

Spatiotemporal dengue fever hotspots associated with climatic factors in Taiwan including outbreak predictions based on machine-learning

Sumiko Anno,¹ Takeshi Hara,² Hiroki Kai,³ Ming-An Lee,⁴ Yi Chang,⁵ Kei Oyoshi,⁶ Yousei Mizukami,⁶ Takeo Tadono⁶

¹Graduate School of Global Environmental Studies, Sophia University, Tokyo, Japan; ²Department of Engineering, Gifu University, Gifu, Japan; ³Department of Research Development, Remote Sensing Technology Centre of Japan, Tokyo, Japan; ⁴Centre of Excellence for Ocean Engineering and Department of Environmental Biology and Fisheries Science, National Taiwan Ocean University, Keelung, Taiwan; ⁵Department of Hydraulic and Ocean Engineering and Institute of Ocean Technology and Marine Affairs, National Cheng Kung University, Tainan, Taiwan; ⁶Earth Observation Research Centre, Japan Aerospace Exploration Agency, Ibaraki, Japan

Correspondence: Sumiko Anno, Graduate School of Global Environmental Studies, Sophia University, 1520 Bldg No. 2, 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan.
Tel. +81.3.3238.4648 - Fax: +81.3.3238.4439.
E-mail: sumiko_anno@sophia.ac.jp

Acknowledgements: the authors would like to thank the Japan Aerospace Exploration Agency (JAXA) for providing the remote sensing data.

Key words: Dengue fever; Spatiotemporal hotspot analysis; Remote sensing; Machine-learning; Taiwan.

See online Appendix for additional figures.

Contributions: SA, data analysis and manuscript writing; TH, data analysis; MAL, dengue data collection; HK, YC, KO, YM, TT, remote sensing data collection and processing.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: the work was supported by a Japan Aerospace Exploration Agency grant, PI No. RA1R803.

Conference presentation: part of this paper was presented at the 4th International Conference on Environmental Science & Technology 2018, March 29-31, 2018, Vienna, Austria.

Received for publication: 3 March 2019.

Revision received: 16 April 2019.

Accepted for publication: 4 June 2019.

©Copyright: the Author(s), 2019
Licensee PAGEPress, Italy
Geospatial Health 2019; 14:771
doi:10.4081/gh.2019.771

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Abstract

Early warning systems (EWS) have been proposed as a measure for controlling and preventing dengue fever outbreaks in countries where this infection is endemic. A vaccine is not available and has yet to reach the market due to the economic burden of development, introduction and safety concerns. Understanding how dengue spreads and identifying the risk factors will facilitate the development of a dengue EWS, for which a climate-based model is still needed. An analysis was conducted to examine emerging spatiotemporal hotspots of dengue fever at the township level in Taiwan, associated with climatic factors obtained from remotely sensed data in order to identify the risk factors. Machine-learning was applied to support the search for factors with a spatiotemporal correlation with dengue fever outbreaks. Three dengue fever hotspot categories were found in southwest Taiwan and shown to be spatiotemporally associated with five kinds of sea surface temperatures. Machine-learning, based on the deep AlexNet model trained by transfer learning, yielded an accuracy of 100% on an 8-fold cross-validation test dataset of longitude-time sea surface temperature images.

Introduction

Dengue is one of the world's most important vector-borne viral diseases, with outbreaks influenced by complex interactions between hosts, vectors, and the dengue virus types, as well as climate, transportation networks, high population density and unplanned urbanization leading to inadequate infrastructures (Gubler, 1998). Climate variability is a major factor affecting its space-time distribution and increasing the burden of dengue in recent years. The variation of sea surface temperatures (SSTs) following El Niño southern oscillation (ENSO) occurrences (Ferreira, 2014; Chretien *et al.*, 2015) has been shown to be a dengue indicator (Méndez-Lázaro *et al.*, 2014) as ENSO can lead to ecological settings favouring emergence or re-emergence of the disease (Johansson *et al.*, 2009a). For example, rainfall and temperature may directly or indirectly affect vector survival, lifespan, development and reproductive rates, which affect the spatiotemporal abundance and distribution of dengue fever (Canyon *et al.*,

1999; Hales *et al.*, 2002; Johansson *et al.*, 2009b; Chan and Johansson, 2012; Fan *et al.*, 2015).

The number of dengue patients has increased and the Ministry of Health and Welfare in Taiwan introduced mandatory case reporting in 1998. The largest dengue fever outbreak occurred in 2015, with 43,774 cases (Taiwan CDC, 2016). Drugs with effect against the dengue viruses are currently in progress (Lim *et al.*, 2013), while vaccine development has accelerated in recent years, with several vaccine candidates close to human clinical trials. Opposition to the use of vaccines observed in industrial as well as developing countries adds to the economic burden of vaccine introduction in the endemic countries (Douglas *et al.*, 2013) as well as safety concerns (Normile, 2017). Currently, as drugs and vaccines are not yet available, early warning systems (EWSs) have been proposed as a potential measure for controlling and preventing disease in an efficient and cost-effective manner (Lowe *et al.*, 2014; Shi *et al.*, 2016). Understanding the pattern of dengue spread and identifying the risk factors would facilitate the development of EWSs and introduce timely precautions for controlling epidemics (Gu *et al.*, 2016). Climate-based modelling is needed because dengue is consistently associated with factors such as temperature and rainfall (Adde *et al.*, 2016). Nevertheless, these correlations remain unclear due to the complex, multi-scale interplay between climate variables and the disease. Further, it is difficult to predict temporally and spatially fluctuating outbreaks using conventional methods requiring identification and assessment of the occurrence of dengue fever with reference to climate variables on a large space-time scale.

In the present study, we conducted a systematic study of dengue fever and its climate associations to identify the risk factors predicting outbreak levels related to climatic factors using satellite-generated remote sensing (RS) data, geographic information systems (GIS) and machine-learning.

Materials and Methods

In this study, the association between SST, rainfall, and temperature were studied as they appear to be the main climatic factors governing the seasonal occurrence of dengue fever in Taiwan.

Study site

Taiwan is an island in the Pacific Ocean situated north of the Philippines, southeast of the Ryukyu Islands of Japan and approximately 161 km from the south-eastern coast of China. It has an area of 35,980 km² and consists of 368 townships (Figure 1). The climate is tropical producing hot and rainy summers and mild winters with the Ocean moderating the hot South and the cooler North. The average temperature for January is 16°C (61°F) in the North and 20°C (68°F) in the South, whereas the average July temperature in both regions is 28°C (82°F). The average rainfall is 257 cm (101 in), ranging from 127 cm (50 in) in the middle of the west coast to more than 635 cm (250 in) on exposed mountain slopes. The winter monsoon brings heavy rains to the northern part of the island between October and March, whereas the summer monsoon brings rain to the South between May and September. The summer months also bring dangerous typhoons and cyclones. Each township has different geographical features, including forested areas, urban areas, agricultural fields as well as different demographic situations.

Case notification data

Dengue fever is a notified infectious disease (NID) in Taiwan since 1998, with all cases registered continuously and systematically by the NID Notification Surveillance System. We obtained the computerized database for 2016 from Taiwan Centers for Disease Control (Taiwan CDC, 2016), which contains all national dengue fever records on a daily basis from 1998, including age, gender, township of residence and time of disease onset. All notifications were summed at the township level by month from 1998 to 2015 and integrated into a GIS for spatiotemporal analysis and subsequent machine-learning.

Sea surface temperature data

Monthly mean SST data, obtained at 4-km resolution from the Advanced Very High Resolution Radiometer (AVHRR) Pathfinder, version 5.3, were collected from the ERDDAP website (<https://coastwatch.pfeg.noaa.gov/erddap>) of the U. S. National Oceanic and Atmospheric Administration (NOAA) for the period 1985-2015. The original dataset was archived in Network

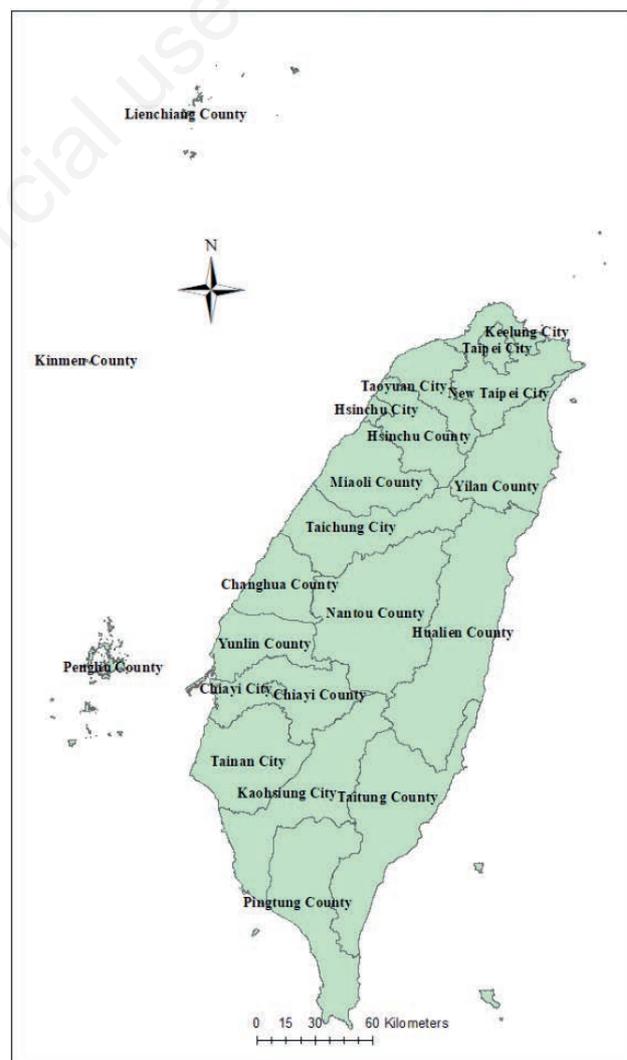


Figure 1. Map of Taiwan with major administrative divisions.

Common Data Form (NetCDF) files with global coverage. We extracted the SST values of interest from the sea around Taiwan (118E-125E; 21.5N-26.5N) using Interactive Data Language (IDL) programming and transformed the data into comma-separated value (CSV) files.

Rainfall data

The rainfall data were the Global Satellite Mapping of Precipitation (GSMaP) product of combined scenes from a plurality of satellites, using the Medium Wave Infrared (MWIR) algorithm (Kubota *et al.*, 2007), obtained from Japan Aerospace Exploration Agency (JAXA) (<http://sharaku.eorc.jaxa.jp/GSMaP/index.htm>). The GSMaP products are integrated hourly at a 0.1-degree grid as the resolution differs slightly depending on which satellite deliver the data. We used the 24-h averaged rainfall derived from the GSMaP hourly product, whose spatial resolution is the same as that of the hourly product. The monthly rainfall data were calculated by multiplying the daily GSMaP measurements by 24 and then working out the average sums from the first to the last day of the month. The monthly rainfall data were spatially interpolated and transformed to a 4-km grid of the UTM projection in NetCDF format. We extracted rainfall values of interest around Taiwan (118E-126E; 21N-27N) using Python programming (<https://www.python.org/>) and transforming them into Excel format. The Global Precipitation Measurement (GPM) project is the result of the continued collaboration between the U.S. National Aeronautics and Space Administration (NASA) and JAXA that previously produced the Tropical Rainfall Measuring Mission (TRMM) (Huffman *et al.*, 2019).

Temperature data

The temperature data used were thermal infrared data observed by the MODerate resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua satellites operated by the NASA and obtained from JAXA from the Satellite Monitoring for Environmental Study (JASMES - <http://www.eorc.jaxa.jp/JASMES/index.html>). The spatial resolution was 1 km and observations were made twice a day per satellite, once during daytime and once at night. Since all satellite orbits do not cross the study area, occasionally measurements were left out. The daily Ocean and Land Surface Temperature (OLST) product, which is spatially aggregated based on a 0.05-degree grid from daytime data captured by the Terra or Aqua satellites, was used. If both datasets were available on a particular day, the newest data were adopted as the product of the day. We used the monthly averaged temperature calculated from the OLST daily product that has the same spatial resolution as the daily product. This was spatially interpolated and transformed to a 4-km grid of the UTM projection in NetCDF format. We extracted temperature values of the sea around Taiwan (118E-126E; 21N-27N) using Python programming and transformed them into Excel format. The processed data were summed by month from 2002 to 2015 and used for spatiotemporal hotspot analysis and machine-learning.

Data analysis

ArcGIS 10.4.1 and ArcGIS Pro 1.4 (ESRI, Redlands, CA, USA) were used to conduct spatiotemporal hotspot analysis for understanding the pattern of dengue fever spread and identifying risk factors, applying Python for machine-learning and prediction of dengue outbreak levels. The assessment of dengue fever and climate associations on a large space-time scale and the identification

of climatic factors affecting this association are necessary for developing a climate-based model for EWS.

We used the concept of time geography, originally introduced by Hägerstrand (1970) and chronicled by Ellegård and Svedin (2012) to eventually be developed into a practically useful application by ESRI under the name space-time cube (STC) (ESRI, 2016a, 2016b). In the STC, the base represents a two-dimensional (2-D) geographical map with height, normally used for expressing altitude, instead used as the time axis. The STC produces a fusion of space and time where appropriate spatial analytical visualization techniques, such as GIS, facilitate understanding (Figure 2A-D).

The workflow comprised seven steps including data processing, emerging hotspot analysis and machine-learning. All data were first entered into Excel (A); converted into points (B); aggregated into space-time bins and point counts tallied (C); converted into a NetCDF data cube format (D); calculating Getis-Ord G_i^* (1995) and Mann-Kendall trend test statistics (Mann, 1945; Kendall and Gibbons, 1990) for each bin in the cube producing maps showing hotspots and coldspots (E); constructing a 3-D SST time series of data using the three axes longitude, latitude and time. Finally, machine-learning was applied to the dataset to produce outbreak predictions (F); and validating the prediction accuracy levels (G). ESRI 2016a, 2016b, 2016c, 2016d provide additional explanation.

Spatiotemporal hotspot analysis

Before conducting spatiotemporal hotspot analysis, we transformed the number of monthly dengue fever cases at the township level including coordinates and the 4-km resolution RS data from Excel format into point data and performed a projection transformation with WGS 84/UTM zone 51N in ArcGIS, v. 10.4.1. The decision on data format was made after comparing resulting hotspot maps after empirically testing bin sizes of 4 to 10 km and letting ArcGIS Pro calculate the time-step timeframes from 1 to 12 months. The emerging map with a 5-km spatial resolution and 6-month time-step interval was similar to the one calculated by ArcGIS (where the time-step interval was one year); moreover, the spatiotemporal trends of the hotspots were similar, reflecting the recent increase in dengue fever outbreaks in Tainan City, Taiwan. Based on these considerations, we converted all data into the NetCDF data cube structure by aggregating the number of dengue fever cases in each township as well as the SST, rainfall and temperature measurements into space-time bins with a spatial resolution of 5 km and 6-month time-step intervals using ArcGIS Pro, v. 1.4. Hotspot analysis, which identifies trends in potential clustering of case densities (counts) or values of the environmental variables studied (SST, rainfall and temperature) in a STC, was performed with the processed data using ArcGIS Pro 1.4 to detect any spatiotemporal patterns. The Getis-Ord G_i^* statistic was first applied for each feature in the dataset. The resultant z-scores and P-values describe whether the cluster features are high or low, while the ArcGIS analysis tool also checks each feature within the context of neighbouring features since a significant hotspot is not only characterized by a high value but also requires surroundings with similar high values. The local sum for a feature and its neighbours is compared proportionally to the sum of all features; when the local sum is very different from the expected local sum, and when that difference is too large to be random, a significant z-score results. The false discovery rate correction can be applied by adjusting significance to account for multiple testing and spatial dependency. A z-score above 1.96 or below -1.96 indicates a sig-

nificant hotspot or a significant coldspot at the significance level of $P < 0.05$ (Harris *et al.*, 2017). To test whether a significant temporal trend exists through each bin's time series of z-scores resulting from the Getis-Ord G_i^* statistic, the Mann-Kendall trend test was applied (Harris *et al.*, 2017). This is a rank correlation analysis for the bin count and its time sequence. The bin value for the first-time period is compared to the bin value for the second-time period. If the first is smaller than the second, the result is +1; in case of the *vice versa* outcome it is -1; and if the two values are the same, the result is zero, a result that would indicate absence of any trend when pairs are summed over time. Based on the variance for the values in the bin time series, the number of ties, and the number of time periods, the observed sum is compared to the expected sum (zero) to determine if the difference is statistically significant or not. The trend for each bin time series is recorded as a z-score and a P-value. The sign associated with the z-score determines if the trend is an increase in bin values (positive z-score) or a decrease in bin values (negative z-score) (ESRI, 2016b). With the hotspot z-score and P-value for each bin and the resulting trend z-score and P-value, the hotspot analysis tool categorized the patterns into 17 different types (Table 1) (ESRI, 2016c).

In addition, spatial link analysis was conducted by joining categories, such as emerging hotspots, coldspots or no pattern detected, based on spatial relationship in order to count the number of townships within a city or county classified into one of the 17 categories and also applied to examine spatiotemporal hotspot associations between dengue outbreaks and the climatic factors under study.

Machine-learning

Mathematically, the term convolution refers to both the result and the process of computing it, *i.e.* two functions produce a third that expresses how the shape of one is modified by the other. This has been practically utilized in the so called deep learning techniques, or artificial convolutional neural network (CNN) approaches, which have been shown to be particularly successful for image classification (Azimi *et al.*, 2018). Convolutional networks are commonly applied for visual imagery analysis and resemble the connectivity pattern seen in biological neurological organization, *e.g.* the mammalian visual cortex, where different neurons respond to stimuli in restricted regions of the visual field in a way that partially overlap so that they cover the entire field. The advantage is the reduction in pre-processing due to the CNNs independence from prior knowledge and human design.

The problem of overfitting, which must be avoided with small datasets like our own, was addressed by transfer learning. Here, we employed the fine-tuned AlexNet model in a Chainer framework (Krizhevsky *et al.*, 2012) to train the model more rapidly. Defined as the original AlexNet approach, this model consists of eight layers comprising five convolutional layers, some of which followed by max-pooling layers, which select the most active neuron in each region, *i.e.* the *winning* neurons, by partitioning layers into quadratic regions of local inhibition (Cireřan *et al.*, 2012). Further, there are three fully connected layers with a final 1000-way softmax except for the input and output layers, which fitted our input image and output categories, respectively. The so called deep learning method utilizes the softmax function as the final output that determines the class. The last three lay-

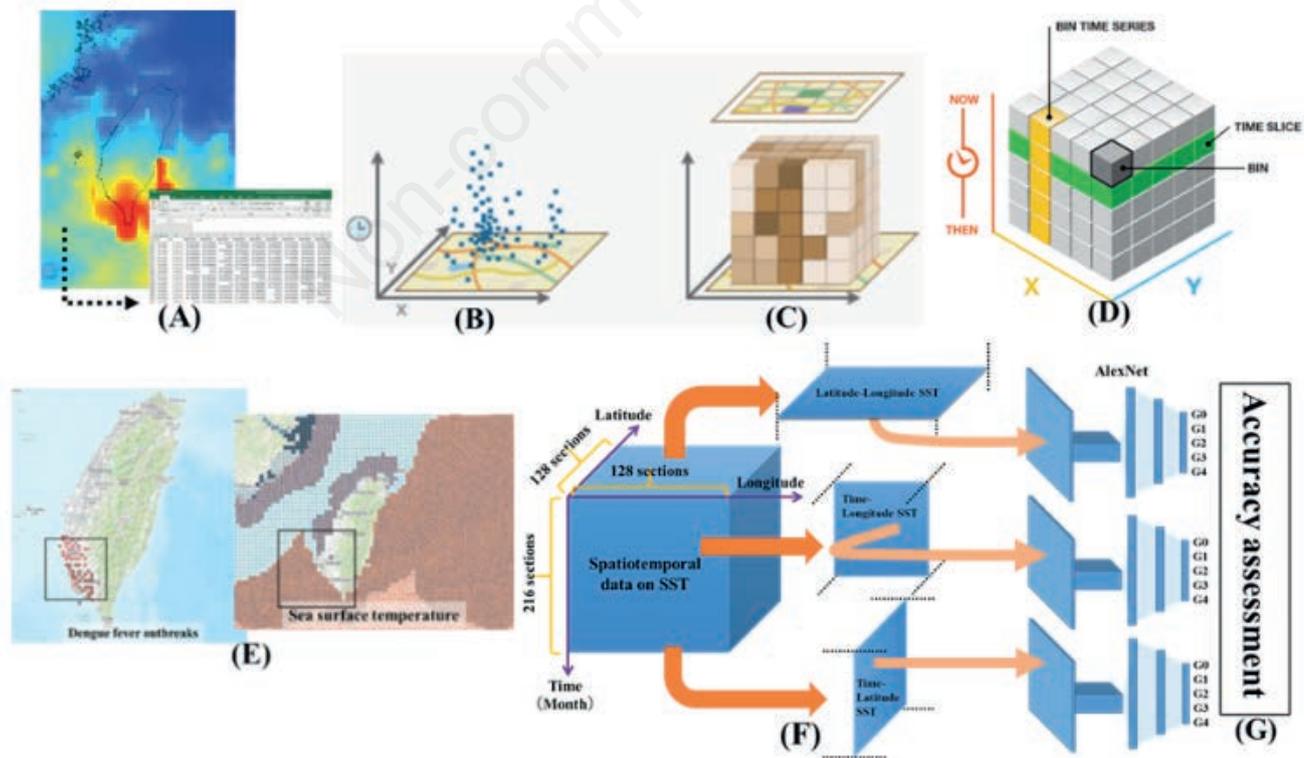


Figure 2. A schematic presentation of the workflow. Panels B-D are reproduced with permission from ESRI, 2016a. ©Copyright 2016 Environmental System Research Institute, Inc.

ers of the pre-trained network net must be fine-tuned for the new classification problem by replacing them with a fully connected layer, a softmax layer and a classification output layer; our model had five nodes to indicate the five levels of outbreaks from G0 to G4. Figure 3A shows a flow diagram of the procedure and Figure 3B the structure of the AlexNet in the classification. As a spatiotemporal correlation was found between dengue fever hotspots and SST hotspots, this dataset was subjected to deep learning. It contained the monthly dengue cases at the township level together with the monthly SST data for Taiwan and surrounding seas from January 2002 to December 2015. We constructed a 3-D time series of SST data using the three axes longitude, latitude and time and plotted the dengue fever data into this coordinate system. The SST values were standardized from 0.0 to 1.0 to show the minimum and maximum values of the SST in the whole dataset. The 3-D data were divided into 2-D data using longitude-time, latitude-time and longitude-latitude for data preparation and input into the fine-tuned AlexNet model. The SST covered the 13.0 to 31.0 °C range. The 2-D data were converted into 256×256 pixels with red-green-blue *pseudo* colour shades by Rainbow3 colour look-up-table showing SST in May 1998 and in October 1999 (Appendix Figure A1) in colours varying from green (low temperatures) to red (high temperatures). The ground appears purple in this scheme since it represents minimum values (temperatures recorded to be less than 13.0 °C). The nearest-neighbour interpolation method was applied to set suitable values in small data lost areas in the data conversion. The dengue outbreak levels were classified based on the maximum number of dengue cases in the same areas indicated by the 2-D SST data. Each cross-section along each axis was extracted and input into the fine-tuned AlexNet model. The final prediction was obtained by applying a dense layer with softmax activation, resulting in probabilities for five levels of dengue outbreak inten-

sities as shown in Table 2. The levels were based on the natural breaks classification method (Werneck, 2008).

The dataset, consisting of the pairs of the 2-D SST data and the dengue outbreak levels, was split into two, one set for training and one for the test. The training data were used to teach the AlexNet CNNs and the test data to assess model accuracy in an unknown situation. For all datasets, one-eighth of the actual data were used for the 8-fold cross-validation, which was estimated to provide reasonable 95% confidence intervals. For the five levels of dengue cases, 588/672 (87.5%) images were used for training and 84/672 (12.5%) for testing. AlexNet CNN was consistently carried out using graphics processing unit programming to achieve greater time efficiency. Two measures were used to avoid overfitting in the training phase: i) a dropout procedure in the AlexNet; and ii) a dataset design creating two groups, a training set and a test set, for the evaluation of the categorizing performances. The model was iterated with both 300 and 1000 epochs (the times when the whole dataset was entered into the model using the pre-determined weights). When training for the classification problem, the errors measured by the softmax cross entropy (SCE) (Goodfellow *et al.*, 2016) were employed to evaluate the CNN weights as loss indices. A learning curve was used to evaluate the performance of the model via the accuracy calculated by a confusion matrix, also known as error matrix that was created based on the actual and the predicted levels. Learning curves for the training and the test phases were a plot of predictive error for training and validation sets over a range of training set sizes. The learning curves for our deep learning model of the five levels were tested for accuracy in each epoch. The learning curves had a perfect accuracy (*i.e.* 1.0), meaning that all input data were classified correctly without any errors.

Table 1. Emerging hotspot/coldspot analysis categories. Reproduced with permission from ESRI, 2016c. ©Copyright 2016 Environmental System Research Institute, Inc.

Pattern (referring to both hotspots and coldspots)	Definition
No pattern detected	Does not fall into any of the hotspot or coldspot patterns defined below
New	Statistically significant location for the final time step, where statistically significant hot or coldspot(s) has not previously been reported
Consecutive	Location with single uninterrupted run of statistically significant hot- or cold spot- bins in the final time-step intervals, which did not exist prior to the final run and where <90% of all bins are statistically significant hot or cold spots
Intensifying	Location that has been statistically significant for hotspot(s) or cold spot(s) for 90% of the time-step intervals, including the final time step, and where the intensity of clustering of high counts at each time step is increasing overall and that the increase is statistically significant
Persistent	Location that has been statistically significant hotspot(s) or cold spot(s) for 90% of the time-step intervals with no discernible trend indicating an increase or decrease in the intensity of clustering of counts over time
Diminishing	Location that has been statistically significant hotspot(s) or cold spot(s) for 90% of the time-step intervals, including the final one, and where the intensity of clustering at each time step is decreasing overall and that this decrease is statistically significant
Sporadic	Location with on and off location of hotspot(s) or cold spot(s), and where <90% of the time-step intervals have been statistically significant hotspots and none of the time-step intervals have been statistically significant coldspots (<i>vice versa</i> for coldspots)
Oscillating	A statistically significant hotspot for the final time-step interval that has a history of also being a statistically significant coldspot during a prior time step, and where <90% of the time-step intervals have been statistically significant hotspots (<i>vice versa</i> for coldspots)
Historic	Not currently hotspot(s), but which has at least 90% of the time-step intervals been statistically significant for hotspot(s), and (<i>vice versa</i> for coldspots)

Results

Hotspot analysis

To better understand the geographic patterns obtained with the four kinds of data (number of dengue cases, SST, rainfall and temperature), we choose to visualize the results of the hotspot analysis performed (Figures 4 and 5, Appendix Figures A2 and A3).

Table 2. The levels of dengue outbreak intensities.

Monthly number of dengue cases	Level
0	1
1-20	2
21-100	3
101-1,000	4
>1,001	5

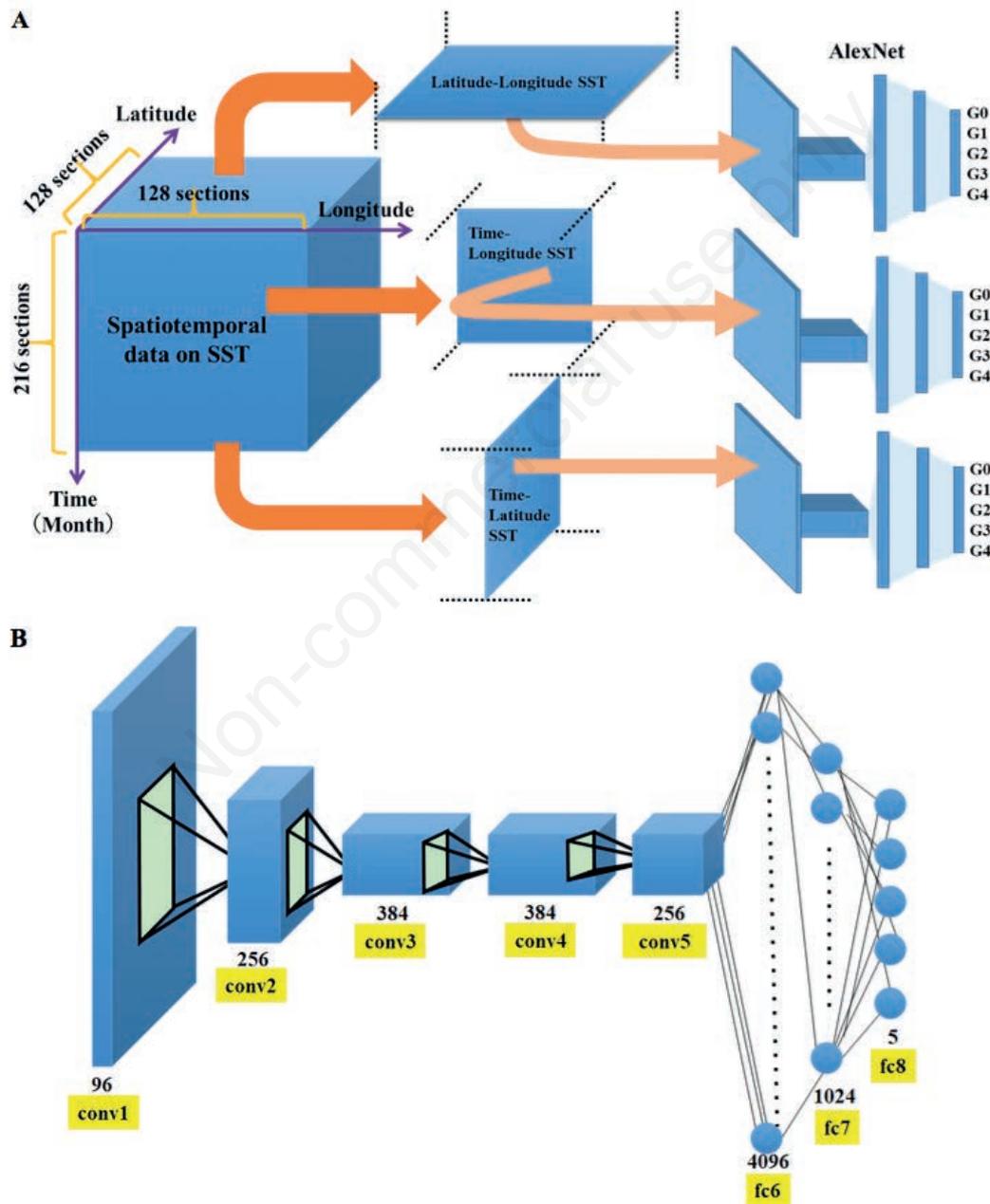


Figure 3. A) The experimental procedure of conducting transfer learning using the AlexNet model. SST, sea surface temperature. B) The architecture of the AlexNet model fine-tuned by transfer learning. Conv, convolutional; fc, fully connected.

The results of the spatial linkages are summarized in Tables 3-5, where the data represent the number of townships within a city or county that were classified as belonging to one or more of the 17 categories. Three hotspot categories (new, sporadic and consecutive) of dengue fever were seen in two cities and one county in southwest Taiwan (Figure 4 and Table 3). We focused on these three kinds, as effective disease control measures can be adapted before the disease has been settled. New hotspots were found in 31 townships in Tainan City and one township in Kaohsiung City, indicating that the dengue outbreaks occurred there after 2015. Sporadic hotspots were present in 21 townships in Kaohsiung City and 10 townships in Pingtung County, indicating that fewer than 12 (25 of the 6-month intervals) of the 14 study years (28 of the 6-month intervals) had significant hotspots. Consecutive hotspots were present in 10 townships in Pingtung County, 9 in Kaohsiung City and two in Tainan City, indicating that significant hotspots were detected continuously in multiple 6-month intervals leading up to the final time-step in the data series.

The highest number of new hotspots were found in Tainan City, the highest number of sporadic ones in Kaohsiung City and the highest number of consecutive ones in Pingtung County. Though the same number of sporadic and consecutive dengue

hotspots were found in Pingtung County, we designated this county a consecutive dengue fever hotspot because the number of consecutive hotspots here was the highest overall.

Turning to SST, we found various kinds of hotspots (intensifying, persistent, diminishing, sporadic, oscillating) along the flow channels of the Kuroshio Current and Kuroshio Branch current, while various kinds of coldspots (intensifying, persistent, diminishing, sporadic, oscillating and also historic) were found along the China coastal flow channels (Figure 5). We identified a spatiotemporal relationship between all dengue hotspots in south-western Taiwan and the five SST categories arranged along the flow channels of the Kuroshio Branch current, with all dengue hotspots facing one of the five kinds of SST hotspot.

Oscillating rainfall hotspots were found in many cities or counties (Table 4 and Appendix Figure A2), not only in Tainan City, Kaohsiung City and Pingtung County, but also in New Taipei City although it was not designated a dengue hotspot. In addition, oscillating rainfall hotspots appeared in 36 townships in Tainan City, 27 in Pingtung County, 26 in Kaohsiung City and 29 in New Taipei City. Rainfall coldspots were neither found in Tainan City nor in Kaohsiung City, while oscillating coldspots with respect to rainfall were found in two townships in Pingtung County. An oscillating

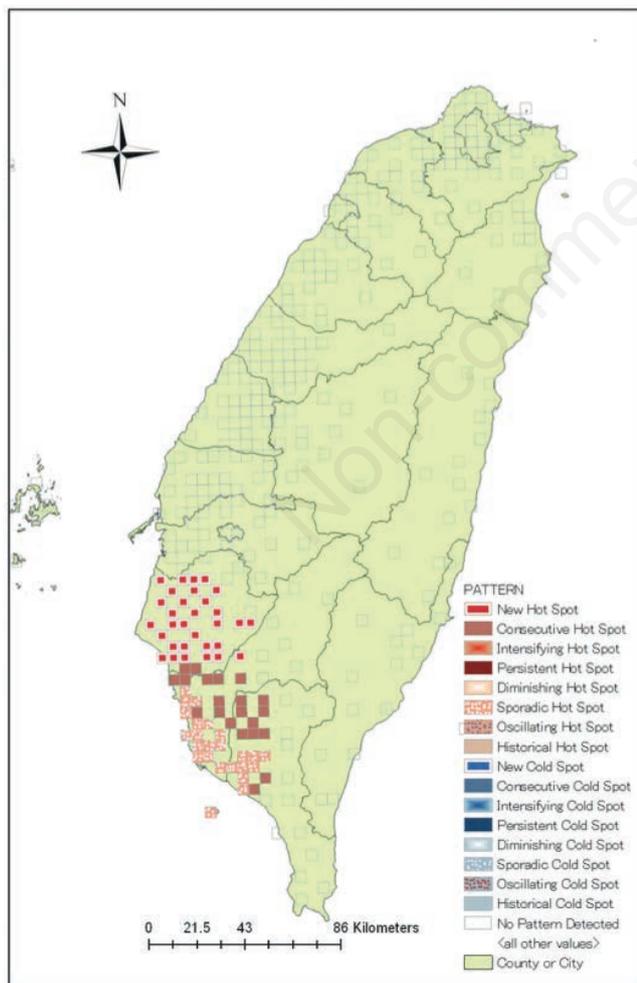


Figure 4. Spatiotemporal hotspot and coldspot patterns for dengue fever.

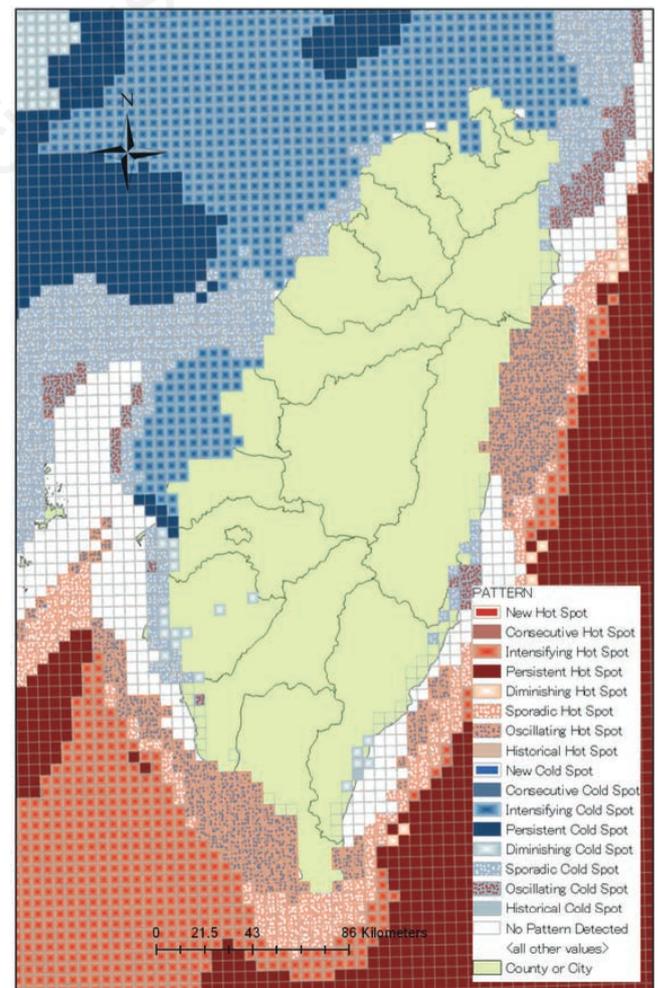


Figure 5. Spatiotemporal hot/cold spot patterns for sea surface temperature.

**Table 3. Dengue fever hotspots detected.**

County	New	Sporadic	Consecutive	No pattern	Total
Changhua County	-	-	-	26	-
Chiayi City	-	-	-	2	-
Chiayi County	-	-	-	18	-
Hsinchu City	-	-	-	3	-
Hsinchu County	-	-	-	13	-
Hualien County	-	-	-	13	-
Kaohsiung City	1	21	9	7	38
Keelung City	-	-	-	7	-
Kinmen County	-	-	-	6	-
Lienchiang County	-	-	-	4	-
Miaoli County	-	-	-	18	-
Nantou County	-	-	-	13	-
New Taipei City	-	-	-	29	-
Penghu County	-	-	-	6	-
Pintung County	-	10	10	13	33
Taichung City	-	-	-	29	-
Tainan City	31	-	2	4	37
Taipei City	-	-	-	12	-
Taitung County	-	-	-	16	-
Taoyuan City	-	-	-	13	-
Yilan County	-	-	-	12	-
Yunlin County	-	-	-	20	-
Total	32	31	21	284	368

The data represent the number of townships within a city or county classified into one of the 17 categories.

Table 4. Rainfall hotspots and coldspots detected.

County	No pattern	Oscillating coldspot	Oscillating hotspot	Persistent coldspot	Sporadic coldspot	Total
Changhua County	12	11	3	-	-	26
Chiayi City	-	-	2	-	-	2
Chiayi County	10	-	8	-	-	18
Hsinchu City	-	-	3	-	-	3
Hsinchu County	-	-	13	-	-	13
Hualien County	-	-	13	-	-	13
Kaohsiung City	12	-	26	-	-	38
Keelung City	-	-	7	-	-	7
Kinmen County	-	-	-	6	-	6
Lienchiang County	-	4	-	-	-	4
Miaoli County	-	-	18	-	-	18
Nantou County	-	-	13	-	-	13
New Taipei City	-	-	29	-	-	29
Penghu County	-	4	-	1	1	6
Pintung County	4	2	27	-	-	33
Taichung City	10	3	16	-	-	29
Tainan City	1	-	36	-	-	37
Taipei City	-	-	12	-	-	12
Taitung County	8	3	5	-	-	16
Taoyuan City	-	-	13	-	-	13
Yilan County	-	-	12	-	-	12
Yunlin County	6	11	3	-	-	20
Total	63	38	259	7	1	368

The data represent the number of townships within a city or county classified into one of the 17 categories.

hotspot was deemed significant for the final time-step interval if it also had a significant coldspot during a prior time-step, with less than 90% of the time-step intervals having had significant hotspots (Table 4). The overlap between the predominant rainfall hotspots with dengue hotspot indicates that rainfall may affect the spatiotemporal occurrence of dengue hotspots.

Oscillating temperature hotspots were found in many cities or counties (Table 5 and Appendix Figure A3), particularly in Pingtung County and Kaohsiung City, also shown to harbour consecutive and sporadic dengue hotspots, respectively. Oscillating temperature hotspots appeared in 28 townships in Pingtung County and 24 in Kaohsiung City. In addition, new temperature hotspots were found in two townships in Kaohsiung City and one township in Pingtung County. Persistent temperature coldspots were prominent in 31 townships in Tainan City, which was also the premier hotspot place for new hotspots. Spatiotemporal relationships were found between the consecutive and sporadic type of dengue hotspots on the one hand and oscillating and new temperature

hotspots on the other, as well as between new dengue hotspots and persistent temperature coldspots.

Machine-learning

The learning curves of the 2-D datasets on the longitude-time axes presented training and test validation accuracies for the network trained from the outset. Both training and test accuracy improved with validation accuracy, reaching 100% and 100% at 300 epochs, and 100% and 81% at 1000 epochs, respectively (Figure 6A and Table 6), on the latitude-time axes they presented training and test validation accuracies of 100% and 75% at 300 epochs, and 100% and 75% at 1000 epochs, respectively (Figure 6B and Table 6), while they presented training and test validation accuracies of 100% and 48% at 300 epochs, and 100% and 26% at 1000 epochs, respectively on longitude-latitude axes (Figure 6C and Table 6).

The learning curves of the 2-D datasets on longitude-time axes presented the higher accuracy of 100%, with 9.54×10^{-7} loss in SCE

Table 5. Temperature hotspots and coldspots detected.

County	New hotspot	Persistent hotspot	Oscillating hotspot	Diminishing hotspot	Persistent coldspot	Sporadic coldspot	Historic coldspot	No pattern	Total
Changhua County	-	12	14	-	-	-	-	-	26
Chiayi City	-	2	-	-	-	-	-	-	2
Chiayi County	-	14	1	1	-	-	-	2	18
Hsinchu City	-	-	-	-	1	2	-	-	3
Hsinchu County	-	-	2	-	6	-	-	5	13
Hualien County	-	10	3	-	-	-	-	-	13
Kaohsiung City	2	-	24	-	6	-	2	4	38
Keelung City	-	1	6	-	-	-	-	-	7
Kinmen County	-	-	3	-	1	-	-	2	6
Lienchiang County	-	-	-	-	4	-	-	-	4
Miaoli County	-	8	7	-	-	-	-	3	18
Nantou County	-	9	4	-	-	-	-	-	13
New Taipei City	-	-	10	-	6	-	-	13	29
Penghu County	-	-	3	-	2	-	-	1	6
Pintung County	1	1	28	-	-	-	-	3	33
Taichung City	-	18	11	-	-	-	-	-	29
Tainan City	-	1	1	-	31	-	1	3	37
Taipei City	-	-	4	-	-	-	-	8	12
Taitung County	-	1	5	-	5	-	-	5	16
Taoyuan City	-	-	-	-	12	-	-	1	13
Yilan County	-	-	2	-	-	1	-	9	12
Yunlin County	-	20	-	-	-	-	-	-	20
Total	3	97	128	1	74	3	3	59	368

The data represent the number of townships within a city or county classified into one of the 17 categories.

Table 6. Overview of the machine-learning accuracy levels.

	Epoch	Longitude-time (%)	Latitude-time (%)	Longitude-latitude (%)
Training	300	100	100	100
	1000	100	100	100
Test	300	100	75	48
	1000	81	75	26

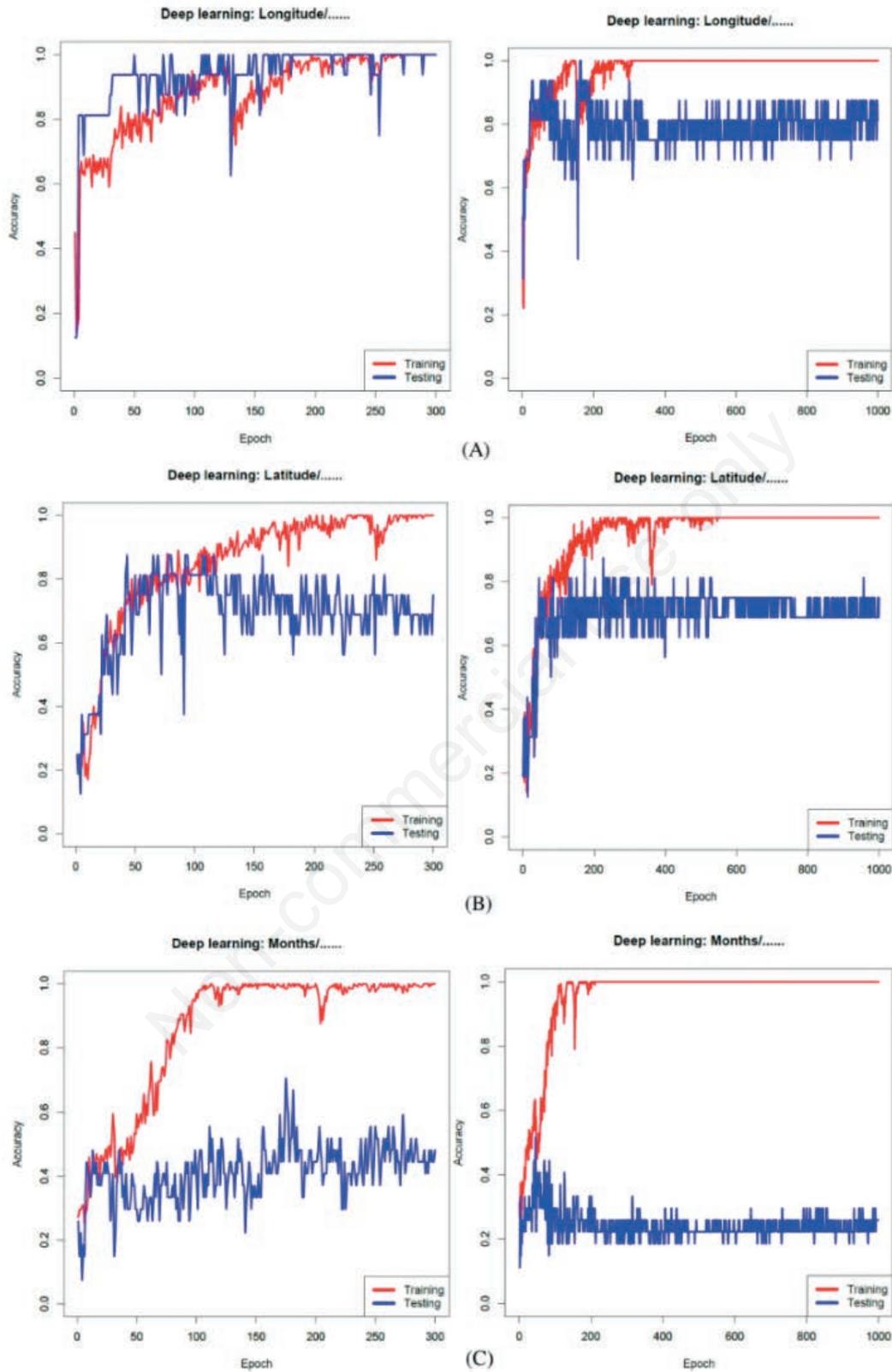


Figure 6. Learning curves for categorizing the accuracy of grading dengue outbreak levels from two combinations of sea surface temperature images of longitude, latitude and time. A) longitude-time datasets; B) latitude-time datasets; C) longitude-latitude datasets; the left images show up to 300 epochs and right side 1000 epochs; red and blue indicate the prediction accuracy in the training and testing phases, respectively.

compared to the other two results (75% with 3.83 SCE and 48% with 7.15 SCE) in the test phase. This suggests that a temporal SST change in the longitudinal direction would contribute to a highly reliable outbreak prediction. The deep learning method showed that longitude features are vital for prediction, and the findings relate to the SST varying vertically (Figure 5) suggesting that SST is a predominant factor with respect to dengue outbreak risk.

Discussion

The hotspot analysis presented here indicates that SST is temporally and spatially associated with dengue fever outbreaks in Taiwan. We found clear relationships between three kinds of hotspots (new, sporadic and consecutive) in southwest Taiwan and five kinds of SST hotspots along the flow channels of the Kuroshio Branch current. However, no dengue hotspots were found in the eastern part of the island facing the five SST hotspots along the Kuroshio Current flow channels. These areas, Tainan City, Kaohsiung City and Pingtung County, are not urbanized and have much smaller populations. Thus, urbanization and population could have contributed to the outbreaks.

We also identified spatiotemporal relationships between these three kinds of dengue hotspots and oscillating hotspots of rainfall, though there was one exception. Rainfall seemed to have something to do with the dengue outbreaks, both temporally and spatially, suggesting that ENSO-related SST and rainfall could be connected in space and time, influencing the spatiotemporal dengue hotspots. With regard to temperature, we found a spatiotemporal relationship between sporadic and consecutive dengue hotspots on the one hand and oscillating and new temperature hotspots on the other. Moreover, there was a spatiotemporal relationship between the new dengue hotspots and the persistent temperature coldspots. Our findings indicate that the differences in the number of temperature hotspots or coldspots influence vector survival and virus replication leading to differences in the dengue hotspot categories.

The presence of dengue and SST hotspots in these locations indicates that significant SST hotspot may influence the spatiotemporal abundance and distribution of dengue fever. Although township and sea surface areas cannot spatially overlap, past studies have shown that ENSO, due to SST fluctuations, is an important driving force with regard to dengue epidemics and can be a candidate indicator for predicting dengue fever outbreaks (Brunkard *et al.*, 2008; Tipayamongkholgul *et al.*, 2009). Our findings based on machine-learning support these past studies as they indicate that SST is a predominant warning factor. For example, the Kuroshio Branch current seems to promote outbreaks as atmospheric convection is active when the SST is high which often results in increased rainfall (Sasaki *et al.*, 2012) that leads larger breeding habitats for the *Aedes* mosquito vector. Considering the results of the hotspot analysis and machine-learning, a logical sequence in which high SST from the Kuroshio Branch current influences an outbreak can be established. Moreover, SST and rainfall may be connected in space and time and play the role of appropriate alarm indicators for predicting forthcoming outbreaks even at smaller spatiotemporal scales. Given the complexity of dengue outbreaks, a comprehensive study may be needed considering the influence of ENSO on rainfall and temperature, along with additional RS data on changing urbanization and population growth.

Conclusions

Integration of spatiotemporal hotspot analysis, RS data, and a machine-learning provide a crucial approach for developing climate-based forecasting models making it possible to deliver an early warning message to public health workers enabling them to promptly launch control or prevention activities. In our study, integration of RS data and machine-learning provided a crucial factor for classifying RS images of SST into the five outbreak levels without feature extraction from the images as required in conventional image analysis methods. Although the results of machine-learning present the possibility of a climate-based model for predicting the dengue outbreak level, they did not predict the spatiotemporal distribution of the outbreaks that would be needed to develop a proper EWS. Network operations based on CNN and convolutional long short-term memory (ConvLSTM) with daily data on dengue and other disease-promoting factors would predict the spatiotemporal distribution of outbreaks better. Such results would lead to a more accurate prediction modelling for effective disease intervention and prevention with the goal of early forecasting within an EWS. Such systems have a great potential also for other vector-borne diseases, such as chikungunya, lymphatic filariasis, Rift Valley fever, yellow fever and zika.

References

- Adde A, Roucou P, Mangeas M, Ardillon V, Desenclos JC, Rousset D, Girod R, Briolant S, Quenel P, Flamand C, 2016. Predicting dengue fever outbreaks in French Guiana using climate indicators. *PLoS Negl Trop Dis* 10:e0004681.
- Azimi SM, Britz D, Engstler M, Fritz M, Mücklich F, 2018. Advanced steel microstructural classification by deep learning methods. *Sci Rep* 8:2128.
- Brunkard JM, Cifuentes E, Rothenberg SJ, 2008. Assessing the roles of temperature, precipitation, and ENSO in dengue re-emergence on the Texas-Mexico border region. *Salud Publica Mex* 50:227-34.
- Canyon DV, Hii JL, Müller R, 1999. Adaptation of *Aedes aegypti* (Diptera: Culicidae) oviposition behavior in response to humidity and diet. *J Insect Physiol* 45:959-64.
- Chan M, Johansson MA, 2012. The incubation periods of dengue viruses. *PLoS One* 7:e50972.
- Chretien JP, Anyamba A, Small J, Britch S, Sanchez JL, Halbach AC, Tucker C, Linthicum KJ, 2015. Global climate anomalies and potential infectious disease risks: 2014-2015. *PLoS Curr* 7:1-19.
- Cireşan D, Meier U, Schmidhuber J, 2012. Multi-column deep neural networks for image classification. 2012 IEEE Conference on Computer Vision and Pattern Recognition. Institute of Electrical and Electronics Engineers (IEEE), New York, NY, pp 3642-3649.
- Douglas DL, DeRoeck DA, Mahoney RT, Wichmann O, 2013. Will dengue vaccines be used in the public sector and, if so, how? Findings from an 8-country survey of policymakers and opinion leaders. *PLoS Negl Trop Dis* 7:e2127.
- Ellegård K, Svedin U, 2012. Torsten Hägerstrand's time-geography as the cradle of the activity approach in transport geography. *J Transp Geogr* 23:17-25.
- ESRI, 2016a. Create Space Time Cube. ESRI, Redlands, CA,



- USA. Available from: <http://desktop.arcgis.com/en/arcmap/10.4/tools/space-time-pattern-mining-toolbox/create-space-time-cube.htm> Accessed: August 2018.
- ESRI, 2016b. Creation of the Space Time Cube (STC) and trend analysis. ESRI, Redlands, CA, USA. Available from: http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/learnmorecreatecube.htm#ESRI_SECTION1_F1EA94A3BA8940E0B56AB08A302D1C08 Accessed: August 2018.
- ESRI, 2016c. Emerging hotspot analysis. ESRI, Redlands, CA, USA. Available from: <http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/learnmoreemerging.htm> Accessed: August 2018.
- ESRI, 2016d. Hotspot analysis by Getis-Ord GI*. ESRI, Redlands, CA, USA. Available from: <http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-statistics-toolbox/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm> Accessed: August 2018.
- Fan J, Wei W, Bai Z, Fan C, Li S, Liu Q, Yang K, 2015. A systematic review and meta-analysis of dengue risk with temperature change. *Int J Environ Res Public Health* 12:1-15.
- Ferreira MC, 2014. Geographical distribution of the association between El Niño south oscillation and dengue fever in the Americas: A continental analysis using geographical information system-based techniques. *Geospat Health* 9:141-51.
- Goodfellow I, Bengio Y, Courville A, 2016. Deep learning. MIT Press, Cambridge, MA.
- Gu H, Leung RK, Jing Q, Zhang W, Yang Z, Lu J, Hao Y, Zhang D, 2016. Meteorological factors for dengue fever control and prevention in South China. *Int J Environ Res Public Health* 13:pii:E867.
- Gubler DJ, 1998. Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev* 11:480-96.
- Hägerstrand T, 1970. What about people in regional science? *Pap Reg Sci Assoc* 24:6-21.
- Hales S, de Wet N, Mairdonaal J, Woodward A, 2002. Potential effect of population and climate changes on global distribution of dengue fever: An empirical model. *Lancet* 360:830-4.
- Harris NL, Goldman E, Gabris C, Nordling J, Minnemeyer S, Ansari S, Lippmann M, Bennett L, Raad M, Hansen M, Potapov P, 2017. Using spatial statistics to identify emerging hot spots of forest loss. *Environ Res Lett* 12:024012.
- Huffman GJ, Pendergrass A, National Center for Atmospheric Research Staff, eds, 2019. The Climate Data Guide: TRMM: Tropical Rainfall Measuring Mission. National Center for Atmospheric Research, Boulder, CO. Available from: <https://climatedataguide.ucar.edu/climate-data/trmm-tropical-rainfall-measuring-mission>
- Johansson MA, Cummings DA, Glass GE, 2009a. Multiyear climate variability and dengue – El Niño southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis. *PLoS Med* 6:1-9.
- Johansson MA, Dominici F, Glass GE, 2009b. Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS Negl Trop Dis* 3:e382.
- Kendall MG, Gibbons JD, 1990. Rank correlation methods. 5th ed. E. Arnold, London - Oxford University Press, New York, NY.
- Krizhevsky A, Sutskever I, Hinton GE, 2012. ImageNet classification with deep convolutional neural networks. *Proc Adv Neural Inf Process Syst* 25:1090-8.
- Kubota T, Shige S, Hashizume H, Aonashi K, Takahashi N, Seto S, Hirose M, Takayabu YN, Ushio T, Nakagawa K, Iwanami K, Kachi M, Okamoto K, 2007. Global precipitation map using satellite borne microwave radiometers by the GSMaP project: production and validation. *IEEE T Geosci Remote* 45:2259-75.
- Lim SP, Wang QY, Noble CG, Chen YL, Dong H, Zou B, Yokokawa F, Nilar S, Smith P, Beer D, Lescar J, Shi PY, 2013. Ten years of dengue drug discovery: progress and prospects. *Antiviral Res* 100:500-19.
- Lowe R, Barcellos C, Coelho CA, Bailey TC, Coelho GE, Graham R, Jupp T, Ramalho WM, Carvalho MS, Stephenson DB, Rodó X, 2014. Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts. *Lancet Infect Dis* 14:619-26.
- Mann HB, 1945. Nonparametric tests against trend. *Econometrica* 13:245-59.
- Méndez-Lázaro P, Muller-Karger FE, Otis D, McCarthy MJ, Peña-Orellana M, 2014. Assessing climate variability effects on dengue incidence in San Juan, Puerto Rico. *Int J Environ Res Public Health* 11:9409-28.
- Normile D, 2017. Safety concerns derail dengue vaccination program. *Science* 358:1514-5.
- Ord K, Getis A, 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geogr Anal* 27:286-306.
- Sasaki YN, Minobe S, Asai T, Inatsu M, 2012. Influence of the Kuroshio in the East China Sea on the early summer (Baiu) rain. *J Clim* 25:6627-45.
- Shi Y, Liu X, Kok SY, Rajarethinam J, Liang S, Yap G, Chong CS, Lee KS, Tan SS, Chin CK, Lo A, Kong W, Ng LC, Cook AR, 2016. Three-month real-time dengue forecast models: an early warning system for outbreak alerts and policy decision support in Singapore. *Environ Health Perspect* 124:1369-75.
- Taiwan CDC, 2016. Taiwan national infectious disease statistics system. Taiwan Centers for Disease Control, Republic of China, Taiwan. Available from: <http://www.cdc.gov.tw/rwd/english>
- Tipayamongkolgul M, Fang CT, Klinchan S, Liu CM, King CC, 2009. Effects of the El Niño-southern oscillation on dengue epidemics in Thailand, 1996-2005. *BMC Public Health* 9:1-15.
- Werneck GL, 2008. Georeferenced data in epidemiologic research. *Ciêns Saúde Colet* 13:1753-66.