

RESEARCH ARTICLE

Prediction of hand, foot, and mouth disease epidemics in Japan using a long short-term memory approach

Kazuhiro Yoshida^{1*}, Tsuguto Fujimoto², Masamichi Muramatsu¹, Hiroyuki Shimizu¹¹ Department of Virology II, National Institute of Infectious Diseases, Tokyo, Japan, ² Department of Fungal Infection, National Institute of Infectious Diseases, Tokyo, Japan* kyoshida12@gmail.com

Abstract

Hand, foot, and mouth disease (HFMD) is a common febrile illness caused by enteroviruses in the *Picornaviridae* family. The major symptoms of HFMD are fever and a vesicular rash on the hand, foot, or oral mucosa. Acute meningitis and encephalitis are observed in rare cases. HFMD epidemics occur annually in Japan, usually in the summer season. Relatively large-scale outbreaks have occurred every two years since 2011. In this study, the epidemic patterns of HFMD in Japan are predicted four weeks in advance using a deep learning method. The time-series data were analyzed by a long short-term memory (LSTM) approach called a Recurrent Neural Network. The LSTM model was trained on the numbers of weekly HFMD cases in each prefecture. These data are reported in the Infectious Diseases Weekly Report, which compiles the national surveillance data from web sites at the National Institute of Infectious Diseases, Japan, under the Infectious Diseases Control Law. Consequently, our trained LSTM model distinguishes between relatively large-scale and small-scale epidemics. The trained model predicted the HFMD epidemics in 2018 and 2019, indicating that the LSTM approach can estimate the future epidemic patterns of HFMD in Japan.

OPEN ACCESS

Citation: Yoshida K, Fujimoto T, Muramatsu M, Shimizu H (2022) Prediction of hand, foot, and mouth disease epidemics in Japan using a long short-term memory approach. PLoS ONE 17(7): e0271820. <https://doi.org/10.1371/journal.pone.0271820>

Editor: Sathishkumar V E, Hanyang University, REPUBLIC OF KOREA

Received: April 14, 2022

Accepted: July 8, 2022

Published: July 28, 2022

Copyright: © 2022 Yoshida et al. 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 author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting information files](#).

Funding: This study was financially supported in part by the Research Program on Emerging and Re-emerging Infectious Diseases from the Japan Agency for Medical Research and Development (AMED; <https://www.amed.go.jp/en/>), that a grant number is 21fk0108084j0003 and 22kf0108627. HS and TF received the support by AMED. The funder had no role in study design, data collection

1. Introduction

Enteroviruses belong to the *Picornaviridae* family and are categorized into 15 species (*Enterovirus A–L* and *Rhinovirus A–C*). More than 100 types of enteroviruses have been identified in humans. They are mainly transmitted by the fecal–oral route or respiratory infection route. A wide variety of clinical symptoms present in enterovirus infections, such as common cold-like illnesses, acute bronchitis, rashes, encephalitis, and paralysis [1–4].

Hand, foot, and mouth disease (HFMD) is a common febrile illness caused by enteroviruses. The main symptoms of HFMD are fever and a vesicular rash on the hand, foot, or oral mucosa, but acute meningitis and encephalitis can occur in rare cases [5, 6]. The most common HFMD pathogens are enterovirus A71 (EV-A71) and coxsackievirus A16 (CV-A16) [7]. In recent years, coxsackievirus A6 (CV-A6) has joined the conventional causative viruses (EV-A71 and CV-A16) as a major causative virus worldwide [8–16]. HFMD epidemics occur

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

annually in Japan, usually in the summer season and usually in children. Large-scale epidemics, mainly associated with CV-A6 [17], have occurred every two years since 2011. In Japan, the prevalence of HFMD and herpangina is monitored at pediatric sentinel sites and reported in the Infectious Disease Weekly Report (IDWR) of the National Institute of Infectious Diseases.

The estimate of the dynamics of HFMD epidemics was conducted in previous studies using the computational algorithm [7, 18–28]. The incidence of HFMD was estimated by using the past epidemics in China [18–24], Thailand [25], Singapore [26], and Japan [7, 27, 28]. In Japan, the time-series susceptible-infected-recovered model [7], the latitude-based approach [27], or the maximum entropy method and the least squares fitting method [28] were used for validating the availability of the models. These studies predicted the dynamic of HFMD epidemic throughout the year in Japan by each model.

To predict the occurrence of HFMD in Japan, we utilize the characteristics of deep learning. Deep learning algorithms are based on neural networks [29], imitation models that mimic the networks of neurons and synapses in the brain. In a neural network, neurons are described as nodes and synapses are represented by the strengths of the connections between nodes. The “nodes” and “synapses” are up-dated during the learning process, building a thinking pattern for problem solving. Among the various types of neural networks are recurrent neural networks (RNNs), convolutional neural networks, feed-forward neural networks, and restricted Boltzmann machines [29–32]. RNNs are suitable for processing time-series data. An RNN has a return value by which a new calculation can refer to the previous calculation process. The representative RNN model is long short-term memory (LSTM), which returns a value called “cell” and calculates the output referring to the past calculation process in the cell [33, 34].

In this study, we trained an LSTM model on HFMD case numbers extracted from the pediatric sentinel sites reported in the IDWR, Japan. Using the trained model, we predicted the epidemic patterns of HFMD in 2018 and 2019 before the actual prevalence of HFMD was revealed, and demonstrated that when the number of reported HFMD cases increases in a certain prefecture, the likelihood of large-scale HFMD outbreaks throughout Japan also increases by showing how the LSTM model learned the IDWR’s data as the characteristic of machine learning. This study contributed the prediction of HFMD outbreaks in Japan therefore the application for public health measures to minimize the large HFMD outbreaks nationwide was expected.

2. Materials and methods

2.1. Statistical data

We analyzed the weekly sentinel data of pediatric HFMD (in the category of number of patients) per prefecture recorded in the IDWR. The IDWR was obtained from websites of the National Institute of Infectious Disease (<https://www.niid.go.jp/niid/en/idwr-e.html>; SENTINEL-REPORTING DISEASES (WEEKLY), which records the number of cases in the current week by prefecture in the Surveillance Data).

2.2. LSTM

The LSTM model was programed as previously described [33, 34] and is a kind of RNN, which is available for using time series data. The LSTM model contains “cell” which is the structure holding the characteristic of the past data as numeric value. The process, that the input is calculated by using the cell and the result is outputted, is basic structure of LSTM model [33]. Additionally, the circuit changing the numeric value in the cell for reflecting the characteristic of the new trend to the output is called as forget gate [34]. In this study, we used the model,

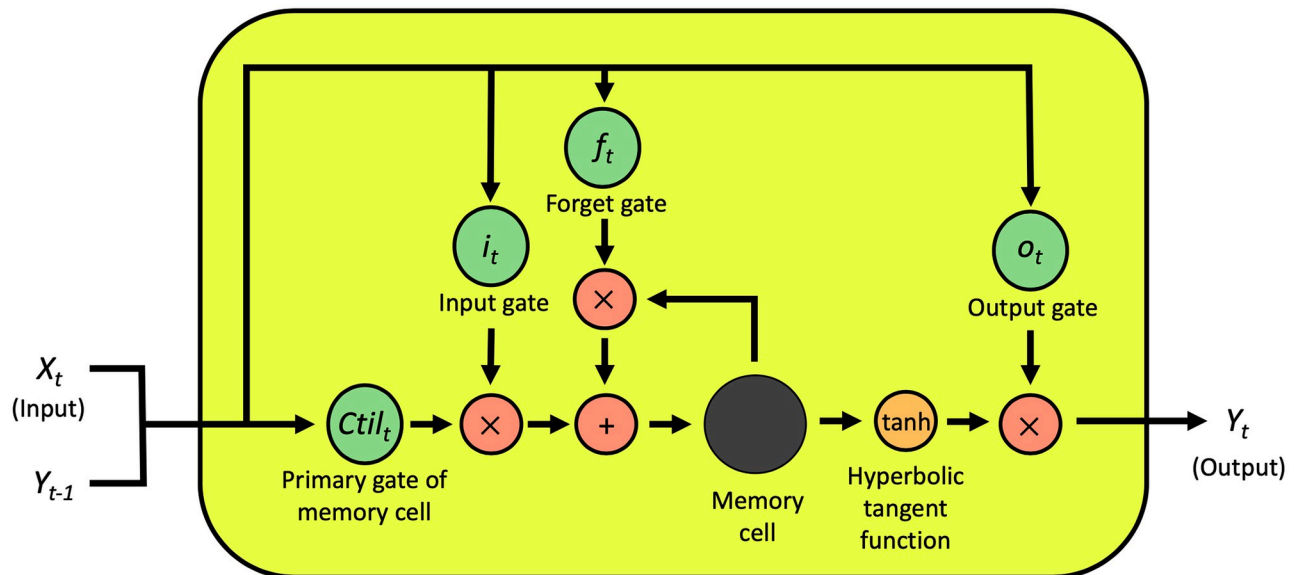


Fig 1. Summary of LSTM model. The formula of each step was described in 2.2. LSTM section.

<https://doi.org/10.1371/journal.pone.0271820.g001>

which the forget gate was added into the basic structure of LSTM (Fig 1). The forget gate (f_t), input gate (i_t), primary gate of memory cell ($Ctil_t$), memory cell (C_t), output gate (o_t), and output (y_t) of the LSTM model were respectively computed as follows:

$$f_t = \delta(W_f \cdot [y_{t-1}, x_t] + b_f), \quad (1)$$

$$i_t = \delta(W_i \cdot [y_{t-1}, x_t] + b_i), \quad (2)$$

$$Ctil_t = \tanh(W_c \cdot [y_{t-1}, x_t] + b_c), \quad (3)$$

$$C_t = f_t \times C_{t-1} + i_t \times Ctil_t, \quad (4)$$

$$o_t = \delta(W_o \cdot [y_{t-1}, x_t] + b_o), \quad (5)$$

$$y_t = o_t \times \tanh(C_t). \quad (6)$$

The symbol of x , W , and b were input, weight parameter, and threshold parameter, respectively. The input and the previous output were multiplied by weight parameter respectively, and threshold parameter was added in f_t , i_t , $Ctil_t$, or o_t step. The signals in f_t , i_t , $Ctil_t$, o_t , or y_t step were converted to an activated signal by an activation function (a sigmoidal function; σ or hyperbolic tangent function; \tanh). The input was normalized and standardized.

2.3. Confirming the appropriate input quantity by simulating the HFMD epidemic of 2015

The LSTM model was trained on pediatric sentinel weekly data by prefecture over the 1999–2014 period. For training, we inputted the epidemiological data of 1 to 6 consecutive weeks into the LSTM model. The model outputted the total number of patients in the following week and also compared the output value with the actual number of patients for training purposes.

We then inputted the data of the five weeks of 2015 to the trained model, which simulated the total number of patients in 2015 using its learned approach. The input was again normalized and standardized. The maximum and minimum output values were adjusted to the maximum and minimum number of patients in 2015, respectively.

2.4. Construction of the trained LSTM model presenting the epidemic pattern of HFMD in 2014–2017

When constructing the trained model, the LSTM model received five weeks of data and outputted the total number of patients four weeks later. The output was simulated from the pediatric sentinel data collected over the 1999–2013 period. Based on the 2014–2017 data, we then simulated the total number of HFMD patients until the epidemic peaked in week 31. During this period, we judged that the simulation results described the actual epidemic. The input was normalized and standardized. The maximum output value in 2017 and the minimum output value in 2014–2017 were adjusted to the maximum number of patients in 2017 and 0, respectively. The output of the simulation results could then be adjusted such that the number of patients was between 0 and the maximum output in the most recent year.

2.5. Prediction of the epidemic pattern in 2018 and 2019

The three independently trained LSTM models were named Trials 10, 26, and 46. The trained models were provided with consecutive five weeks of data and outputted the total number of HFMD patients four weeks later (at week 25). The input was normalized and standardized. The maximum output value in 2019 and the minimum output value in 2014–2019 were adjusted to the maximum number of patients in 2019 and 0, respectively. The output value of the simulation results was then adjusted such that the number of patients was between 0 and the maximum output value in the most recent year.

2.6. Contribution of each prefecture's epidemic against the nationwide HFMD outbreak of 2015

We replaced a number of HFMD patients per prefecture in the data of 2014 and 2015 aggregated in the IDWR reports, and inputted the modified data to the three independently trained models for simulating the total number of HFMD patients. The input data covered five weeks and the trained models outputted the total number of HFMD patients four weeks later. The input was normalized and standardized. We also summed the differences in the simulation results (weeks 9–31) obtained before and after modifying the input data. The actual numbers of HFMD patients by prefecture in 2015 were displayed as a heat map.

2.7. Root mean squared error (RMSE)

The accuracy of the simulation result was evaluated by the RMSE, defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - y_t)^2}. \quad (7)$$

In this expression, Y_t is the actual number of HFMD patients (standardized and normalized) and y_t denotes the simulation results. The RMSE defines the size of the prediction error in the output of the trained LSTM model.

2.8. Coefficient of determination (R^2)

The accuracy of the simulation result was also evaluated by the R^2 , defined as

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (8)$$

The y_t is the actual number of patients and \bar{y} is the average of the y_t . The \hat{y}_t denotes the simulation results. The scale of the y_t and the \hat{y}_t were matched for the calculation of R^2 . The R^2 defines how the simulation results can be explained by the actual number of patients.

3. Results

3.1. Simulation of HFMD epidemic pattern in 2015 for confirming the appropriate input quantity

To confirm that the input quantity maximized the simulation performance, we simulated a number of HFMD patients in 2015 based on the IDWR reports of 1999–2014 (Fig 2). The LSTM model was trained on the weekly pediatric sentinel data by prefecture during the 1999–2014 period, extracted from the IDWR in Japan. More specifically, we inputted the number of HFMD patients over 1–6 consecutive weeks into the LSTM model. The model then outputted the total number of patients in the following week. The outputs of the LSTM model fairly

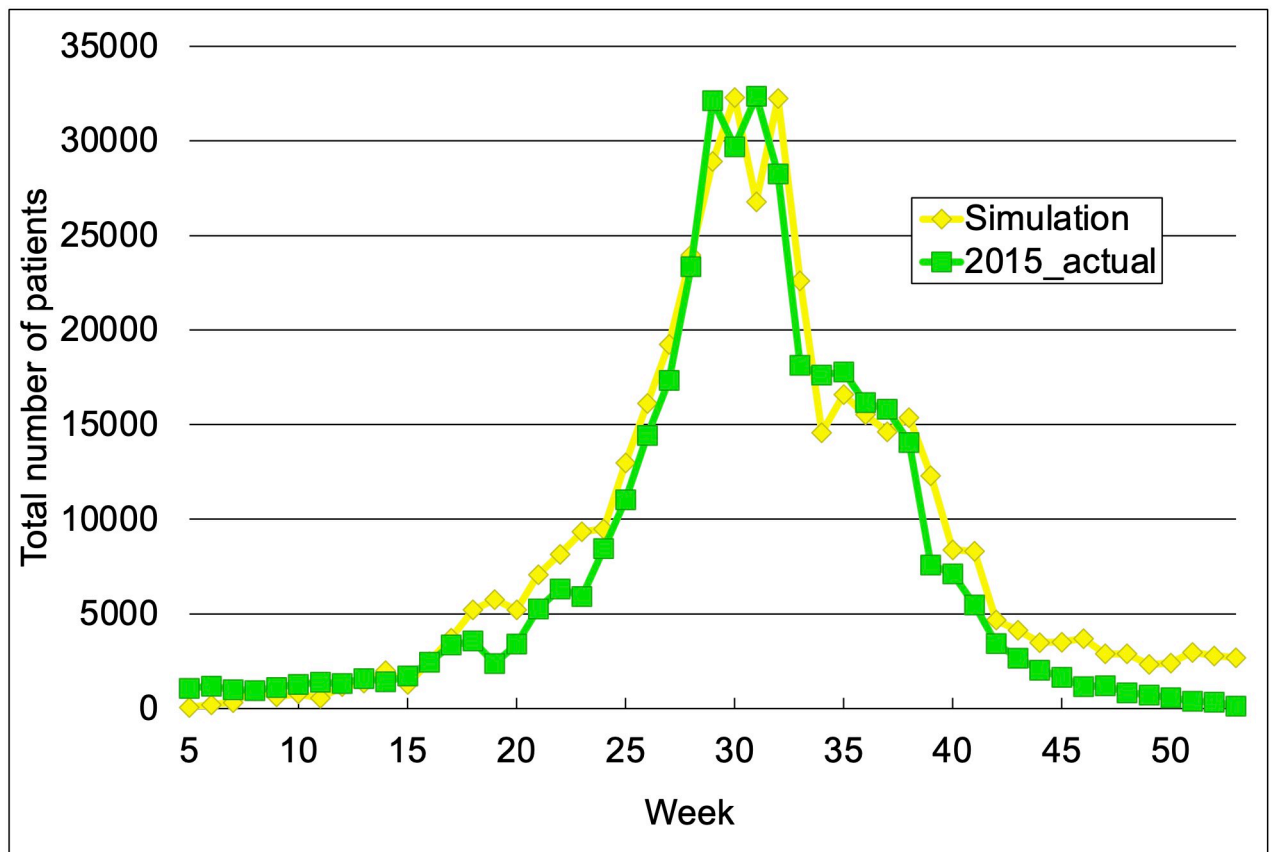


Fig 2. Simulation result of HFMD epidemic in 2015 (using 5 weeks of data as input). The maximum and minimum output values were adjusted to the maximum and minimum number of patients in 2015, respectively.

<https://doi.org/10.1371/journal.pone.0271820.g002>

matched the actual number of HFMD cases. To estimate the epidemic pattern in 2015, we inputted the data of 1–6 weeks in 2015 to the trained LSTM model. The model then simulated the total number of HFMD patients in 2015 using its learned method.

The simulation results almost matched the actual prevalence of HFMD cases. Among the 1–6 weeks of data input, five weeks of data achieved the best correspondence between the simulated and actual results ($RMSE = 8.53 \times 10^{-7}$, $R^2 = 0.944$, and the evaluated period was week 5–53). Concretely, the simulated number of patients increased from around week 15, peaked in weeks 30–32, and decreased from week 33.

3.2. Simulation of HFMD epidemic 2–5 weeks after the data input period

To verify that the LSTM model could simulate the near-future increase in the number of HFMD patients, we simulated the number of HFMD cases at 2–5 weeks after the data input period in 2015 (see [S1 Fig](#)). The model was trained on five weeks of consecutive data over the 1999–2014 period as described for [Fig 2](#) in the previous subsection. The model outputted the total number of patients 2–5 weeks later. In the 2015 simulation, the number of patients 2–5 weeks later was computed by the learned approach. The simulation results one week later ([Fig 2](#)) and 2–4 weeks later better matched the actual results than the simulation result of five weeks later (the RMSEs and the R^2 are listed in [S1](#) and [S2](#) Tables). Therefore, the model could accurately simulate the number of HFMD patients up to four weeks after the data input period.

3.3. Construction and validation of the trained LSTM model presenting the epidemic patterns of 2014–2017

To predict the epidemic patterns of HFMD in 2018 and 2019, we trained the LSTM model to simulate the epidemic patterns of HFMD over the 2014–2017 period ([Fig 3](#)). During training, the model was inputted with five consecutive weeks of data over the 1999–2013 period and outputted the total number of patients four weeks later. To judge whether the model can describe the true epidemic, we simulated the total number of HFMD patients during 2014–2017 until the epidemic peaked in week 31. Tables [1](#) and [2](#) gave the RMSEs and the R^2 of the total numbers of HFMD patients from week 1 to week 31 in the years 2014–2017 predicted in the best-performance trials of the three independently trained LSTM models (Trials 10, 26, and 46). The epidemic curves described by the trained models were similar to those of the actual epidemic. The simulation results of 2015 and 2017 showed sharp increases from around week 25, which were absent in the simulations of 2014 and 2016. Therefore, the epidemic trends were simulated one month before the epidemic peak.

3.4. Prediction of epidemic patterns of HFMD in 2018 and 2019

To check whether the model can predict a yet-unreported HFMD epidemic, we predicted the epidemic patterns of HFMD in 2018 and 2019 using the three independently trained LSTM models ([Fig 4](#)). When predicting the 2018 and 2019 epidemics, we employed the method used for training the LSTM model. The three trained models predicted the epidemic pattern of 2018, which was slightly larger in scale than the 2014 and 2016 epidemics. The 2018 predictions were consistent with the actual HFMD patient numbers until week 25 ([Fig 5](#)). The three trained models also predicted the epidemic pattern of 2019, which was identical in scale to the 2015 and 2017 epidemics. The scales of the predicted and actual case numbers in 2019 were well matched. The RMSEs and the R^2 of the 2018 and 2019 predictions are listed in Tables [3](#) and [4](#), respectively. Judging from these results, the trained LSTM model could predict the epidemic patterns of HFMD in 2018 and 2019.

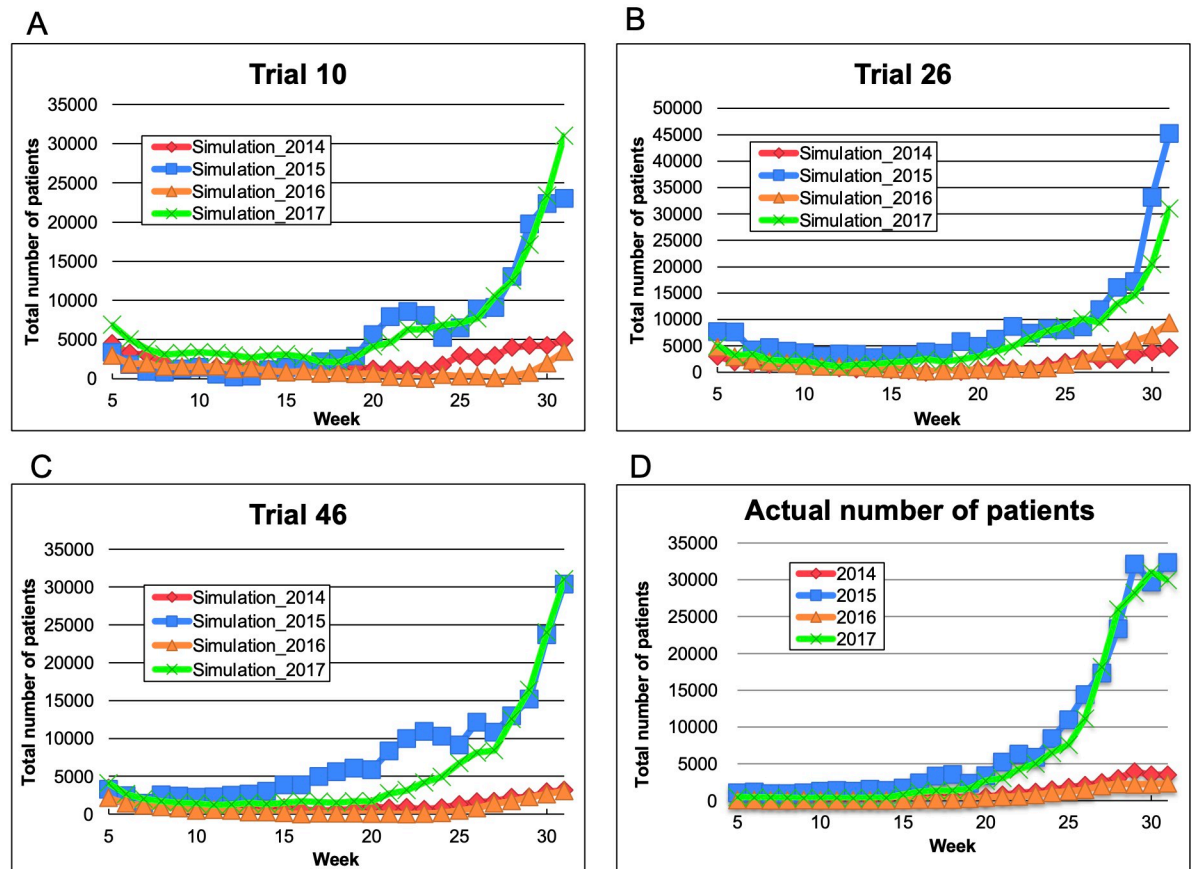


Fig 3. Epidemic patterns in 2014–2017 simulated by the three independently constructed models (obtained in the top-performing trials). The maximum output value in 2017 and the minimum output value in 2014–2017 were adjusted to the maximum number of patients in 2017 and 0, respectively: (A) trial 10; (B) trial 26; (C) trial 46; (D) actual number of HFMD patients. Diamonds: 2014; squares: 2015; triangles: 2016; crosses: 2017.

<https://doi.org/10.1371/journal.pone.0271820.g003>

3.5. Influence size of each prefecture’s epidemic against the nationwide HFMD outbreak in Japan

To verify the size of the influence of each prefecture’s epidemic against the nationwide outbreak, we investigated how the LSTM model, which contained the machine learning algorithm, learned the number of HFMD patients in each prefecture. Here, we modified the numbers of HFMD patients in each prefecture in 2014 (a small-scale season) and 2015 (a large-scale season) obtained from the IDWR reports before input to the models, and inputted the modified data to the three independently trained models. The total simulated numbers of

Table 1. RMSEs of the simulation over the 2014–2017 period ^a.

	2014	2015	2016	2017
Trial 10	5.62×10^{-7}	5.38×10^{-6}	7.36×10^{-7}	2.46×10^{-6}
Trial 26	6.00×10^{-7}	6.53×10^{-6}	5.22×10^{-7}	6.97×10^{-6}
Trial 46	9.11×10^{-7}	7.42×10^{-6}	4.56×10^{-7}	7.16×10^{-6}

^a Calculated using the summed differences in weeks 5–52 or weeks 5–53

<https://doi.org/10.1371/journal.pone.0271820.t001>

Table 2. R^2 of the simulation over the 2014–2017 period ^a.

	2014	2015	2016	2017
Trial 10	-0.493	0.787	-1.389	0.785
Trial 26	0.408	0.741	-6.611	0.776
Trial 46	0.307	0.779	0.257	0.811

^a Calculated using the summed differences in weeks 5–31

<https://doi.org/10.1371/journal.pone.0271820.t002>

HFMD patients based on the modified 2014 and 2015 data were compared with the simulation results (weeks 9–31) based on the unmodified data. After summing the differences between the results obtained before and after modifying the input data, we found that among 47 prefectures, replacing the number of patients in Osaka Prefecture maximized the variability of the simulation. The results of other urban areas (Tokyo, Aichi, and Fukuoka) were less variable than those of Osaka (Figs 6 and 7). Moreover, the patient numbers began increasing earlier in Osaka Prefecture than in other prefectures during the actual HFMD outbreak of 2015 (Fig 8).

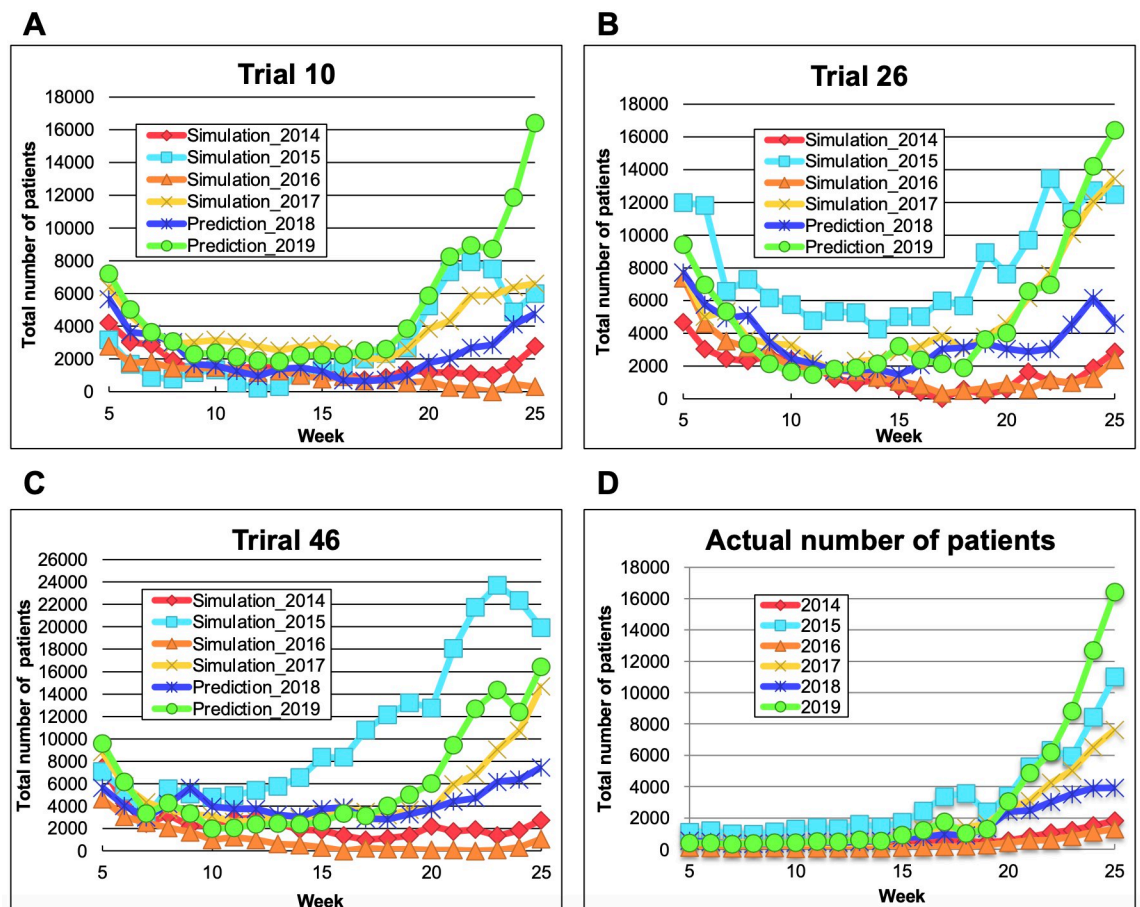


Fig 4. Epidemic patterns in 2018 and 2019 predicted by the three independently trained models (in trials 10, 26, and 46). (A–C): The maximum output value in 2019 and the minimum output value in 2014–2019 were adjusted to the maximum number of patients in 2019 and 0, respectively; (D) Actual numbers of patients until week 25.

<https://doi.org/10.1371/journal.pone.0271820.g004>

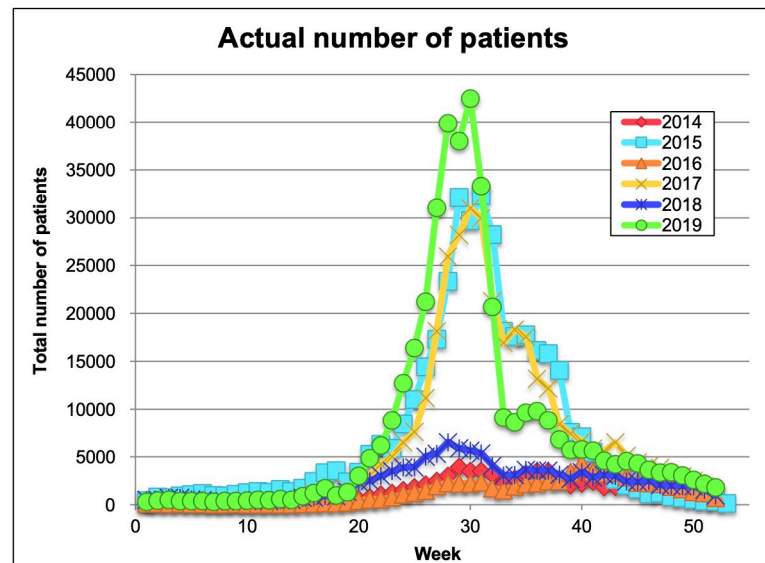


Fig 5. Actual epidemic patterns in 2014–2019. Actual numbers of patients in weeks 1–52 or 1–53.

<https://doi.org/10.1371/journal.pone.0271820.g005>

This result indicates that the machine learning algorithm detected the increasing pattern in Osaka as a characteristic feature of large-scale epidemics.

4. Discussion

We trained LSTM models to simulate the epidemic patterns of HFMD based on IDWR data in Japan. The models were trained on the HFMD epidemic data of 1999–2013 and simulated the epidemic patterns in 2014–2017 before they were revealed. Furthermore, the models predicted the HFMD epidemic patterns of 2018 and 2019.

HFMD cases in Japan are caused by various types of enteroviruses. The major causative agents were EV-A71 and CV-A16, but the number of cases attributed to CV-A6 has increased since 2011, implying an epidemiological change around 2010 [35]. CV-A6 mainly contributed the large-scale epidemics in 2011, 2013, 2015, and 2017. In a previous study, the dynamics of numbers of patients infected by different type of enteroviruses were explored [36], while the present study aimed to predict an HFMD outbreak, regardless of enterovirus type. Additionally, the learning step of LSTM model included the period when EV-A71, CV-A16, or CV-A6 was major pathogen (i.e., the 1999–2013 season), therefore the learned LSTM model could predict each number of reported HFMD patients, who were infected with enteroviruses including the types isolated until 2013. As the infected patients and epidemic pattern were changed by reflecting the epidemic situation of each type of enteroviruses, the epidemic situations were

Table 3. RMSEs of the 2018 and 2019 predictions ^a.

	2018	2019
Trial 10	2.94×10^{-7}	6.66×10^{-7}
Trial 26	2.09×10^{-7}	8.53×10^{-7}
Trial 46	1.43×10^{-7}	2.18×10^{-6}

^a Calculated using the summed differences in weeks 6–25

<https://doi.org/10.1371/journal.pone.0271820.t003>

Table 4. R² of the 2018 and 2019 predictions ^a.

	2018	2019
Trial 10	-0.109	0.758
Trial 26	-2.632	0.715
Trial 46	-4.454	0.456

^a Calculated using the summed differences in weeks 6–25

<https://doi.org/10.1371/journal.pone.0271820.t004>

also reflected into reported HFMD patients and the model. Additionally, it was considered that the LSTM model learned the change of the major pathogen with CV-A6 has increased since 2011 by forget gate in the model. By using the learned LSTM model, the HFMD outbreak in 2019 was showed at Fig 4.

To understand the epidemic patterns of HFMD, we employed LSTM because this algorithm has been widely applied in predictions of influenza, chicken pox, scarlet fever, and malaria [37–40]. It was indicating that use of LSTM model was appropriate for predicting HFMD epidemic pattern.

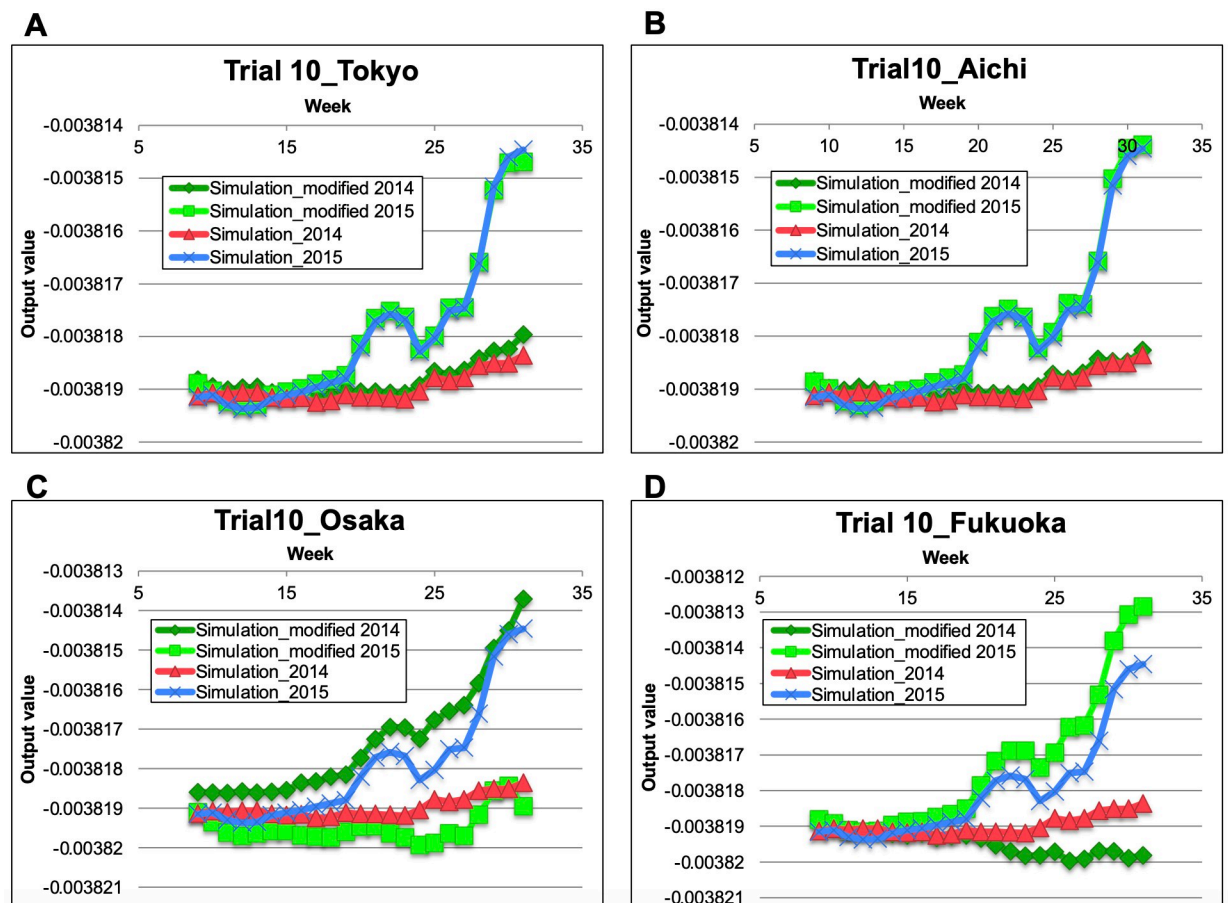


Fig 6. Example of simulation results of each prefecture’s epidemic by using the modified data. The three independently trained models were inputted with the pediatric sentinel data of 2014 and 2015 after replacing some data in Tokyo (A), Aichi (B), Osaka (C), and Fukuoka (D). Diamonds: modified 2014 data, squares: modified 2015 data, triangles: original 2014 data, crosses: original 2015 data.

<https://doi.org/10.1371/journal.pone.0271820.g006>

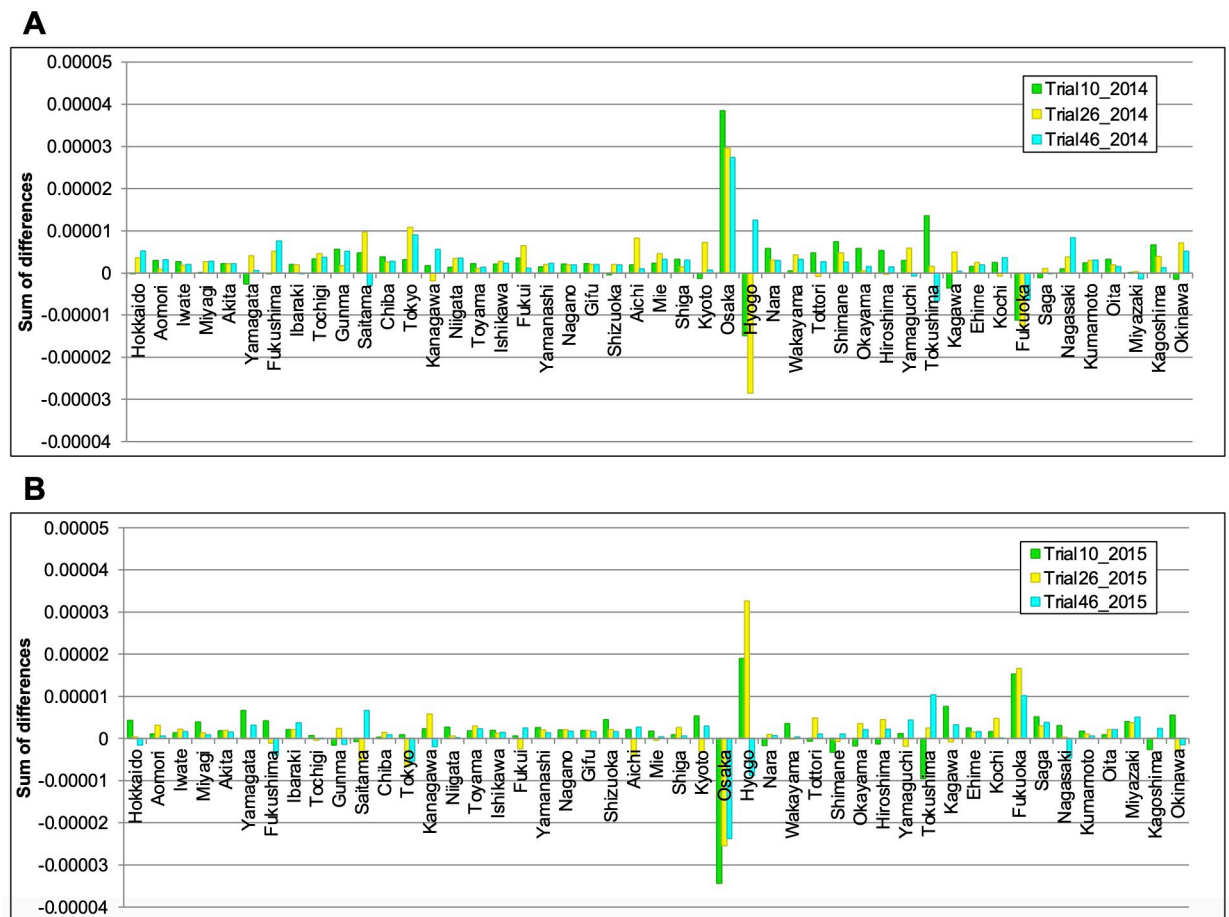


Fig 7. Size of influence of each prefecture's epidemic against the HFMD outbreak of 2015. (A) and (B): Summed differences (weeks 9–31) between the simulation results before and after modifying the 2014 and 2015 data, respectively.

<https://doi.org/10.1371/journal.pone.0271820.g007>

Recently, LSTM-based simulations of HFMD epidemics have also been reported in China [21–23]. These reports applied an LSTM model with a forget gate and connected cell conditions [21], or with a forget gate [22, 23], the availability of these models was showed. In this study, we used the same model at Gu's and Zhang's reports [22, 23], and the LSTM model showed similar performance when the epidemiological data of HFMD in Japan was used. Moreover, HFMD epidemics in Beijing have also been predicted by an algorithm with a gated recurrent unit (GRU) [24], suggesting that the algorithm using GRU was available as same as the LSTM algorithm for predicting the HFMD epidemic. Whereas the yearly patient numbers have remained fairly stable in China, large and small-scale epidemics in Japan have repeated at two-yearly intervals since 2011. Therefore, by predicting the epidemic scales of HFMD using the LSTM model based on IDWR data, we could possibly implement preventive measures at pediatric hospitals, clinics, and communities against large HFMD outbreaks.

The machine learning analysis revealed a unique feature of the Japanese HFMD data, namely, that the HFMD patients in Osaka Prefecture most affected the computational simulation results. With a conventional method, new characteristic in statistical data was found through human, although the LSTM model investigated the new characteristic automatically using “auto-update” learning algorithm. Moreover, we found out the characteristic of Osaka's

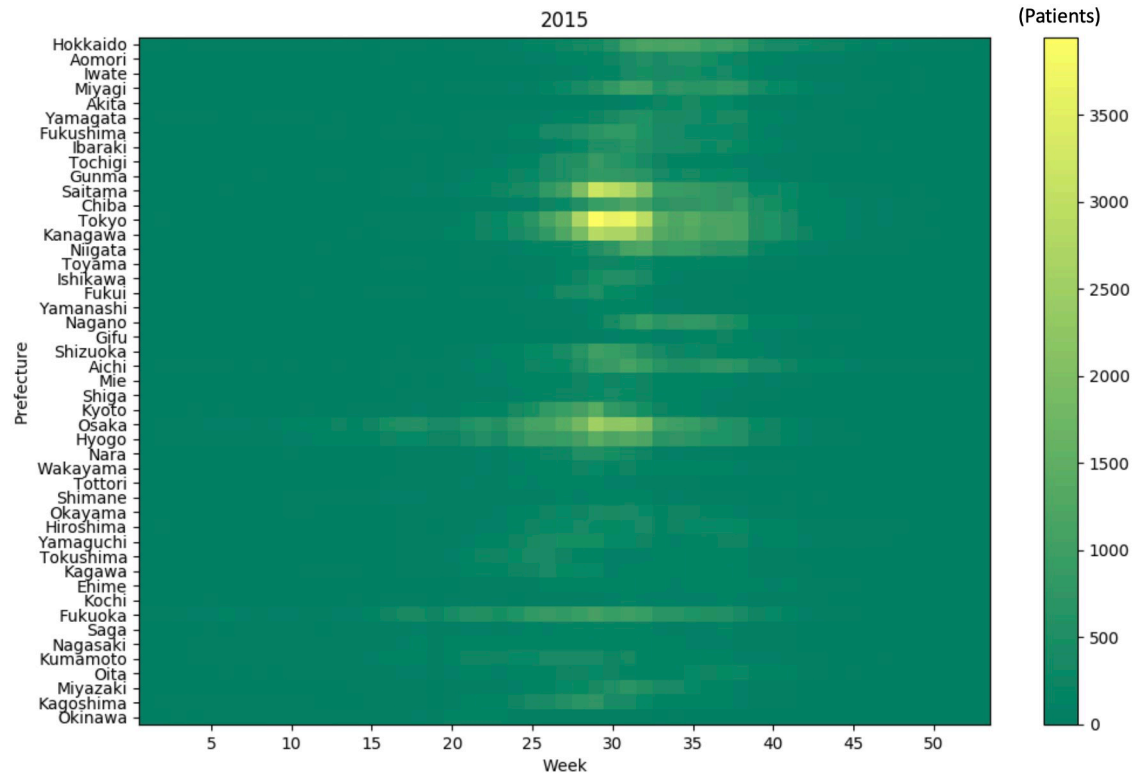


Fig 8. Heat map of the actual HFMD patient numbers in 2015. The actual patient numbers by prefecture were displayed.

<https://doi.org/10.1371/journal.pone.0271820.g008>

data in the process of using the LSTM model. It was indicating that new epidemiological characteristic, which could not be found out through human, could be sought by unrepresented approach in epidemiology.

Previous studies showed the prediction of HFMD epidemics at some prefectures in Japan, using the latitude-based approach [27], or the maximum entropy method and the least squares fitting method [28]. These studies compared the results of some prefectures calculated under the same conditions, while in this study, we compared the actual and hypothetical HFMD epidemics, which the number of patients at each prefecture was changed, of the large-scale epidemic. The variability of the simulation was maximum when the number of patients of Osaka prefecture in 2014 and 2015 were replaced.

In this study, the characteristic of Osaka's data was revealed only by analyzing the weekly recorded numbers of HFMD patients in each prefecture of Japan, and the LSTM algorithm didn't use the others factor such as Osaka's geographical location, population, temperature, or humidity. Therefore, the relationship between the characteristic of Osaka's data and the factors, except the number of HFMD patients, were indefinite. The fact that Osaka is geographically located in the center of Japan and is the third most populous prefecture in Japan may have had an impact.

In Figs 2–4, and S1 Fig, we compared the actual number of patients and the results estimated by LSTM model. By using the data in Figs 2–4, and S1 Fig, RMSE and R^2 were calculated for evaluating the accuracy of the simulation results. The values of R^2 from 2014 and 2016's simulation in Fig 3 were minus because it was suggested that the small-scale epidemic in 2014 and 2016 provided the trend, that the residual sum of squares was larger than the total

sum of squares. Additionally, we compared relatively the estimated results at week 25 in Fig 4 for predicting the scale of epidemics in 2018 and 2019.

5. Conclusion

In this study, we predicted HFMD outbreaks in Japan using LSTM algorithm. We showed the dynamics of HFMD epidemic 4 weeks later from the numbers of weekly HFMD cases reported IDWR of the National Institute of Infectious Diseases. We predicted HFMD epidemic in 2018 and 2019, and the results of the prediction and the actual reported patients were almost match. Additionally, we found the increasing pattern in Osaka was the characteristic feature of large-scale epidemics.

This study is available for the control of HFMD outbreaks.

Supporting information

S1 Method. Simulation of HFMD patients 2–5 weeks later.

(DOCX)

S1 Fig. Simulation of HFMD patients 2–5 weeks later in 2015. The LSTM model was trained on the same data as Fig 2. The model was trained on the data of consecutive five weeks and outputted the total number of HFMD patients 2–5 weeks later. The projected numbers of patients 2–5 weeks later were simulated by the same training approach (one week later: interval 0 in Fig 2; two weeks later: interval 1; three weeks later: interval 2; four weeks later: interval 3; five weeks later: interval 4). The input was normalized and standardized. The maximum and minimum output values at interval 0 were adjusted to the maximum and minimum number of patients in 2015, respectively.

(TIF)

S1 Table. RMSEs of the simulation results 2–5 weeks later.

(DOCX)

S2 Table. R² of simulations 2–5 weeks later.

(DOCX)

Acknowledgments

We thank the staff at prefectural and municipal public health institutes for conducting continuous surveillance for public health purposes in Japan. For the surveillance data, we also appreciate the efforts provided by the NIID Infectious Disease Surveillance Center's members. The authors would like to thank Enago (www.enago.jp) for the English language review.

Author Contributions

Conceptualization: Kazuhiro Yoshida.

Data curation: Kazuhiro Yoshida.

Formal analysis: Kazuhiro Yoshida.

Funding acquisition: Tsuguto Fujimoto, Hiroyuki Shimizu.

Investigation: Kazuhiro Yoshida.

Methodology: Kazuhiro Yoshida.

Project administration: Kazuhiro Yoshida.

Resources: Kazuhiro Yoshida.

Software: Kazuhiro Yoshida.

Supervision: Masamichi Muramatsu, Hiroyuki Shimizu.

Validation: Tsuguto Fujimoto, Hiroyuki Shimizu.

Visualization: Kazuhiro Yoshida.

Writing – original draft: Kazuhiro Yoshida.

Writing – review & editing: Tsuguto Fujimoto, Masamichi Muramatsu, Hiroyuki Shimizu.

References

1. Baggen J, Thibaut HJ, Strating JRPM, van Kuppeveld FJM. The life cycle of non-polio enteroviruses and how to target it. *Nat Rev Microbiol*. 2018; 16: 368–381. <https://doi.org/10.1038/s41579-018-0005-4> PMID: 29626210
2. Pons-Salort M, Parker EPK, Grassly NC. The epidemiology of non-polio enteroviruses: recent advances and outstanding questions. *Curr Opin Infect Dis*. 2015; 28: 479–87. <https://doi.org/10.1097/QCO.000000000000187> PMID: 26203854
3. Sinclair W, Omar M. Enterovirus. 2021. <https://www.ncbi.nlm.nih.gov/books/NBK562330/>
4. The Picornavirus Pages. The Pirbright Institute, UK. https://www.picornaviridae.com/sg3_ensavirinae/enterovirus/enterovirus.htm
5. Ventarola D, Bordone L, Silverberg N. Update on hand-foot-and-mouth disease. *Clin Dermatol*. 2015; 33: 340–6. <https://doi.org/10.1016/j.clindermatol.2014.12.011> PMID: 25889136
6. Repass GL, Palmer WC, Stancampiano FF. Hand, foot, and mouth disease: Identifying and managing an acute viral syndrome. *Cleve Clin J Med*. 2014; 81: 537–43. <https://doi.org/10.3949/ccjm.81a.13132> PMID: 25183845
7. Takahashi S, Metcalf CJE, Arima Y, Fujimoto T, Shimizu H, Doorn HRV, et al. Epidemic dynamics, interactions and predictability of enteroviruses associated with hand, foot and mouth disease in Japan. *J R Soc Interface*. 2018; 15: 20180507. <https://doi.org/10.1098/rsif.2018.0507> PMID: 30209044
8. Solomon T, Lewthwaite P, Perera D, Cardoso MJ, McMinn P, Ooi MH. Virology, epidemiology, pathogenesis, and control of enterovirus 71. *Lancet Infect Dis*. 2010; 10: 778–790. [https://doi.org/10.1016/S1473-3099\(10\)70194-8](https://doi.org/10.1016/S1473-3099(10)70194-8) PMID: 20961813
9. Chang YK, Chen KH, Chen KT. Hand, foot and mouth disease and herpangina caused by enterovirus A71 infections: a review of enterovirus A71 molecular epidemiology, pathogenesis, and current vaccine development. *Rev Inst Med Trop Sao Paulo*. 2018; 60: e70. <https://doi.org/10.1590/S1678-9946201860070> PMID: 30427405
10. Chen X, Tan X, Li J, Jin Y, Gong L, Hong M, et al. Molecular epidemiology of coxsackievirus A16: intra-type and prevalent intertype recombination identified. *PLoS ONE*. 2013; 10: e82861. <https://doi.org/10.1371/journal.pone.0082861> PMID: 24340064
11. Mizuta K, Abiko C, Aoki Y, Ikeda T, Matsuzaki Y, Hongo S, et al. Molecular epidemiology of Coxsackievirus A16 strains isolated from children in Yamagata, Japan between 1988 and 2011. *Microbiol Immunol*. 2013; 57: 400–5. <https://doi.org/10.1111/1348-0421.12041> PMID: 23668614
12. Di B, Zhang Y, Xie H, Li X, Chen C, Ding P, et al. Circulation of Coxsackievirus A6 in hand-foot-mouth disease in Guangzhou, 2010–2012. *Virology*. 2014; 11: 157. <https://doi.org/10.1186/1743-422X-11-157> PMID: 25178398
13. Kanbayashi D, Kaida A, Hirai Y, Yamamoto SP, Fujimori R, Okada M, et al. An epidemic of hand, foot, and mouth disease caused by Coxsackievirus A6 in Osaka City, Japan, in 2017. *Jpn J Infect Dis*. 2019; 72: 334–336. <https://doi.org/10.7883/yoken.JJID.2018.532> PMID: 31061358
14. Kimmis BD, Downing C, Tying S. Hand-foot-and-mouth disease caused by Coxsackievirus A6 on the rise. *Cutis*. 2018; 102: 353–356. Available from: <https://cdn.mdedge.com/files/s3fs-public/Document/November-2018/CT102005353.PDF> PMID: 30566537
15. Koh WM, Bogich T, Siegel K, Jin J, Chong EY, Tan CY, et al. The epidemiology of hand, foot and mouth disease in Asia: a systematic review and analysis. *Pediatr Infect Dis J*. 2016; 35: e285–300. <https://doi.org/10.1097/INF.0000000000001242> PMID: 27273688
16. Xing W, Liao Q, Viboud C, Zhang J, Sun J, Wu JT, et al. Epidemiological characteristics of hand-foot-and-mouth disease in China, 2008–2012. *Lancet Infect Dis*. 2014; 14: 308–318.

17. Gonzalez G, Carr MJ, Kobayashi M, Hanaoka N, Fujimoto T. Enterovirus-associated hand-foot and mouth disease and neurological complications in Japan and the rest of the world. *Int J Mol Sci*. 2019; 20: 5201. <https://doi.org/10.3390/ijms20205201> PMID: 31635198
18. Zhuang ZC, Kou ZQ, Bai YJ, Cong X, Wang LH, Li C, et al. Epidemiological research on hand, foot, and mouth disease in mainland China. *Viruses*. 2015; 7: 6400–11. <https://doi.org/10.3390/v7122947> PMID: 26690202
19. Bo YC, Song C, Wang JF, Li XW. Using an autologistic regression model to identify spatial risk factors and spatial risk patterns of hand, foot and mouth disease (HFMD) in mainland China. *BMC Public Health*. 2014; 14: 358. <https://doi.org/10.1186/1471-2458-14-358> PMID: 24731248
20. Fu T, Chen T, Dong ZB, Luo SY, Miao Z, Song XP, et al. Development and comparison of forecast models of hand-foot-mouth disease with meteorological factors. *Sci Rep*. 2019; 9: 15691. <https://doi.org/10.1038/s41598-019-52044-5> PMID: 31666565
21. Wang Y, Xu C, Zhang S, Yang L, Wang Z, Zhu Y, et al. Development and evaluation of a deep learning approach for modeling seasonality and trends in hand-foot-mouth disease incidence in mainland China. *Sci Rep*. 2019; 9: 8046. <https://doi.org/10.1038/s41598-019-44469-9> PMID: 31142826
22. Gu J, Liang L, Song H, Kong Y, Ma R, Hou Y, et al. A method for hand-foot-mouth disease prediction using GeoDetector and LSTM model in Guangxi, China. *Sci Rep*. 2019; 9: 17928. <https://doi.org/10.1038/s41598-019-54495-2> PMID: 31784625
23. Zhang R, Guo Z, Meng Y, Wang S, Li S, Niu R, et al. Comparison of ARIMA and LSTM in forecasting the incidence of HFMD combined and uncombined with exogenous meteorological variables in Ningbo, China. *Int J Environ Res Public Health*. 2021; 18: 6174. <https://doi.org/10.3390/ijerph18116174> PMID: 34200378
24. Wang Y, Cao Z, Zeng D, Wang X, Wang Q. Using deep learning to predict the hand-foot-and-mouth disease of enterovirus A71 subtype in Beijing from 2011 to 2018. *Sci Rep*. 2020; 10: 12201. <https://doi.org/10.1038/s41598-020-68840-3> PMID: 32699245
25. Verma S, Razzaque MA, Sangtongdee U, Arpnanondt C, Tassaneetrithep B, Arthan D, et al. Hand, foot, and mouth disease in Thailand: a comprehensive modelling of epidemic dynamics. *Comput Math Methods Med*. 2021; 2021: 6697522. <https://doi.org/10.1155/2021/6697522> PMID: 33747118
26. Chen Y, Chu CW, Chen MIC, Cook AR. The utility of LASSO-based models for real time forecasts of endemic infectious diseases: a cross country comparison. *J Biomed Inform*. 2018; 81: 16–30. <https://doi.org/10.1016/j.jbi.2018.02.014> PMID: 29496631
27. Tang JH, Chan TC, Shigematsu M, Hwang JS. Latitude-based approach for detecting aberrations of hand, foot, and mouth disease epidemics. *BMC Med Inform Decis Mak*. 2015; 15: 113. <https://doi.org/10.1186/s12911-015-0236-5> PMID: 26703896
28. Sumi A, Toyoda S, Kanou K, Fujimoto T, Mise K, Kohei Y, et al. Association between meteorological factors and reported cases of hand, foot, and mouth disease from 2000 to 2015 in Japan. *Epidemiol Infect*. 2017; 145: 2896–2911. <https://doi.org/10.1017/S0950268817001820> PMID: 28826420
29. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; 521: 436–44. <https://doi.org/10.1038/nature14539> PMID: 26017442
30. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012; 25. Available from: <https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>
31. Svozil D, Kvasnicka V, Pospichal J. Introduction to multi-layer feed-forward neural networks. *Chemometr Intell Lab Syst*. 1997; 39: 43–62. [https://doi.org/10.1016/S0169-7439\(97\)00061-0](https://doi.org/10.1016/S0169-7439(97)00061-0)
32. Sutskever I, Hinton G, Taylor G. The recurrent temporal restricted boltzmann machine. *Adv Neural Inf Process Syst*. 2008; 21. Available from: <https://www.semanticscholar.org/paper/The-Recurrent-Temporal-Restricted-Boltzmann-Machine-Sutskever-Hinton/0228810a988f6b8f06337e14f564e2fd3f6e1056>
33. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997; 9: 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735> PMID: 9377276
34. Gers FA, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM. *Neural Comput*. 2000; 12: 2451–2471. <https://doi.org/10.1162/089976600300015015> PMID: 11032042
35. Hand, foot, and mouth disease and herpangina, 2007 to September 2017 (week 38), Japan. *IASR*. 2017; 38: 191–193. Available from: <https://www.niid.go.jp/niid/en/iasr-vol38-e/865-iasr/7617-452te.html/>
36. Pons-Salort M, Grassly NC. Serotype-specific immunity explains the incidence of diseases caused by human enteroviruses. *Science*. 2018; 361: 800–803. <https://doi.org/10.1126/science.aat6777> PMID: 30139872
37. Zhang J, Nawata K. A comparative study on predicting influenza outbreaks. *Biosci Trends*. 2017; 11: 533–541. <https://doi.org/10.5582/bst.2017.01257> PMID: 29070762

38. Zhu X, Fu B, Yang Y, Ma Y, Hao J, Chen S, et al. Attention-based recurrent neural network for influenza epidemic prediction. *BMC Bioinformatics*. 2019; 20: 575. <https://doi.org/10.1186/s12859-019-3131-8> PMID: [31760945](https://pubmed.ncbi.nlm.nih.gov/31760945/)
39. Zhang J, Nawata K. Multi-step prediction for influenza outbreak by an adjusted long short-term memory. *Epidemiol Infect*. 2018; 146: 809–816. <https://doi.org/10.1017/S0950268818000705> PMID: [29606177](https://pubmed.ncbi.nlm.nih.gov/29606177/)
40. Chae S, Kwon S, Lee D. Predicting infectious disease using deep learning and big data. *Int J Environ Res Public Health*. 2018; 15: 1596. <https://doi.org/10.3390/ijerph15081596> PMID: [30060525](https://pubmed.ncbi.nlm.nih.gov/30060525/)