Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Identification of Significant Climatic Risk Factors and Machine Learning Models in Dengue Outbreak Prediction

Felestin Yavari Nejad
Universiti Malaya
Kasturi Dewi Varathan (kasturi@um.edu.my)
Universiti Malaya https://orcid.org/0000-0003-3421-4501

Research article

Keywords: Risk Factor, Dengue, Outbreak Prediction Model, TempeRain Factor

Posted Date: October 7th, 2019

DOI: https://doi.org/10.21203/rs.2.15755/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Background: Dengue fever is a widespread viral disease and one of the world's main pandemic vector-borne infections and serious hazard to humanity. According to the World Health Organization (WHO), the incidence of dengue has grown dramatically worldwide in recent decades. The WHO currently estimates an annual incidence of 50–100 million dengue infections worldwide. Until today there is no tested vaccine or treatment to stop or prevent dengue fever thus the importance of dengue outbreak prediction is significant. The current issue in dengue outbreak prediction is accuracy. There are a limited number of studies that look at in depth analysis of climate factors in dengue outbreak prediction.

Methods: In this study, the most significant and important climatic factors that contribute to dengue outbreak were identified. These factors were used as input parameters on machine learning models. The models were trained and evaluated based on four-year data from January 2010 to December 2013 in Malaysia.

Results: This work provides two main contributions. A new risk factor, which was called TempeRain Factor (TRF), was determined and used as an input parameter for dengue prediction outbreak model. Moreover, the TRF was applied to demonstrate that its strong impact on dengue outbreaks. Experimental results showed that Support Vector Machine (SVM) with the newly identified meteorological risk factor in this study resulted in higher accuracy of 98.09% and reduced the root mean square error to 0.098 for predicting dengue outbreak.

Conclusions: This research managed to explore on the factors that are being used in dengue outbreak prediction systems. The main contribution of this paper is in identifying new significant factors that contribute in dengue outbreak prediction. From the evaluation, we managed to obtain a significant improvement in accuracy of the machine-learning model in dengue outbreak prediction.

Introduction

Pandemic infectious diseases are spreading in many geographical areas. According to WHO reports, dengue fever is one of the most important mosquito-borne diseases. Dengue is a common problem and one of the deadliest infectious diseases worldwide. WHO has identified dengue as the major rapidly spreading mosquito-borne virus-like illness. Thus, this disease is a threat and presents severe risk for human populations in numerous tropical and sub-tropical regions [1–6]. Health organizations should have a prediction and early warning system to control and monitor dengue fever [7]. Member states in three WHO regions regularly reported the annual number of cases increased from 2.2 million in 2010 to 3.2 million in 2015 [8].

Moreover, WHO estimated an annual projection of 50–100 million dengue infections worldwide. Furthermore, annual mortality of approximately 20,000–22,000 deaths caused by dengue fever has been reported [8,9]. Contrary to yellow fever or other mosquito-borne diseases, a vaccine or treatment against all serotypes of dengue virus is not available, and no antiviral drug for treating dengue fever has been reported [10]. The only alternative is to prevent or control the outbreak of this disease.

The accuracy of the prediction system for outbreaks is the primary and important concern for controlling dengue fever [11]. Establishing related risk factors are critical for prediction systems [12]. Given that climate factors play main role in this disease, identifying the relation between weather information and incidence of dengue outbreak is a main task in the establishment of an accurate prediction system for future outbreak [13,14]. In this study, important climatic risk factors such as temperature, relative humidity, and amount of rainfall were also examined. The current accuracy for prediction systems ranges from 82.39% to 97.05% [12,15–20].

This study is important, because it identifies the critical climatic risk factors in dengue outbreak prediction (TempeRain Factor). Then, the identified critical factors (TRF) were applied in prediction models that increased the accuracy of the prediction and reduced the error of the prediction model. This process is expected to especially help the authorized organization or decision makers in health organization, governments, and others to be aware and plan better prevention program in the near future.

Background

Related works

A recent study from the WHO indicated that 390 million dengue infections occur annually (95% credible interval of 284–528 million), of which 96 million (67–136 million) are manifested clinically with any severity of disease [21]. Another study on the prevalence of dengue has estimated that 3.9 billion people in 128 countries are at risk of infection from dengue viruses [22]. As of December 2018, the Ministry of Health, Malaysia (MOH) has recorded approximately 80,615 cases with 147 death cases, compared with 19,884 cases in December 2011 with 36 deaths [23]. The number of cases increased by approximately four folds. Moreover, by the end of March 2019, a total of 39,805 cases of dengue with 64 deaths were reported in Malaysia compared to March 2018 with 16,917 cases with 34 deaths [24].

Various early warning and monitoring systems are currently implemented to monitor dengue outbreak worldwide. Dengue prediction models have been previously investigated, but some of these models still have limitations on obtaining high accuracy in dengue outbreak prediction [11,25]. Different models and techniques have been integrated in designing several models for predicting dengue outbreak. Several studies have established prediction models for dengue outbreak using artificial neural networks [12].

Researches have also used hybrid model for outbreak prediction. The hybrid model is an example of integrated model and there are many model which are based on genetic algorithm to determine the weight in a neural network model [11,13,14,20,26]. In Singapore the researchers found significant correlated dengue cases with climatic variables through a Poisson Regression Model [27]. A researcher [17] developed a dengue outbreak prediction system in Singapore and obtained 90% accuracy. Thitiprayoonwongse established another prediction system based on a decision tree and obtained 96.7% accuracy [18]. Models of dengue outbreak prediction system in Malaysia showed accuracies of 96.27% and 82.39% [12,20].

Vulnerability maps of dengue incidences have been generated in Malaysia, resulting in the development and implementation of visualized and predictive modeling using Geographic Information System (GIS) for dengue in Selangor, Malaysia [28]. In Indonesia, the dengue outbreak prediction for GIS-based early warning system achieved an accuracy of 97.05% [15]. Another study from the National Taipei University of Technology used the C-Support Vector Classification to forecast dengue fever epidemics in Taiwan, and the accuracy for shuttle RBF kernel type was 90.5% [16]. In 2015, Loshini et al. predicted localized dengue incidences in Malaysia using an ensemble system for identification and found that ensemble models have better prediction power compared with single model [29].

Prediction of dengue outbreak is crucial globally, because this infectious disease remains as one of the main issues in many countries [11,26,30,31]. Table 1 listed the studies on different models of dengue outbreak prediction with distinct climatic risk factors. The star (*) in the columns in Table 1 shows the risk factors used by the different studies.

Table 1: Risk Factor for Dengue Outbreak Prediction Model

Most of these studies on dengue were from Asian countries, such as East-West Asia and regions in the Pacific Ocean. According to the WHO, countries in East-West of Asia, such as Malaysia, Singapore, Taiwan, Indonesia, Bangladesh, and Thailand, are critical areas for dengue fever. Most studies have shown that temperature and rainfall have direct and important effects on dengue outbreaks [14,20,26,30,31].

Moreover, changing climatic factors, such as increasing temperature, rainfall, and humidity are the most influential driving forces of dengue transmission [31]. A study had correlated dengue cases with climatic variables for the city of Singapore and the model worked on dengue cases were considered as dependent variable, while climatic variables, such as rainfall, maximum and minimum temperature, and relative humidity, were independent variables [27]. Based on the grade of each risk factor used in the 22 studies shown in Table1, most studies primarily used total rainfall (17 studies), average temperature (16 studies), relative humidity (15 studies), minimum temperature (11 studies), and maximum temperature (10 studies) as the input of the prediction model. However, none of the studies were focusing on detailed analysis on the factors and none looked into detailed relationship that could exist between the factors.

This research worked on detailing out the factors and identifying the association that exists between the identified factors which contribute to the dengue outbreak prediction system. The detailed factors will then be used as input values for dengue outbreak prediction.

Methods

This section explains in detail the methodology used for this research that includes dataset used, analysis made, new factor identification integrated input factors, evaluation with machine learning models, and evaluation method. Fig 1 illustrates and shows the conceptual framework of our research of this research.

Fig 1: Conceptual Framework on Identifying Significant Climate Factors in Dengue Outbreak Prediction

Data is retrieved from two official sources which are Ministry of Health Malaysia and Malaysia Meteorological Department. The data are combined and cleaned accordingly. The pre-processed data is analysed and new detailed factors are identified. The factors are then integrated and feed as integrated input factors to different machine learning models and evaluated. The following sections describe each part of the processes involved in this framework in detail.

Dataset

Data were collected from two different resources. We obtained weekly dengue cases data based on two different federal territories, which are Kuala Lumpur and Putrajaya from January 2010 to December 2013. This data is obtained from the reports of Disease Control Division, Ministry of Health. Weather data of Kuala Lumpur and Putrajaya is retrieved from the Malaysia Meteorological Department (MMD) for the period of January 2010 to December 2013. Thus, a total 209 weeks of confirmed dengue cases and meteorological data were evaluated in this study. However, approximately 8% of the data are missing in the datasheets of the MMD for the study period. Thus, we obtained the missing data for the same period from the US Weather Channel Interactive, which provides information on Malaysian meteorological data. The data were fitted at the same time with the Putrajaya-Cyberjaya station in Malaysia. Only minimum temperature, maximum temperature, average temperature, minimum humidity, and rainfall were selected, since many studies have emphasized these factors as the most important risk factors for dengue outbreak prediction models as shown in Table1.

Analysis

Weather data from the MMD provide daily weather information, and incidence of dengue cases are published weekly by the MOH. Thus, the data are normalized on weekly basis. Weather and meteorological factors play important roles in the incidence of dengue fever. Thus, the dataset was analyzed, and the relationship between the incidence of dengue cases and weather information was determined weekly using Pearson correlation coefficient (PCC). Pearson product-moment correlation coefficient (sometimes referred to as the PPMCC or PCC or Pearson's r) is a measure of the linear dependence between two variables X and Y (equation 1). This method is an important evaluation method, providing a value between +1 and −1, where 1 indicates the total positive linear correlation, 0 shows no linear correlation, and −1 indicates total negative linear correlation. This measure is widely used in the sciences [50].

Significant Factors Identification

The most significant climate factors are identified based on correlation analysis on the dataset as shown in Table 2. The analysis shows the highest correlation exists between minimum temperature and cumulative rainfall with the incidence of dengue case were determined in different weeks.

Table 2: Correlation between Dengue Incidence Cases and Climate Factors

Minimum temperature and daily rainfall are the most significant dengue weather based risk factors [38,51,52,53].The average minimum temperature can be calculated as follow (equation 2) (2)

where *i* is the number of week from which the average minimum temperature would be calculated. The cumulative rainfall for week *i* can be calculated using equation 3, as follows:

(3)

where *i* is the desired week from which the total rainfall would be calculated, cumulative rainfall week (i) is the final calculation, week (i−1) is the week prior to week (i), and total rainfall for the recent week (0) is the rainfall amount of the current week.

Table 3 illustrates the PCCs between weather variables and incidence of dengue cases. The positive higher numbers, which were underlined and highlighted, showed the highest correlation and coefficients between the weather parameter with the incidence of dengue fever. In Table 3, the results for seven weeks prior to the current week and the optimum value for the average minimum temperature (0.499) are shown.

The highest value for the cumulative rainfall (0.0071) was obtained for two weeks prior to the current week (Table 3).

Table 3: Pearson's Correlation Coefficient between Climatic Factors with Incidence of Dengue Cases

Thus, based on the correlation analysis, the average minimum temperature of week 5 (prior to the current week) and cumulative rainfall for week 2 (prior to the current week) have higher correlation with dengue cases. These two factors will be known as TempeRain Factor (TRF) and

will be used as part of input parameters for dengue outbreak prediction. The combination of the factors is shown in Fig 2.

Fig 2: Components of the TempeRain Factor (TRF)

Cumulative of rainfall for 2 weeks prior to the current week is identified as a significant factor as this tallies with the life cycle of an aedes aegypti which take approximately 2 weeks [38,51,52,53,54,55]. Thus, it clearly shows that dengue outbreak could happen right after the aedes aegypti mosquito completes its life cycle and become an adult.

Predicting using machine learning model

Once the significant factors have been identified, the research is preceded in predicting the number of dengue cases. In order to predict this, we have tested five machine learning models using input factors with TRF and input factors without TRF. Table 4 shows the detailed input factors and descriptions.

Table 4: Input Factors With and Without TRF

Based on the high output result [16,56], we have selected the Support Vector Machine (SVM), RBF Tree, DecisionTable, Native Bayes, and Bayes Net models to evaluate the factors using the WEKA [57]. We used full training set to validate the performance of the model.

Evaluation Metrics

There are some accuracy measures and parameter on the basis of which we can evaluate the performance of classifiers. Along with them there are some accuracy and error measures that are used to find out how far the predicted value is from actual known value [58]. The Confusion Matrix is a useful tool for analyzing how well your classifier can recognize tuples of different classes which is used in WEKA.

The sensitivity and specificity measures can be used to calculate accuracy of classifiers. Sensitivity is also referred to as the true positive rate (the proportion of positive tuples that are correctly identified), while Specificity is the true negative rate (that is, the proportion of negative tuples that are correctly identified).

Equation 4 shows how the accuracy was calculated from Confusion Matrix.

Accuracy = 100* (TP+TN)(TP+FP+TN+FN)

(4)

To demonstrate the error rate, we used Root Mean Squared Error (RMSE) [50,58]. RMSE is also used to identify the strengths in model evaluation. Optimizing RMSE during model calibration may provide small error variance but at the expense of significant model bias [50,59]. This statistic is determined as follows (equation 5):

(5)

where P_i and O_i are known as the experimental and forecasted values, respectively, and n is the total number of test data.

Results And Discussion

Table 5 illustrates the results from the five machine learning models with and without the TRF input factors. Enhanced results and reduced errors were obtained using weather data (as external risk factors for dengue fever outbreak prediction model), by applying machine learning models (as data analyzer), and adding newly identified factors (TRF).

Table 5: Machine Learning Classifier Models Using the Full Training Set (With TempeRain Factor-TRF)

Thus, the proposed factors and machine learning model is beneficial in predicting the number of dengue cases. The results also showed that models that include TRF had higher accuracies compared with those without TRF. The highest accuracy was obtained by SVM with TRF (98.086 %) and with very low RMSE (0.098).

Table 6: Benchmarking with Past Studies

Table 6 shows the accuracy of SVM with TRF compared with other models. The proposed model with TRF achieved highest accuracy of 98.086% compared with other models.

Conclusion and Future Work

Thus, we identified a new significant risk factor called TRF, which combined the average minimum temperature at five weeks prior to the current week and cumulative rainfall at two weeks prior to the current week. The TRF significantly contributed to dengue outbreak prediction. Utilizing accurate and appropriate input factors for outbreak prediction could provide enhanced and more precise results for model output. We used various machine learning models to apply the identified significant factor in predicting the dengue outbreak.

The integration of the factors in the SVM model resulted in significant accuracy of 98.086%. This accuracy showed that using TRF in SVM model outperformed all other outbreak prediction models. Moreover, the RMSE of 0.098 of the proposed system was also lower than in other models. We strongly believe that using the TRF can lead to better outbreak prediction system. In our future studies, we will test the with different prediction models. Moreover, future researches should emphasize in exploration of other hidden and important risk factors for dengue outbreak prediction.

We had some limitations in this research and the most important one is data availability. This is due to the privacy issue and regulation set by the Ministry of Health Malaysia. Although there are many risk factors for dengue outbreak but due to the time and cost and accessibility limitation, we only focused on detailed analysis on temperature and rain risk factors for dengue outbreak.

Abbreviations

FN: False Negative

FP: False Positive

GIS: Geographic Information System

MMD: Malaysian Meteorological Department

MOH: Ministry of Health Malaysia

PCC: Pearson correlation coefficient

PPMCC: Pearson Product-Moment Correlation Coefficient

RBF: Radial basis function

RMSE: Root Mean Squared Error

SVM: Support Vector Machine

TN: True Negative

TP: True Positive

TRF: TempeRain Factor

WHO: World Health Organization

Declarations

Availability of data and material

The completed combined datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

The dengue confirmed case data that support the findings of this study are available in Ministry of Health Malaysia, [http://www.moh.gov.my/index.php/database_stores/store_view/1]

The weather data that support the findings of this study are obtained from Malaysian Meteorological Department. Data are available from the authors upon reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

Research University Grant-Faculty Program (GPF011D–2019).

Authors' contributions

Felestin Yavari Nejad contributed on the related works, experiments and analysis of the studies. Kasturi Dewi Varathan contributed in method and discussions.

Acknowledgements

We would like to thank Research University Grant-Faculty Program (GPF011D–2019) for funding this research.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

References

1.Holmes, E. C., Tio, P. H., Perera, D., Muhi, J., & Cardosa, J. (2009). Importation and co-circulation of multiple serotypes of dengue virus in Sarawak, Malaysia. Virus Research, 143 (1), 1–5. doi:10.1016/j.virusres.2009.02.020

2.Wongkoon, S., Jaroensutasinee, M., & Jaroensutasinee, K. (2012). Development of temporal modeling for prediction of dengue infection in Northeastern Thailand. Asian Pacific Journal of Tropical Medicine, 5 (3), 249–253.

3. Chen, S. C., & Hsieh, M. H. (2012). Modeling the transmission dynamics of dengue fever: Implications of temperature effects. Science of the Total Environment, 431, 385–391. doi:10.1016/j.scitotenv.2012.05.012

4.Chinikar, S., Ghiasi, S. M., Shah-Hosseini, N., Mostafavi, E., Moradi, M., Khakifirouz, S., Rasi Varai, F. S., Rafigh, M., Jalali, T., Goya, M. M., Shirzadi, M. R., Zainali, M. & Fooks, A. R. (2013). Preliminary study of dengue virus infection in Iran. Travel Medicine and Infectious Disease, 5(3), 166–169. doi:10.1016/j.tmaid.2012.10.001

5.Juanarita, J., Azmi, M. N. R., Azhany, Y. & Liza-Sharmini, A. T. (2012), Dengue related maculopathy and foveolitis. Asian Pacific Journal of Tropical Biomedicine, 2(9), 755–756. doi: 10.1016/S2221–1691(12)60223–8

6.WHO/TDR. (2009), Dengue: guidelines for diagnosis, treatment, prevention and control—New edition. Geneva: World Health Organization.

7. Abeyrathna, M. P. A. R., Abeygunawrdane, D. A., Wijesundara, R. A. A. V., Mudalige,V. B., Danaja, M., Kaushalya, M., Sriganesh, L., Madushi, B., Shehan, P. (2016). Dengue Propagation Prediction using Human Mobility. Moratuwa Engineering Research Conference (MERCon). 156–161.

8.World Health Organization (WHO). (2016). Weekly epidemiological record. Factsheet117, 30(91), 349–364. Available: http://www.who.int/mediacentre/factsheets/fs117/en

9.Ibrahim, A., Zin, N. A.M., Ashaari, N. S. (2011). Simulation Model for Predicting Dengue Fever Outbreak. World Academy of Science, Engineering and Technology, International Journal of Computer, Information Science and Engineering 5(11).

10.Kuhn, K., Campbell-Lendrum, D., Haines, A., Cox, J. (2005). Using climate to predict infectious a disease Epidemics. Geneva, Switzerland: World Health Organization (WHO) Document Production Services.

11.Husin, N. A., Mustapha, N., Sulaiman, M. N., & Yaakob, R. (2012). A hybrid model using genetic algorithm and neural network for predicting dengue outbreak. 4th Conference on. doi:10.1109/DMO.2012.6329793

12. Aburas, H. M., Cetiner, B. G. and Sari, M., (2010). Dengue confirmed-cases prediction: A neural network model. Expert Systems with Applications, 37(6), 4256–4260. doi:10.1016/j.eswa.2009.11.077

13.Mathulamuthu, S. M., Asirvadam, V. S., Dass, S. C., Gill, B. S., Loshini, T. (2016). Predicting Dengue Incidences Using Cluster Based Regression on Climate Data. Control System, Computing and Engineering (ICCSCE), 2016 6th IEEE International, 245–250, doi:

10.1109/ICCSCE.2016.7893579

14.Soemsap, T., Wongthanavasu, S., Satimai, W. (2014) Forecasting Number of Dengue Patients Using Cellular Automata Model. Proceedings of the International Electrical Engineering Congress, doi: 10.1109/iEECON.2014.6925876.

15.Tazkia, R. A. K., Narita, V., Nugroho, A. S. (2016). Dengue Outbreak Prediction for GIS based Early Warning System. International Conference on Science in Information Technology (ICSITech), doi: 10.1109/ICSITech.2015.7407789

16.Rahmawati, D. & Huang, Y. P. Using C-support Vector Classification to Forecast Dengue Fever Epidemics in Taiwan. (2016). International Conference on System Science and Engineering (ICSSE) National Chi Nan University, Taiwan; July 7–9. 978–1–4673–8966–2/16

17.Hii YL,. (2013). Climate and Dengue Fever: Early warning based on temperature and Rainfall. Umeå University Medical Dissertations. New Series No 1554, ISSN 0346–6612, ISBN 978–91–7459–589–5.

18.Thitiprayoonwongse, D., Suriyaphol, P., Soonthornphisaj, N. (2012). Data Mining of Dengue Infection Using Decision Tree. Latest Advances in Information Science and Applications, Entropy, 2, 2 154–159. doi: 10.1109/ICSITech.2015.7407789

19.Tanner, L., Schreiber, M., Low, JGH., Ong, A., Tolfvenstam, T., et al. (2008). Decision Tree Algorithms Predict the Diagnosis and Outcome of Dengue Fever in the Early Phase of Illness. PLoS Negl Trop Dis 2(3): e196. doi:10.1371/journal.pntd.0000196

20.Ibrahim, F., Faisal, T., Mohamad Salim, M. I. & Taib, M. N., (2010). Non-invasive diagnosis of risk in dengue patients using bioelectrical impedance analysis and artificial neural network. Medical & Biological Engineering & Computing, 48(11), 1141–1148. doi: 10.1007/s11517–010–0669-z

21.Bhatt S., Gething PW., Brady OJ., Messina JP., Farlow AW., Moyes CL. et.al. (2013). The global distribution and burden of dengue. Nature.496:504–5077. doi:10.1038/nature12060.

22. Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG et al. (2012).

Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl Trop Dis. 2012;6(8):e1760. doi:10.1371/journal.pntd.0001760.

23.Ministry of Health Malaysia (MOH). Dengue Fever And Chikungkunya Situation, Retrieved from http://www.moh.gov.my/index.php/database_stores/store_view/17, Available [Access March 2017]

24.World Health Organization (WHO), Distribution of dengue, worldwide, (2018). Average number of suspected or confirmed dengue cases reported to WHO, 2010–2016. Retrieved from www.who.int/denguecontrol/epidemiology/en, Available [Access March 2019]

25.Andrick,B., Clark, B., Nygaard, K., Logar, A. and Penaloza, M. (1997). Infectious Disease

and Climate Change: Detecting Contributing Factors and Predicting Future Outbreaks. Geoscience and Remote Sensing, 1997. IGARSS ’97. doi: 10.1109/IGARSS.1997.609159

26.Korstanje, M., George, B., (2016). Media constructions of fear in the outbreak of an epidemic disease: The case of dengue fever in Argentina, International Journal of Emergency Services, 5(1), 95–104, doi: 10.1108/IJES–01–2016–0001

27.Pinto, E., Coelho, M., Oliver, L., and Massad, E. (2011). The influence of climate variables on dengue in Singapore. International Journal of Environmental Health Research 21(6): 415–426. doi: 10.1080/09603123.2011.572279.

28.Mathur, N., Asirvadam V. S., Sarat. C. (2016). Generating Vulnerability Maps of Dengue Incidences for Petaling District in Malaysia, 12th International Colloquium on Signal Processing & its Applications (CSPA2016). doi: 10.1109/CSPA.2016.7515836

29.Loshini T., Vijanth S. Asirvadam, Sarat C. Dass. Balvinder S. Gill. Predicting Localized Dengue Incidences using Ensemble System Identification. (2015) International Conference on Computer, Control, Informatics and Its Applications (IC3INA). pp:6–11.

doi: 10.1109/IC3INA.2015.7377737

30.Burattini, MN., Chen, M., Chow, A., Coutinho, FAB., Goh, KT., Lopez, LF., Ma, S.,

Massad, E., (2008). Modelling the control strategies against dengue in Singapore. Epidemiol Infect. 136(3), 309–319. doi:: 10.1017/S0950268807008667

31.Mochammad, C. R., Achmad, B., Tri, H. (2016). Comparison of Montecarlo Linear and Dynamic Polynomial Regression in Predicting Dengue Fever Case. Knowledge Creation and Intelligent Computing (KCIC). doi: 10.1109/KCIC.2016.7883649.

32.Jesavel A. Iguchi, Xerxes T. Seposo and Yasushi Honda. (2018). Meteorological factors affecting dengue incidence in Davao, Philippines, BMC Public Health (2018) 18:629. doi: 10.1186/s12889–018–5532–4

33.Paul KK, Dhar-Chowdhury P, Haque CE, Al-Amin HM, Goswami DR, Kafi MAH, et al. (2018). Risk factors for the presence of dengue vector mosquitoes, and determinants of their prevalence and larval site selection in Dhaka, Bangladesh. PLoS ONE 13(6): e0199457. doi:

10.1371/journal.pone.0199457

34.Hu Suk Lee, Hung Nguyen-Viet, Vu Sinh Nam, Mihye Lee, Sungho Won, Phuc Pham Duc and Delia Grace. (2017). Seasonal patterns of dengue fever and associated climate factors in 4 provinces in Vietnam from 1994 to 2013. BMC Infectious Diseases (2017) 17:218. Doi: 10.1186/s12879–017–2326–8

35.Datoc, H. I., Caparas, R., Caro, J. (2016). Forecasting and Data Visualization of Dengue spread in the Philippine Visayas Island group. 7th International Conference on Information, Intelligence, Systems & Applications (IISA), doi: 10.1109/IISA.2016.7785420.

36.Xiang, J., Hansen, A., Liu, Q., Liu, X., Tong, M. X., Sun, Y., & Weinstein, P. (2016). Association between dengue fever incidence and meteorological factors in Guangzhou, China, 2005–2014. Environmental Research, 153, 17–26. doi.10.1016/j.envres.2016.11.009

37.Hai-Yan Xu, Fu, X., Lee, L. K. H., Ma, S., Goh, K. T., Wong, J., & Lim, C. L. (2014). Statistical modeling reveals the effect of absolute humidity on dengue in Singapore. PLoS Negl Trop Dis, 8(5), e2805. doi: 10.1371/journal.pntd.0002805

38.Lung, C. C., Hwa L. Y., (2014). Impact of meteorological factors on the spatiotemporal patterns of dengue fever incidence. Environment International 73: 46–56.

39.Maha Bouzid, Felipe J Colón-González, Tobias Lung, Iain R Lake and Paul R Hunter. (2014). Climate change and the emergence of vector-borne diseases in Europe: case study of dengue fever. BMC Public Health 2014 14:781. doi:10.1186/1471–2458–14–781

40.Felipe J., Colón-González, Fezzi, C., Lake, I. R., & Hunter, P. R. (2013). The effects of weather and climate change on dengue. PLoS Negl Trop Dis, 7(11), e2503. doi: 10.1371/journal.pntd.0002503

41. Cheong, Y. L., Burkart, K., Leitão, P. J., & Lakes, T. (2013). Assessing weather effects on

dengue disease in Malaysia. International journal of environmental research and public health, 10(12), 6319–6334. doi:10.3390/ijerph10126319

42. Dom, N. C., Hassan, A. A., Latif, Z. A., & Ismail, R. (2013). Generating temporal model using climate variables for the prediction of dengue cases in Subang Jaya, Malaysia. Asian Pacific Journal of Tropical Disease, 3(5), 352–361. doi: 10.1016/S2222–1808(13)60084–5, Chicago

43.Hii YL, Zhu H., Ng N., Ng LC., Rocklöv J. (2012). Forecast of Dengue Incidence Using Temperature and Rainfall. PLoS Negl Trop Dis 6(11): e1908. doi:10.1371/journal.pntd.0001908

44.Zhaoxia Wang, Chan, H. M., Hibberd, M. L., & Lee, G. K. K. (2012). Delayed Effects of Climate Variables on Incidence of Dengue in Singapore during 2000–2010. APCBEE Procedia, 1, 22–26. doi: 10.1016/j.apcbee.2012.03.005

45.Rachel, L., Bailey, T. C., Stephenson, D. B., Graham, R. J., Coelho, C. A. S., Carvalho, M. Sá., Barcellos, C. (2011). Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. Computers & Geosciences. 37(3), 371–381. doi:10.1016/j.cageo.2010.01.008.

46.Halide Halmar. (2010). Assessing Quality and Value of Predictive Models for Dengue Hemorrhagic Fever Epidemics. Nova Publisher, New York.

47.Cetiner, B. G., Sari, M., & Aburas, H. M. (2009, May). Recognition of dengue disease patterns using artificial neural networks. In 5th International Advanced Technologies Symposium (IATS'09) 359–362.

48.Rachata,N., Charoenkwan, P., Yooyativong, T., Chamnongthal, K., Lursinsap, C. & Higuchi, K. (2008). Automatic Prediction System of Dengue Haemorrhagic-Fever Outbreak Risk by Using Entropy and Artificial Neural Network. Communications and Information

Technologies, 2008 (ISCIT). pp210–214. doi: 10.1109/ISCIT.2008.4700184.

49.Promprou, S., Jaroensutasinee, M., & Jaroensutasinee, K. (2005). Climatic Factors Affecting Dengue Haemorrhagic Fever Incidence in Southern Thailand.

50.Moriasi D. N., Arnold, J. G., Van Liew, M. W. Bingner, R. L., Harmel, R. D., Veith, T. L. (2007). Model Evaluation Guidelines For Systematic Quantification Of Accuracy In Watershed Simulations. Transactions of The Asabe. 50(3): 885–900. doi: 10.13031/2013.23153.

51.Christophers, S. R. (1960). Aedes aegypti (L.) the yellow fever mosquito. Its life history.In: Bionomics and Structure. Cambridge Univ. Press, Cambridge. 133, (3463), 1473–1474. doi: 10.1126/science.133.3463.1473-a

52.Yang HM, Macoris MLG, Galvani KC, Andrighetti MTM, Wanderley DMV.(2009). Assessing the effects of temperature on the population of Aedes aegypti, the vector of dengue. Epidemiol Infect 137: 1188–1202. doi: 10.1017/S0950268809002040.

53.Ahmad R, Wong YC, Zamre I, Lee HL, Zurainee MN. (2009). The effect of extrinsic

incubation temperature on development of dengue serotype 2 and 4 viruses in Aedes aegypti (L.). Southeast Asian J Trop Med Public Health 40(5): 942–650.

54.Watts, D. M., Burke, D. S., Harrison, B. A., Whitmire, R. E., & Nisalak, A. (1987). Effect of

temperature on the vector efficiency of Aedes aegypti for dengue 2 virus. The American

journal of tropical medicine and hygiene, 36(1), 143–152. doi: 10.4269/ajtmh.1987.36.143

55.Chan, M., Johansson, M. A. (2012) The Incubation Periods of Dengue Viruses. PLoS ONE 7(11): e50972. doi:10.1371/journal.pone.0050972

56.Fathima, S. and Hundewale, N. (2011). Comparison of Classification Techniques-SVM and Naives Bayes to predict the Arboviral Disease-Dengue, International Conference on Bioinformatics and Biomedicine Workshops, doi: 10.1109/BIBMW.2011.6112426

57.Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I.H (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations. 11(1): 10–18.

58.Nasa.Ch and Suman (2012). Evaluation of Different Classification Techniques for WEB Data. International Journal of Computer Applications (0975—8887). 52(9): 34–40.

59.Boyle, D. P., H. V. Gupta, & Sorooshian, S. (2000). Toward improved calibration of

hydrologic models: Combining the strengths of manual and automatic methods. Water Resources Res. 36(12): 3663–3674. doi: 10.1029/2000WR900207

60.Ahmad R, Suzilah I, Wan Najdah WMA, Topek O, Mustafakamal I, Lee HL (2018) Factors

determining dengue outbreak in Malaysia. PLoS ONE 13(2): e0193326. https://doi.org/10.1371/journal.pone.0193326

61.Saha, S. (2016). Combined committee machine for classifying dengue fever. In Microelectronics, Computing and Communications (MicroCom), 2016 International Conference on. pp. 1–6. doi: 10.1109/MicroCom.2016.7522585.

Tables

Table 1: Risk Factor for Dengue Outbreak Prediction Model

Author(s)	Year	Geographical Data Used	Temperature			Humidity	Rainfall					Mean
			Min	Avg	Max	Relative (Mean)	Cumulative Rainfall	Total Rainfall	Max 24-h Rainfall	Max 1-H Rainfall	Bi-Weekly	
Jesavel A. *et al.* [32]	2018	Philippines		*			*					
Paul KK. *et al.* [33]	2018	Bangladesh		*		*		*				
Hu S.L *et al.* [34]	2017	Vietnam		*				*				
Datoc *et al.* [35]	2016	Philippine		*		*		*				
Xiang j. *et al.* [36]	2016	China	*		*	*						*
Hai-Yan Xu *et al.* [37]	2014	Singapore	*	*	*	*		*				*
Lung C.C *et al.* [38]	2014	Taiwan	*	*	*			*	*	*	*	
Maha B. et al. [39]	2014	Europe	*		*	*		*				
Felipe J. *et al.* [40]	2013	Mexico	*		*			*				
Cheong Y.L *et al.* [41]	2013	Malaysia ,	*	*	*	*					*	*
Hii Yien Lin [17]	2013	Singapore		*			*					
Dom N.C *et al.* [42]	2013	Malaysia		*		*		*				
Hii Yien Ling *et al.* [43]	2012	Singapore		*			*					
Zhaoxia Wang *et al.*[44]	2012	Singapore	*			*		*				
Chen *et al.* [3]	2012	Bangladesh	*		*	*		*				
Husin, N. A *et al.* [11]	2012	Malaysia						*				
Rachel *et al.* [45]	2011	Brazil		*		*		*				
Aburas, H. M. *et*	2010	Singapore		*		*		*				

al. [12]												
Halide Halmar [46]	2010	Indonesia	*	*	*	*		*				
Cetiner *et al.* [47]	2009	Turkey		*		*		*				
Rachata *et al.* [48]	2008	Thailand	*	*	*	*		*				
Promprou S. *et al.* [49]	2005	Thailand	*	*	*	*		*				
Total			11	16	10	15	3	17	1	1	2	3

Table 2: Correlation between Dengue Incidence Cases and Climate Factors

Temperature			Mean relative Humidity	Rainfall
Minimum Temperature	Mean Temperature	Maximum Temperature		
0.447	0.339	0.316	-0.176	**-0.020**

Table 3: Pearson's Correlation Coefficient between Climatic Factors with Incidence of Dengue Cases

	Average Minimum Temperature	Cumulative Rainfall
Current Week	0.447	–0.0201
1 Week Prior	0.465	0.0065
2 Week Prior	0.480	**0.0071**
3 Week Prior	0.494	–0.0005
4 Week Prior	0.498	–0.0123
5 Week Prior	**0.499**	–0.0139
6 Week Prior	0.489	–0.0045
7 Week Prior	0.476	0.0020

Table 4: Input Factors With and Without TRF

Input Factors without TRF		Input Factors with TRF	
Type	Parameter Description	Type	Parameter Description
Weather Factors	Minimum temperature (°C)	Weather Factors	
	Mean temperature (°C)		Mean temperature (°C)
	Maximum temperature (°C)		Maximum temperature (°C)
	Mean relative humidity (%)		Mean relative humidity (%)
	Cumulative of rainfall (mm)		
		TRF Factors	**Average of minimum temperature 5 weeks before the current week (°C)**
			Cumulative of rainfall for 2 weeks prior to the current week (mm)

Table 5: Machine Learning Classifier Models Using the Full Training Set (With TempeRain Factor-TRF)

Models		Accuracy (%)	Root Mean Squared Error (RMSE)
SVM	With TRF	**98.086**	0.098
	Without TRF	83.254	0.290
RBF Tree	With TRF	85.168	0.245
	Without TRF	80.861	0.268
Decision Table	With TRF	83.254	0.267
	Without TRF	80.861	0.283
Naïve Bayes	With TRF	83.732	0.280
	Without TRF	78.947	0.293
Bayes Net	With TRF	83.254	0.263
	Without TRF	80.861	0.275

Table 6: Benchmarking with Past Studies

Authors	Year	Model	Accuracy (%)
Ahmad R. *et al.* [60]	2018	Correlation and Autoregressive Distributed Lag Model	84.9
Tazkia *et al.* [15]	2016	GIS based by Naïve Bayes	97.05
Saha S. [61]	2016	Multilayer Perceptron and Support Vector Machine	96.56
Rahmawati and Huang [16]	2016	C-SVM Kernel and RBF	90.5
Hii Yien Ling [17]	2013	Poisson Multivariate Regression Models	90
Thitiprayoonwongse *et al.*[18]	2012	Decision Tree	96.7
Aburas *et al.* [12]	2010	Artificial Neural Networks	82.39
Ibrahim *et al.* [20]	2010	Bioelectrical Impedance Analysis (BIA) and Artificial neural network (ANN).	96.27
Rachata *et al.* [48]	2008	Automatic Prediction System by Using Entropy and Artificial Neural Network	85.92
Our Proposed Model		**SVM using TempeRain Factor (TRF)**	**Accuracy = 98.086**

Figures

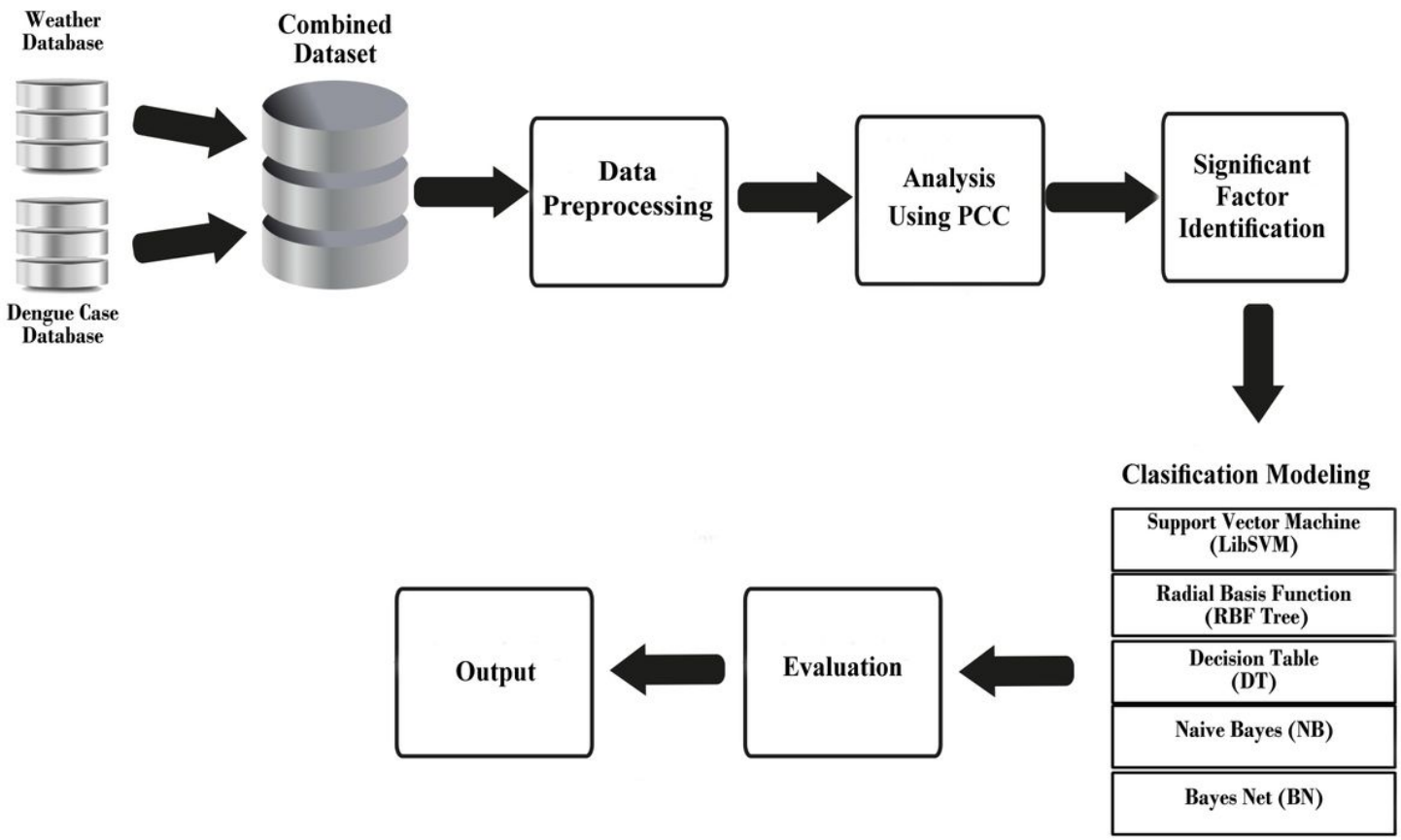

Figure 1

Conceptual Framework on Identifying Significant Climate Factors in Dengue Outbreak Prediction

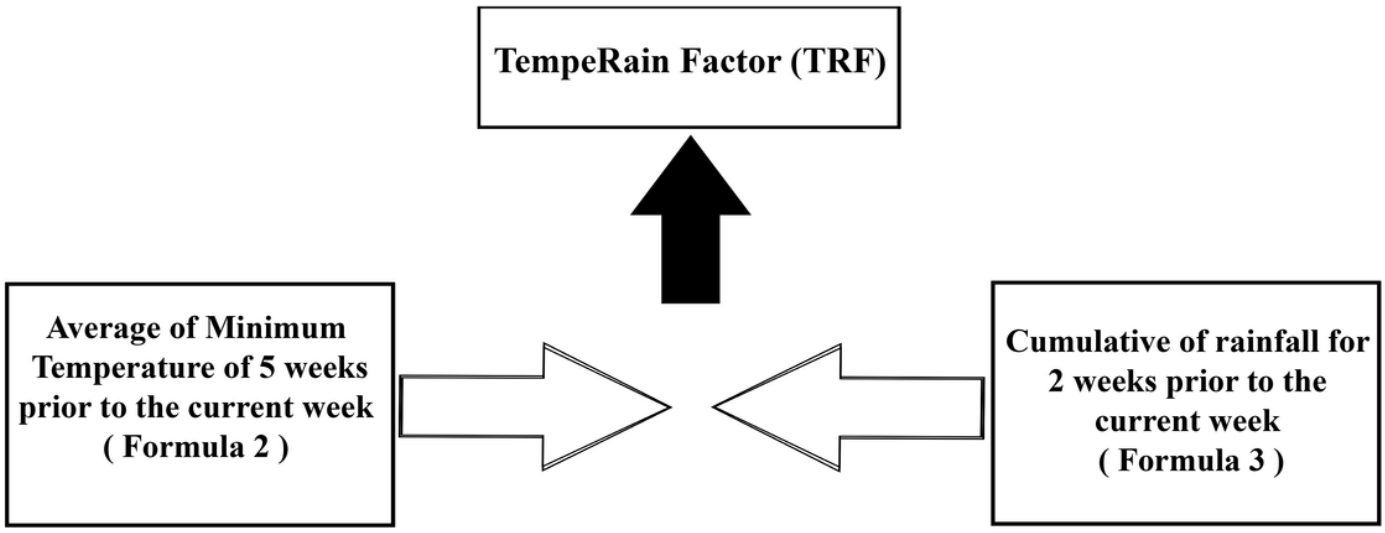

Figure 2

Components of the TempeRain Factor (TRF)