Predicting Dengue Outbreaks using Local Weather Factors and the North Atlantic Oscillation Index

Guohun Zhu*1, Ying Jiang*2, and Zejiao Chen*3

*1School of ITEE The University of Queensland, Brisbane, QLD 4067, Australia
E-mail: g.zhu@uq.edu.au
*2Hefei Guihuai Intelligent Technology Ltd, Hefei 230011, China
E-mail: jying1102@163.com
*3Changchun Cancer Hospital, Changchun 130012, China

Abstract. Climate is an important contributing factor in the outbreak and spread of dengue fever because it significantly affects the density and distribution of the mosquitoes carrying dengue virus. Dengue forecasting models based solely on local weather factors have had limited success. This paper proposes a novel dengue fever outbreak prediction model based on local weather factors and the multivariate North Atlantic Oscillation (NAO) index. The local weather data includes weekly temperature, precipitation, and humidity in San Juan, Puerto Rico from 1990 to 2013. Pacific/North American pattern (PNA) and NAO combined with the local weather are forwarded into a Support Vector Machine (SVM) to predict dengue outbreak. Statistical analysis shows that the outbreak of dengue cases has a strong negative-correlation with NAO indices and a strong positive-correlation with local temperatures and humidity. Classification results show that the accuracy of the dengue outbreak can only achieve 53.6% using weather data nine weeks in advance. However, combining local weather factors and NAO data in 15-week ahead, the model can predict dengue outbreak with 77.1% accuracy.

Keywords: Dengue fever, outbreak prediction, lag cross-correlation

1. INTRODUCTION

Dengue fever is a debilitating viral disease carried by mosquitoes that have spread widely in tropical and sub-tropical areas of the world in recent years. The World Health Organization (WHO) has reported that incidents of dengue fever have increased from 2.2 million in 2010 to 3.2 million in 2015 [1]. It could affect the COVID-19 identification during the pandemic period [2]. Recently, major epidemics have occurred in Southeast Asia, the Americas and the Western Pacific [1]. Thus, it is an important and worthwhile challenge to predict the outbreak of dengue accuracy in order to take preventative measures, minimize risks of dengue infection, and to limit the spread of the disease.

It is widely acknowledged that climate is an important factor in the spread of dengue. A number of studies [3], have shown that high temperature and humidity can increase the likelihood of dengue fever outbreak. High rainfall has also been shown to be a significant factor in the incidence of dengue [4, 5]. Temperature and precipitation have also been found to have strong coherence with dengue incidence [6]. Geostatistical techniques have been applied to study dengue outbreaks and determined that outbreaks usually occur during the rainy season in Queensland [7].

Based on these reports, a number of researchers have investigated predictive models for dengue fever outbreaks. For example, a fuzzy mining method was applied to local weather data to predict dengue outbreak in Philippines with 62.7% sensitivity [8]. A logistic regression method was applied to forecast the dengue fever with 83% sensitivity [9]. Dengue cases were reported to associate with rainfall [10]. A decision tree method showed that a threshold level of rainfall amount is an essential condition for dengue transmission in the tropics area, such as in Northern Queensland [11].

Machine learning approaches have been also applied to predict dengue outbreaks. Based on local climate, such as temperature, humidity and rainfall, artificial neural network (ANN) methods have been used to identify dengue outbreaks in Singapore [12] and Sri Lanka [13], respectively. Fuzzy classified methods have been applied in Philippines and South Korea [8, 14] with local climate data. ENSO indices were also applied in dengue outbreak forecasting in Costa Rica [15]. However, some forecast models, such as the Support Vector Machine (SVM) method in Noumea [16], specifically exclude the ENSO indices on dengue outbreaks.

However, recent studies have shown that dengue fever distribution is potentially impacted by global climate change. Siriyasatien et al. reported that dengue outbreaks are influenced by El Niño-Southern Oscillation (ENSO) [4]. During ENSO events, there are significant changes in the amount and intensity of rainfall in the tropics, especially over Southeast Asia and northern South America. In Australia, a decrease in the ENSO was strongly associated with an increase in dengue fever cases [17]. It was a weakly relationship between dengue cases and ENSO index in Puerto Rico [18].

In summary, previous studies have shown that the number of dengue cases is strongly related to local climate data and the ENSO index, but ENSO is rarely applied as a classifying feature. Until now, there are no public reports on the Atlantic Oscillation (NAO) index that studies dengue outbreaks.

This paper proposes a novel dengue outbreak forecast model based on local weather data and NAO index. The period of dengue outbreaks is from 1990/4/30 to 2013/4/23 in San Juan, Puerto Rico. NAO indices (rather than ENSO indices) are selected because San Juan is in the Northern Atlantic. Then a correlation analysis is applied to investigate whether the local weather factors or NAO indices are significantly related to the number dengue cases. Lastly, all extracted features are forwarded into a support vector machine to conduct the classification and evaluate the modelling performance.

2. METHODOLOGY

2.1. Data Collection and Preprocessing

The daily weather data and weekly dengue case data for the San Juan region (Puerto Rico) were acquired from the Dengue Forecasting Project on the Epidemic Prediction Initiative Website [1]. The weather data includes nine parameters: daily minimum temperature, daily maximum temperature, daily average temperature, daily dew temperature, daily temperature difference, daily precipitation, daily relative humidity, and daily specific humidity. Three daily global NAO indices (AO, NAO and PNA) were acquired from the data available in Climate Prediction Center [2] of the National Weather Service.

Because the local weather data is recorded daily and dengue cases are collected weekly, the climate data is converted into weekly values by averaging the daily values for a week or selecting the maximum or minimum values. Thus, weekly minimum temperature, weekly maximum temperature, weekly average temperature, weekly dew temperature, weekly temperature difference, weekly precipitation, weekly relative humidity, and weekly specific humidity were calculated.

In general, the number of incidences may not represent the outbreak seriously because a large population could accept a large outbreak number. Thus, this paper labels the records outbreak of dengue cases to five classes (Very Low, Low, Medium, High, Very High). The ranges are listed in Table 1.

Table 1-Class ranges for dengue cases

classes	Label	Ranges of Dengue cases
five	VL	<=10
	L	11—30
	M	31—50
	H	51—90
	VH	>90

[1] https://predict.phiresearchlab.org/legacy/dengue/index.html
[2] http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml

2.2. Support Vector Machine

Support Vector Machine (SVM) is a popular type of Machine Learning method that constructs an optimal hyperplane in the sample feature space, and applies it to achieve classification of the training sample.

There are two hyperspaces in a SVM. One is a linear space discrimination, another is a nonlinear classification based on a “kernel” function. There are four kernel functions: linear, polynomial kernel, radical basis function (RBF) and sigmoid. In this study, the optimal RBF kernel was assigned a value of 0.81, because dengue data does not satisfy normal distributions[19]. The incidence of dengue is the forecast object in my research. In this research, the radial basis function (also known as Gaussian kernel) is used as the kernel function Because the experiment results in our previous work shows that the kernel function has a high performance on identify the dengue fever [18]. The radial basis function form is.

$$k(x, x_i\} = e^{(-r|x-x_i|^2)} \qquad (1)$$

where r is nuclear parameter, x_i is the sample factor of the support vector, x is a predictor for vector.

3. RESULTS

3.1. Data Collection and Preprocessing

Figure 1 shows the dengue cases in San Juan and three NAO indices: AO, NAO and PNA trends, the number of dengue cases reported weekly. The period is from 4 April 1990 to 23 April 2013. The higher number (>200) were occurred in 1994.9-1994.12, 1998.7-1998.8, 2010.7-2010.9 and 2012.11-2013.1.

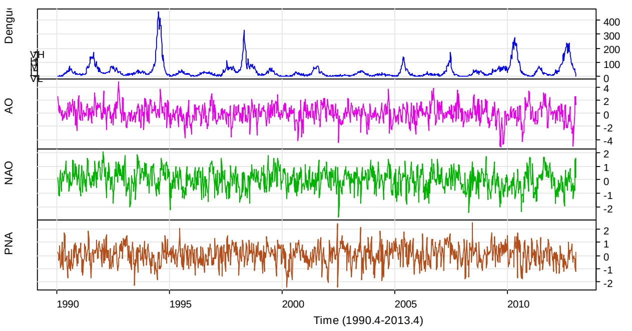

Fig.1 The weekly dengue cases in San Juan on Top panel, the y-axis are labeled with five levels (VL, L, M, H, and VH), following panels are AO, NAO, and PNA

It is difficult to observe which NAO indices has strong correlation with Dengue cases in Fig.1. However, the delay lag between dengue instances and weather factors can be extracted by cross-correlation as shown in Fig.2. As for maximum temperature, minimum temperature, average temperature (ignore from Fig. 2), dew temperature, temperature difference (ignore from Fig 2), relative humidity, specific humidity, are shown to be strongly positive correlated with dengue instance. However, it was exhibited a strong anti-correlation between dengue instances and AO, NAO, and PNA, where precipitation was weekly correlated with dengue instances.

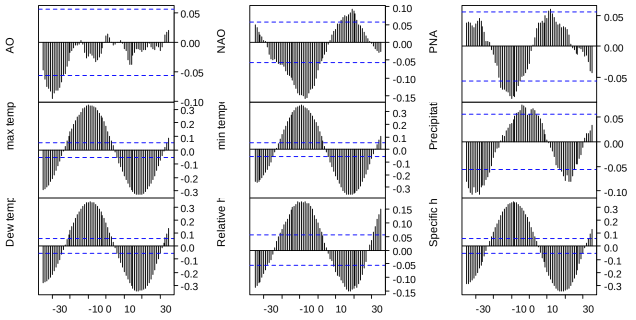

Fig.2 Weekly lag cross-correlation between dengue cases and AO, NAO, PNA, Max. temperature, Min. temperature precipitation, dew temperature, relative humidity and specific humidity.

The detail of the correlation coefficient of the dengue cases and the weekly average local weather and NAO indices are shown in Table 1. It is shown that there exists anti-correlation between dengue cases and global factors (NAO) ($p<0.01$). In contrast, a similar study of dengue in Northern Queensland of Australia which also shows SOI has a 12-week positive correlation impact on dengue cases [20]. It is also found that the 9-week ahead temperatures and specific humidity are positive correlated with dengue cases. The daily minimum temperature is strongest corresponding to dengue cases.

Table 2-Class ranges for dengue cases

Features data (weekly)	Lag	**Correlation coefficient to *dengue cases***	**p-value**
AO	-28	-0.0984	0.0008
NAO	-8	-0.1517	1.495e-07
PNA	-10	-0.0853	0.0049
Max temperature	-9	0.3282	2.2e-16
Mean temperature	-9	0.3399	2.2e-16
Min temperature	-9	0.3510	2.2e-16
Dew Temperature	-9	0.3429	2.2e-16
Difference Temperature	-9	0.0625	0.0326
Precipitation	-28	-0.1085	0.0002
Relative Humidity	-11	0.1778	7.157e-10
Specific Humidity	-9	0.3441	2.2e-16

3.2. 15-Week in Advance to Predicted Dengue Outbreak

The predicated data was divided into two sets. One is the training set, which comprises data from 2000 to 2012. The second set is the testing set, which comprises data for the period from 2012 to 2013. Each week was assigned a label according to Table 1 (e.g., VL, L, M, H, VH), based on the number of dengue cases in that week. Two periods are tested, one is 14 weeks, another is 8 weeks.

Twelve features: maximum temperature, minimum temperature, average temperature, dew temperature, dew temperature, temperature difference, precipitation, relative humidity, specific humidity and three NAO indices, were forwarded into a SVM for training and testing. Both 8-week and 14-week in advance classifications were performed respectively. The results are show in Table 2.

Table 3 shows the confusion matrix for 5-class classification with NAO indices (accuracy of 77.2%). The positive predicative value (PPV) is highest for outbreak of dengue cases but it has lower sensitivities.

Table 3: Confusion matrix for 5-class prediction

	VL	**L**	**M**	**H**	**VH**	**PPV**
VL	269	16	12	9	1	0.876
L	34	421	57	55	72	0.659
M	3	2	114	4	3	0.905
H	0	2	1	83	0	0.965
VH	0	0	0	0	29	1.000
Sens.	.879	.955	.620	0.50	0.276	

Fig. 3 shows two receiver operating characteristic (ROC) curves analysis for the classifying results. The area under ROC (AUC) is larger, the performance is better. In the Fig. 3, the arears under the curve are 0.8993 and 0.7849 respectively.

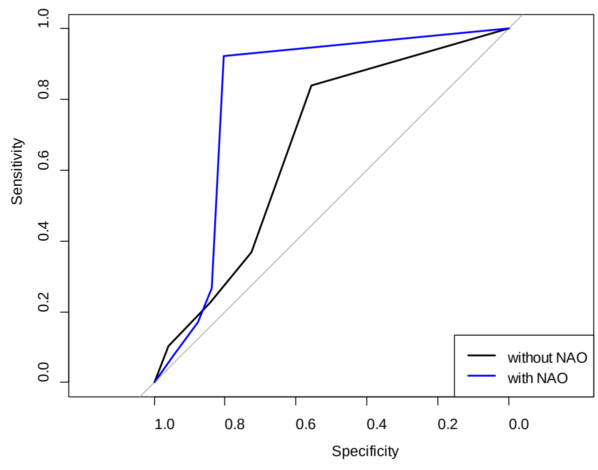

Fig. 3 9-weeks ROC curve for two type classifying, one is with NAO and black line denotes AUC without NAO.

4. DISCUSSION

The first row of Fig. 2 shows that the NAO indices are weakly anti-correlated with the dengue cases. In contrast, Max. temperature, Min. temperature and Dew temperature are the factors with the highest impact on the dengue incidences. This is probably the reason that many researchers believe that the ENSO/NAO indices have no effect on dengue transmission [6]. In fact, previous analysis of San Juan data using Monte Carlo simulation results indicate that rainfall strongly modulates the dengue cases [21].

However, Table 2 shows quite clearly that using SVM with only local weather data performs quite poorly at predicting dengue outbreaks – accuracy is 56.7% for 9-weeks ahead and 52.7% for 14-weeks in advance. However, if the NAO indices are included in the input to the SVM, the performance improved to 77.2% accuracy for 9-week ahead classification and 76.1% accuracy for 15-week in advance class classification. Thus, the NAO indices are a significant factor when predicting dengue outbreaks in San Juan.

Table 4: Some prediction results listed, where ahead period used only minimum lag week numbers.

Reference (Year)	Modelling Method	Dengue Regions	Ahead period test data (min-lag)	Sensitivity or ACC
Buczak et al. (2014) [8]	Fuzzy classification 2-group	Philippines	4-week	62.7%
Nishanthi et al. (2014) [13]	ANN	Sri Lanka	1-week	68.5% (ACC)
Kesorn et al. (2015) [22]	SVM 2-class	Thailand	No mentioned	87.4% (ACC)
Zhang et al. (2016) [23]	Binomial regression	Guangzhou	5-week	78.83% (sensitivity)
Chan et al. (2015) [9]	Logistic regression	Taiwan	2-week	80% (sensitivity)
Huang et al. (2015) [20]	Tree	Northern Queensland	2-week	80.2%
Our method	SVM 5-class	San Juan	9-week	76.1% (ACC)
			14-week	77.1% (ACC)

Table 4 shows that using both local weather data and NAO indices in the SVM for dengue outbreak forecast, is more ahead than all previous methods. Certainly, dengue cases outbreak pattern should be not the same in different area. However, our results clearly reveal that NAO indices (or ENSO indices for the Southern Hemisphere) should be employed together with local weather data when developing predictive models for dengue fever outbreaks.

This paper is the first time to investigate the impact of including NAO indices as well as local weather data, in dengue outbreak forecast modelling. The correlation analysis showed that local weather data (such as temperature, humidity, and rainfall) has no linear relation with dengue incidence but that NAO indices have a weak relationship with dengue outbreaks. However, our SVM-based classification results show that if both NAO indices plus local weather data are used for the modelling, predictive performance can be improved. A positive predicator value of 0.965 was achieved when predicting dengue cases in the higher cases (>50), which is higher accuracy than previous approaches.

One drawback of this approach is only used local weather data and NAO indices to evaluate the dengue outbreaks. However, it could be potentially migrated and applied to predict dengue fever outbreaks in other high risk areas such as South East Asia and North Queensland, using a combination of local weather data and ENSO indices (for Southern Hemisphere and Pacific Ocean regions). This is a research area for future validation.

5. DISCUSSION

This paper is the first time to investigate the impact of including NAO indices as well as local weather data, in dengue outbreak forecast modelling. The correlation analysis showed that local weather data (such as temperature, humidity, and rainfall) has no linear relation with dengue incidence but that NAO indices have a weak relationship with dengue outbreaks. However, our SVM-based classification results show that if both NAO indices plus local weather data are used for the modelling, predictive performance can be improved. A positive predicator value of 0.965 was achieved when predicting dengue cases in the higher cases (>50), which is higher accuracy than previous approaches.

Although this approach has only been evaluated using local weather data from the San Juan region in Puerto Rico and NAO indices, it could easily be migrated and applied to predict dengue fever outbreaks in other high-risk areas such as South East Asia and North Queensland. This might be a research area for future validation using a combination of local weather data and ENSO indices (for Southern Hemisphere and Pacific Ocean regions).

REFERENCES:

[1] World Health Organization. “Dengue and severe dengue,” No. WHO-EM/MAC/032/E. World Health Organization. Regional Office for the Eastern Mediterranean, 2016.

[2] L. T. Lam, Y. X. Chua, D. H. Tan. “Roles and challenges of primary care physicians facing a dual outbreak of COVID-19 and dengue in Singapore,” Family Practice. May 2020.

[3] S. Sharmin, K. Glass, E. Viennet, and D. Harley, "Interaction of mean temperature and daily fluctuation influences dengue incidence in Dhaka, Bangladesh," PLOS Negl Trop Dis, vol. 9, no. 7, p. e0003901, 2015.

[4] P. Siriyasatien, A. Phumee, P. Ongruk, K. Jampachaisri, and K. Kesorn, "Analysis of significant factors for dengue fever incidence prediction," BMC bioinformatics, vol. 17, no. 1, p. 1, 2016.

[5] M. A. Johansson, D. A. Cummings, and G. E. Glass, "Multiyear climate variability and dengue—El Nino southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis," PLoS Med, vol. 6, no. 11, p. e1000168, 2009.

[6] W. Hu, A. Clements, G. Williams, S. Tong, and K. Mengersen, "Spatial patterns and socioecological drivers of dengue fever

transmission in Queensland, Australia," Environmental health perspectives, vol. 120, no. 2, p. 260, 2012.

[7] A. L. Buczak et al., "Prediction of high incidence of dengue in the Philippines," PLoS Negl Trop Dis, vol. 8, no. 4, p. e2771, 2014.

[8] T.-C. Chan, T.-H. Hu, and J.-S. Hwang, "Daily forecast of dengue fever incidents for urban villages in a city," International journal of health geographics, vol. 14, no. 1, p. 1, 2015.

[9] S. Naish, P. Dale, J. S. Mackenzie, J. McBride, K. Mengersen, and S. Tong, "Climate change and dengue: a critical and systematic review of quantitative modelling approaches," BMC infectious diseases, vol. 14, no. 1, p. 1, 2014.

[10] X. Huang, A. C. Clements, G. Williams, G. Milinovich, and W. Hu, "A threshold analysis of dengue transmission in terms of weather variables and imported dengue cases in Australia," Emerging microbes & infections, vol. 2, no. 12, p. e87, 2013.

[11] H. M. Aburas, B. G. Cetiner, and M. Sari, "Dengue confirmed-cases prediction: A neural network model," Expert Systems with Applications, vol. 37, no. 6, pp. 4256-4260, 2010.

[12] P. Nishanthi, A. Perera, and H. Wijekoon, "Prediction of Dengue Outbreaks in Sri Lanka using Artificial Neural Networks," International Journal of Computer Applications, vol. 101, no. 15, 2014.

[13] A. L. Buczak et al., "Fuzzy association rule mining and classification for the prediction of malaria in South Korea," BMC medical informatics and decision making, vol. 15, no. 1, p. 47, 2015.

[14] D. Fuller, A. Troyo, and J. C. Beier, "El Nino Southern Oscillation and vegetation dynamics as predictors of dengue fever cases in Costa Rica," Environmental Research Letters, vol. 4, no. 1, p. 014011, 2009.

[15] E. Descloux et al., "Climate-based models for understanding and forecasting dengue epidemics," PLoS Negl Trop Dis, vol. 6, no. 2, p. e1470, 2012.

[16] W. Hu, A. Clements, G. Williams, and S. Tong, "Dengue fever and El Nino/Southern Oscillation in Queensland, Australia: a time series predictive model," Occupational and environmental medicine, vol. 67, no. 5, pp. 307-311, 2010.

[17] M. R. Jury, "Climate influence on dengue epidemics in Puerto Rico," International Journal of Environmental Health Research, vol. 18, no. 5, pp. 323-334, 2008.

[18] Y. Jiang, G. Zhu, and X. Dang, "Predication of Dengue Outbreak based on Poisson Regression and Support Vector Machine," in The 7th International Symposium on Computational Intelligence and Industrial Applications (ISCIIA2016), Beijing, 2016, 2016, pp. 1-6.

[19] X. Huang, A. C. Clements, G. Williams, G. Devine, S. Tong, and W. Hu, "El Niño-Southern Oscillation, local weather and occurrences of dengue virus serotypes," Scientific reports, vol. 5, 2015.

[20] C. W. Morin, A. J. Monaghan, M. H. Hayden, R. Barrera, and K. Ernst, "Meteorologically driven simulations of dengue epidemics in San Juan, PR," PLoS Negl Trop Dis, vol. 9, no. 8, p. e0004002, 2015.

[21] K. Kesorn et al., "Morbidity rate prediction of dengue hemorrhagic fever (DHF) using the support vector machine and the Aedes aegypti infection rate in similar climates and geographical areas," PloS one, vol. 10, no. 5, p. e0125049, 2015.

[22] Y. Zhang et al., "Developing a Time Series Predictive Model for Dengue in Zhongshan, China Based on Weather and Guangzhou Dengue Surveillance Data," PLoS Negl Trop Dis, vol. 10, no. 2, p. e0004473, 2016.