employed metric that measures the differences between values predicted by a model or estimator and the observed values. It is calculated as the square root of the average squared error. MAPE quantifies accuracy as a percentage and can be computed as the cumulative absolute percent error for each time frame, obtained by dividing the absolute difference between the actual and predicted values by the actual values. In other words, it represents the mean error in percentage terms. For further details and mathematical examples, refer to the R package MLmetrics (https://cran.r-project.org/web/packages/MLmetrics/).

**Seasonal patterns**

In order to assess the changes in incidence during the COVID-19 pandemic in 2020, we aimed to analyze the time series by decomposing it into different components, each representing a distinct underlying pattern category. Typically, a time series can be divided into three components: a trend-cycle component, a seasonal component, and a residual component (which encompasses any other variations in the time series). However, for the purpose of this study, our focus was solely on the seasonal component, examining weekly and monthly patterns to better understand the impact of COVID-19 on the incidence cases of Herpangina and Hand, Foot, and Mouth Disease (HFMD).

**Statistical analysis**

All statistical analyses, along with data visualization, were conducted using the R programming language (version 4.0.2, https://cran.r-project.org). We utilized the modet ime package (version 0.6.1, https://cran.r-project.org/web/packages/modet ime) for data cleaning and the timetk package (version 2.6.1, https://cran.r-project.org/web/packages/timetk) for simulating and predicting machine learning models. Additionally, the ggplot2 package (version 3.3.5, https://cran.r-project.org/web/packages/ggplot2) was employed for creating visualizations of the data.

**Ethical approval**

The retrospective nature of this study necessitated the removal of patients' information from our database to ensure confidentiality and privacy. The research protocol received approval from the ethical committees of all participating hospitals, demonstrating adherence to ethical guidelines and standards.
The Effect of COVID-19 on Infectious Disease to Outpatient of Children: A Machine Learning Study

Result

The Incident case of Herpangina and HFMD from 2018 to 2021

During the period from January 1, 2018, to June 30, 2021, a total of 84,628 cases of Herpangina and 19,700 cases of Hand, Foot, and Mouth Disease (HFMD) were reported in study hospital. In Hangzhou, the first COVID-19 outbreaks were recorded in January 2020, and as of June 30, 2021, there were a cumulative total of 1,340 COVID-19 cases. (Table 1)

Table1: The reported cases of Herpangina, HFMD and COVID-19 between 2018 and 2021.

<table>
<thead>
<tr>
<th>Year</th>
<th>Herpangina</th>
<th>HFMD</th>
<th>COVID-19</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>39108</td>
<td>10761</td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>25002</td>
<td>5348</td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td>8654</td>
<td>2168</td>
<td>1290</td>
</tr>
<tr>
<td>2021</td>
<td>11864</td>
<td>1423</td>
<td>50</td>
</tr>
</tbody>
</table>

Before 2020, both Herpangina and Hand, Foot, and Mouth Disease (HFMD) exhibited a certain degree of seasonal periodicity in their reported incidences. Herpangina had higher incidence cases compared to HFMD, but they shared similar epidemic peaks in terms of incidence. Prior to 2020, the first high peak occurred in June and July, followed by a sharp decrease, while the second smaller peak was observed in September and October (refer to Supplement 1).

Supplement 1

Following the outbreak of COVID-19 in 2020, the number of cases for both Herpangina and HFMD significantly decreased compared to previous years. Moreover, the peaks in incidence were observed to shift to November and December. As preventive and control measures were implemented for COVID-19, the number of locally confirmed COVID-19 cases gradually declined, reaching a zero trend since April 2020. Consequently, the reported cases of COVID-19 remained relatively low. However, there was a notable increase in the number of Herpangina and HFMD cases, particularly in June (see Figure 1).
The Effect of COVID-19 on Infectious Disease to Outpatient of Children: A Machine Learning Study

To analyze the trends and patterns in the number of Herpangina and Hand, Foot, and Mouth Disease (HFMD) cases in hospitals from January 2018 to December 2019, we employed machine learning techniques such as ARIMA, Random Forest, and XGBOOST. These models were trained on the training set and evaluated on the testing set.

In the case of Herpangina, the XGBOOST model demonstrated high accuracy on the testing set, with minimal differences observed among the Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) values (see Table 2). On the other hand, the ARIMA model displayed lower accuracy compared to the Random Forest model.

When analyzing the forecasting figure (Figure 2A), it was observed that the ARIMA model underestimated the number of cases, while the Random Forest model overestimated them. Notably, only the XGBOOST model was able to accurately predict the fluctuating values over time.

Table 2: The performance of Machine learning models in predicted and observed value in testing set.

<table>
<thead>
<tr>
<th></th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Herpangina</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIMA (0,1,1)</td>
<td>15.77</td>
<td>2.08</td>
<td>35.55</td>
<td>17.68</td>
</tr>
<tr>
<td>Random Forest</td>
<td>11.77</td>
<td>1.55</td>
<td>35.32</td>
<td>14.12</td>
</tr>
<tr>
<td>XGBOOST</td>
<td>7.95</td>
<td>1.05</td>
<td>21.63</td>
<td>10.25</td>
</tr>
<tr>
<td><strong>HFMD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIMA (0,1,1)</td>
<td>3.25</td>
<td>0.95</td>
<td>3.98</td>
<td>79.13</td>
</tr>
<tr>
<td>Random Forest</td>
<td>4.95</td>
<td>1.45</td>
<td>5.64</td>
<td>116.04</td>
</tr>
<tr>
<td>XGBOOST</td>
<td>4.08</td>
<td>1.20</td>
<td>4.95</td>
<td>88.71</td>
</tr>
</tbody>
</table>

Figure 1: The reported cases of Herpangina, HFMD and COVID-19 between 2018 and 2021

Figure 2: The fitted values by Machine learning models of Herpangina, HFMD in testing set.
We also conducted time series forecasting of HFMD cases using the three models. Among them, XGBOOST exhibited the best performance, as indicated in Table 2. Although all of the model predictions tended to overestimate the cases in the testing set (Figure 2B), the XGBOOST model effectively captured the fluctuating changes on a daily basis.

**Numbers of cases With Herpangina and HFMD, during, and after the COVID-19**

To assess the impact of COVID-19 on the incidence of Herpangina and Hand, Foot, and Mouth Disease (HFMD), we conducted further analysis of the periodic patterns in the time series data, focusing on monthly and weekly observations of both observed and predicted cases during and after the COVID-19 outbreak.

In 2020, for Herpangina, the number of patients seeking treatment at hospitals remained at a low level. However, after the 40th week, outpatient cases started to increase, with the highest peak observed in December (IQR: 62 [54, 71]). The best-fitted XGBOOST model predicted the highest peaks in the 27th (IQR: 240 [238, 244]) and 28th (IQR: 238 [237, 245]) weeks, as well as in June (IQR: 179 [195, 207]) and July (IQR: 237 [187, 241]). Comparing the predicted and observed cases, a 92.3% and 93.6% decrease was observed in June and July, respectively. In the first half of 2021, the observed outpatient cases increased over time, reaching the highest point in the 22nd week (IQR: 421 [405, 486]) and May (IQR: 221 [117, 221]). The XGBOOST model predicted outpatient cases in the 22nd week (IQR: 119 [119, 166]) and May (IQR: 86 [67, 110]) (Figure 3AB).

![Figure 3: The seasonal patterns of observed cases and predicted values in 2020 and first half of 2021 among Herpangina.](image_url)
For HFMD in 2020, a similar pattern was observed, with the number of patients presenting at hospitals remaining at a low level until the 38th week when outpatient cases began to increase. This pattern was also evident on a monthly basis, with an increase starting from August and reaching its peak in December (IQR: 20 [17, 25]). The best-fitted XGBoost model predicted two peaks in 2020, occurring in the 28th-30th and 39th weeks. On a monthly scale, two peaks were observed in July (IQR: 29 [26, 31]) and September (IQR: 22 [18, 30]). Comparing the predicted and observed cases, a 96.5% and 68.1% decrease was observed in July and September, respectively. Again, in the first half of 2021, the observed outpatient cases exhibited a U-shaped curve, initially decreasing and then increasing after the 17th week. On a monthly basis, the highest case count was recorded in January (IQR: 11 [8, 17]) and May (IQR: 17 [12, 22]). The predicted outpatient cases from the XGBoost model also showed a U-shaped curve over time, with the highest case counts in January (IQR: 9 [6, 14]) and May (IQR: 14 [13, 16]) (Figure 4AB).

**Figure 4:** The seasonal patterns of observed cases and predicted values in 2020 and first half of 2021 among HFMD.

**Discussion**

This study utilized machine learning models to simulate and predict the outpatient cases of Herpangina and Hand, Foot, and Mouth Disease (HFMD) before, during, and after the COVID-19 epidemic, covering the period from January 2018 to May 2021. This is the first study to examine the impact of the COVID-19 epidemic on these two infectious diseases among children in Hangzhou. The findings provide valuable insights for the government in implementing targeted strategies against COVID-19 and understanding the subsequent effects on infectious diseases among children.
Our study revealed clear seasonal variability in Herpangina cases prior to the COVID-19 outbreak. The highest peak occurred in June and July, with a smaller peak observed in September and October. However, during the COVID-19 epidemic, these seasonal patterns disappeared, likely due to the implementation of government strategies such as school closures, community management measures, home isolation, and the closure of leisure facilities [11]. It is well known that Herpangina primarily affects children under the age of 10 [17]. Worldwide, cases of Herpangina have shown seasonal variation, with studies from America indicating higher infection rates during the summer and fall months [26]. Moreover, in tropical climates, these diseases tend to occur during the rainy seasons, and a recent fatal outbreak reported in Japan also demonstrated strong seasonality for Herpangina [27]. Climate factors may contribute to the prevalence of the disease [28]. Our findings suggest that Herpangina exhibits seasonality before the COVID-19 outbreak, and after effectively controlling COVID-19, Herpangina cases start to peak in the summer once again.

Our machine learning XGBOOST model demonstrates strong performance in predicting the incidence of Herpangina for the year 2020 and the first half of 2021. According to our predictions, there was a peak in June to July 2020, followed by a gradual increase in cases during the summer of 2021. This can be attributed to the seasonality patterns of Herpangina observed in 2018 and 2019, as our XGBOOST model learned these regular patterns from the historical incidence data. However, the actual number of Herpangina cases in 2020 was significantly lower, with smaller peaks observed until November 2020. We tentatively attribute this increase in cases to the reopening of kindergartens, schools, and entertainment venues in the second half of 2020, following the successful control of the COVID-19 outbreak [29]. Furthermore, the predicted cases and real data of Herpangina exhibited a similar pattern in the first half of 2021. Based on historical data, the peak incidence of Herpangina typically occurs in the summer, especially from June to July, which is consistent with previous studies [3]. The increase in cases observed in 2021 further indicates the successful control of the COVID-19 epidemic in the first half of the year.

HFMD has emerged as one of the leading causes of child mortality and a significant public health concern in China. Numerous studies have described the epidemiological, virological, and pathological features of this epidemic. For instance, the largest pediatric infectious diseases center in Shanghai reported 28,058 outpatient cases of HFMD from 2007 to 2010, with the highest peak occurring in May [30]. Studies conducted in Singapore and Japan have also shown a certain degree of seasonal periodicity for HFMD, with some exhibiting one epidemic peak and others showing two epidemic peaks [31]. Our results further support the understanding that HFMD is a pediatric enteroviral disease that predominantly occurs during the summer.

The annual fluctuations in HFMD cases were accurately predicted by our seasonal models. This observation aligns with previous studies that extensively utilized the SEIR (Susceptible/Exposed/Infectious/Removed) model and ARIMA model to track fluctuations and seasonal patterns [32][33]. Niu et al. proposed that the control measures implemented during the COVID-19 outbreak may have played a role in limiting the spread of HFMD in China [10]. They detected a decrease in the effective reproduction number of HFMD in selected cities since the outbreak of COVID-19, with a subsequent rebound in the incidence and effective reproduction number of HFMD in the second half of 2020. Our research findings are consistent with Niu’s study, but we additionally demonstrate the decrease in HFMD cases in 2020 using a machine learning model to capture seasonal changes. According to our model, during the COVID-19 epidemic in 2020, the number of HFMD patients decreased dramatically and subsequently returned to a high peak in the first half of 2021. This suggests that the government’s prevention and control measures for COVID-19 also had a positive impact on reducing the transmission of HFMD.

COVID-19 has had a profound impact on global health and the economy. However, the response measures implemented to control the spread of COVID-19, such as physical distancing, self-isolation, and community lockdowns, can inadvertently affect the transmission of other infectious diseases, including Kawasaki disease, coxsackievirus, and human enterovirus diseases [34][10]. During the COVID-19 outbreak, governments implemented stringent control measures to contain the epidemic, diverting resources towards COVID-19 response efforts. This shift in focus limited access to preventive care for other infectious diseases affecting children. We believe that the government-led public health response to COVID-19, with its proactive control measures and strategies in China [8], may have had a positive impact on the control of other infectious diseases among children during subsequent outbreaks. However, in countries where the COVID-19 situation remains challenging and weaker control measures are in place, other infectious diseases continue to pose a public health problem and coexist alongside COVID-19.

The present study has several limitations that should be acknowledged. Firstly, the study was limited to a single hospital and conducted over a relatively short period. However, it is worth noting that our findings are consistent with similar studies conducted in other cities in China. Secondly, our analysis focused solely on the observed outpatient data for Herpangina and HFMD on a daily basis. We did not account for the strong seasonal periodicity of these diseases, which is influenced by factors such as average air temperature and other meteorological variables. Therefore, the impact of meteorological factors before, during, and after the COVID-19 epidemic was not considered in our study.
Thirdly, the assessment of the impacts of COVID-19 on Herpangina and HFMD is based on the assumption that, in the absence of a COVID-19 outbreak, these diseases would exhibit fluctuations over time in a seasonal pattern. It is important to note that our study did not directly measure the effects of COVID-19 on the incidence of Herpangina or HFMD. Despite these limitations, we believe that this study provides valuable insights into the potential impact of COVID-19 in reducing the transmission of other infectious diseases among children, as well as the application of machine learning approaches to study Herpangina and HFMD. Future studies should aim to replicate this assessment of COVID-19’s impact on other infectious diseases affecting children in different cities or regions.

Conclusion

Based on the analysis of Herpangina and HFMD cases from 2018 to 2021, the machine learning model exhibited good performance in simulating and predicting the incidence of these diseases. During the COVID-19 period, there was a noticeable decrease in the number of cases for both Herpangina and HFMD. However, in the first half of 2021, the incidence of these diseases began to return to normal levels. These findings indicate that the outbreak of COVID-19 had a significant impact on the transmission of Herpangina and HFMD among children.

Conflict of Interest

The authors declare no conflict of interest.

References


**Citation:** Li D, Xiang D, Zhang SX, Zheng JX. The Effect of COVID-19 on Infectious Disease to Outpatient of Children: A Machine Learning Study. *SVOA Paediatrics* 2023, 2:4, 102-112.

**Copyright:** © 2023 All rights reserved by Zheng JX, 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.