

RESEARCH ARTICLE

Open Access



# Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction

Felestin Yavari Nejad and Kasturi Dewi Varathan\*

## Abstract

**Background:** Dengue fever is a widespread viral disease and one of the world's major pandemic vector-borne infections, causing serious hazard to humanity. The World Health Organisation (WHO) reported that the incidence of dengue fever has increased dramatically across the world in recent decades. WHO currently estimates an annual incidence of 50–100 million dengue infections worldwide. To date, no tested vaccine or treatment is available to stop or prevent dengue fever. Thus, the importance of predicting dengue outbreaks is significant. The current issue that should be addressed in dengue outbreak prediction is accuracy. A limited number of studies have conducted an in-depth analysis of climate factors in dengue outbreak prediction.

**Methods:** The most important climatic factors that contribute to dengue outbreaks were identified in the current work. Correlation analyses were performed in order to determine these factors and these factors were used as input parameters for machine learning models. Top five machine learning classification models (Bayes network (BN) models, support vector machine (SVM), RBF tree, decision table and naive Bayes) were chosen based on past research. The models were then tested and evaluated on the basis of 4-year data (January 2010 to December 2013) collected in Malaysia.

**Results:** This research has two major contributions. A new risk factor, called the TempeRain factor (TRF), was identified and used as an input parameter for the model of dengue outbreak prediction. Moreover, TRF was applied to demonstrate its strong impact on dengue outbreaks. Experimental results showed that the Bayes Network model with the new meteorological risk factor identified in this study increased accuracy to 92.35% for predicting dengue outbreaks.

**Conclusions:** This research explored the factors used in dengue outbreak prediction systems. The major contribution of this study is identifying new significant factors that contribute to dengue outbreak prediction. From the evaluation result, we obtained a significant improvement in the accuracy of a machine learning model for dengue outbreak prediction.

**Keyword:** Risk factor, Dengue, Outbreak prediction model, TempeRain factor

## Background

Pandemic infectious diseases are spreading in many geographical areas. The World Health Organisation (WHO)

has reported that dengue fever is one of the most important mosquito-borne and deadliest infectious diseases which have been caused by the dengue virus. Accordingly, this disease is a threat and poses severe risk to human populations in tropical and subtropical regions [1–7]. Member states in the three WHO regions regularly reported an increase in the annual number of cases from 2.2 million in 2010 to 3.2 million in 2015 [8]. A recent

\*Correspondence: [kasturi@um.edu.my](mailto:kasturi@um.edu.my)  
Department of Information System, Faculty of Computer Science  
and Information Technology, University of Malaya, Kuala Lumpur, Malaysia



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

study from WHO indicated that 390 million dengue infections occur annually (95% credible interval of 284–528 million); among which, 96 million (67–136 million) are manifested clinically with any severity of the disease [9, 10]. There is a research that works on how dengue has changed global threat by vector-borne disease and will help decision-makers worldwide to better prepare for and respond to future changes in dengue risk for the years 2020, 2050 and 2080 [11].

As of December 2018, the Ministry of Health (MOH) of Malaysia has recorded approximately 80,615 dengue cases with 147 deaths compared with 19,884 cases in December 2011 with 36 deaths [12]. The number of cases increased approximately fourfold. By the end of March 2019, 39,805 cases of dengue with 64 deaths were reported in Malaysia compared with 16,917 cases with 34 deaths in March 2018 [13].

The alternative solutions would be to prevent and control the outbreak of this disease. One of the ways in achieving this is by having a good prediction on the existence of the outbreak. This kind of predictions helps the higher authorities as well as individuals to devise plans and handle the situation in an efficient and effective manner [14, 15]. The accuracy of a prediction system for outbreaks is the primary and important concern for controlling dengue fever [14]. Therefore, establishing related risk factors is critical for prediction systems [16]. Given that climate factors play a key role in this disease, identifying the relation between weather information and dengue outbreak incidence is a major task in establishing an accurate prediction system for future outbreaks [17–19]. In the current study, important climatic risk factors, such as temperature, relative humidity and rainfall amount, were examined. The current accuracy for prediction systems based on climate factors ranges from 82.39 to 90.5% [16, 20–25].

This research aims to describe the dengue prediction system accuracy and the level of risk factors that contribute to a dengue outbreak prediction system and identify the associations amongst new climate risk factors. The detailed factors are then used as inputs for predicting dengue outbreaks.

### Related works

Various early warning and monitoring systems are currently implemented to monitor dengue outbreaks worldwide. Dengue prediction models have been previously investigated, but some of these models still exhibit limitations in achieving high accuracy in dengue outbreak prediction [14, 15]. Different models and techniques have been integrated into the design of several models for predicting dengue outbreaks. A number of studies have also established prediction models for dengue outbreaks using artificial neural networks [16].

Hybrid models have been used in outbreak prediction research. A hybrid model is an example of an integrated model, and many models based on genetic algorithms are available to determine the weight in a neural network model [14, 17–19, 25, 26]. In Singapore, researchers found significant correlated dengue cases with climatic variables by using a Poisson regression model [27]. One researcher [22] developed a dengue outbreak prediction system in Singapore and obtained 90% accuracy. There are research that was established by using decision tree in prediction system [23]. One research in Indonesia was concerned with dengue outbreak prediction using a GIS-based early warning system [20]. Another study from the National Taipei University of Technology used C-support vector classification to forecast dengue fever epidemics in Taiwan, and the accuracy of radial basis function (RBF) model was 90.5% [21].

Vulnerability maps of dengue incidences have been generated in Malaysia, resulting in the development and implementation of visualised and predictive modelling using geographic information systems (GIS) for dengue fever in Selangor, Malaysia [28]. There are different models of dengue outbreak prediction systems in Malaysia have achieved different accuracies [16, 25]. In 2015, [29] predicted localised dengue incidences in Malaysia using an ensemble system for identification and found that ensemble models exhibit better prediction power than a single model [29]. The prediction of dengue outbreaks is crucial worldwide because this infectious disease remains as a major issue in many countries [14, 26, 30, 31]. Table 1 lists studies on different models of dengue outbreak prediction with distinct climatic risk factors. The asterisk (\*) in the columns of the table denotes the risk factors used in different studies.

Most studies on dengue fever were conducted in Asian countries, such as Malaysia, Singapore, Taiwan, Indonesia, Bangladesh and Thailand, are critical areas for dengue fever. Most studies have shown that temperature and rainfall directly and significantly affect dengue outbreaks [15, 18, 25, 26, 30, 31].

Moreover, changing climatic factors, such as increasing temperature, rainfall and humidity, are the most influential driving forces of dengue virus transmission [31]. One study correlated dengue cases with climatic variables in the city of Singapore and the model for dengue cases was considered the dependent variable; meanwhile, climatic variables, such as rainfall, maximum and minimum temperatures and relative humidity, were considered independent variables [27]. On the basis of the grade of each risk factor used in the 22 references listed in Table 1, most studies primarily used total rainfall (17 studies), average temperature (16 studies), relative humidity (15 studies), minimum temperature (11 studies)

**Table 1** Risk factors used in different researches for dengue outbreak prediction models from 2005 to 2018

| References | Technique  | Year | Geographical data used | Temperature |     | Humidity |                 | Rainfall            |                  | Mean |
|------------|--|------|------------------------|-------------|-----|----------|-----------------|---------------------|------------------|------|
|            |  |      |                        | Min         | Avg | Max      | Relative (mean) | Cumulative rainfall | Max 1-H rainfall |      |
| [32]       | Wavelet coherence analysis/quasi-Poisson regression combined with distributed lag nonlinear model (DLNM) | 2018 | Philippines            | *           |     |          |                 |                     | *                |      |
| [33]       | Generalized linear model   | 2018 | Bangladesh             | *           |     | *        |                 |                     | *                |      |
| [34]       | Negative binomial regression (NBR)/generalized estimating equation (GEE)                                 | 2017 | Vietnam                | *           |     | *        |                 |                     | *                |      |
| [35]       | Artificial neural network (ANN)  | 2016 | Philippine             | *           |     | *        |                 |                     | *                |      |
| [36]       | Distributed lag non-linear models (DLNM)/generalised estimating equation models (GEE)                    | 2016 | China                  | *           | *   | *        |                 |                     |                  | *    |
| [37]       | Spearman rank correlation/distributed lag non-linear model (DLNM)  | 2014 | Singapore              | *           | *   | *        |                 |                     | *                | *    |
| [38]       | Distributed lag nonlinear model (DLNM) and Markov random fields  | 2014 | Taiwan                 | *           | *   | *        |                 | *                   | *                | *    |
| [39]       | Generalized additive model (GAM)   | 2014 | Europe                 | *           |     | *        |                 |                     | *                |      |
| [40]       | Generalized additive model (GAM)   | 2013 | Mexico                 | *           |     | *        |                 |                     | *                |      |
| [41]       | Poisson generalized additive model/distributed non-linear lag model (DLMN)                               | 2013 | Malaysia,              | *           | *   | *        |                 |                     | *                | *    |
| [22]       | Poisson multivariate regression models   | 2013 | Singapore              | *           |     |          |                 | *                   |                  |      |
| [42]       | Autoregressive integrated moving average (ARIMA)   | 2013 | Malaysia               | *           |     | *        |                 |                     | *                |      |
| [43]       | Poisson multivariate regression  | 2012 | Singapore              | *           |     | *        |                 | *                   |                  |      |
| [44]       | Spearman's rank correlation coefficient (SRCC)   | 2012 | Singapore              | *           |     | *        |                 |                     | *                |      |
| [3]        | Vector-host transmission model   | 2012 | Taiwan                 | *           |     | *        |                 |                     | *                |      |
| [14]       | Neural network and genetic algorithm   | 2012 | Malaysia               | *           |     | *        |                 |                     | *                |      |
| [45]       | Generalised linear model (GLM)/Bayesian framework using Markov chain Monte Carlo (MCMC)                  | 2011 | Brazil                 | *           |     | *        |                 |                     | *                |      |
| [16]       | Artificial neural networks (ANN)   | 2010 | Singapore              | *           |     | *        |                 |                     | *                |      |
| [46]       | Multiple regression and discriminant analysis techniques/Peirce skill score                              | 2010 | Indonesia              | *           | *   | *        | *               | *                   | *                | *    |

**Table 1** (continued)

| References | Technique   | Year | Geographical data used |     |     | Temperature     |                | Humidity            |                | Rainfall          |                  |           | Mean |
|------------|---|------|------------------------|-----|-----|-----------------|----------------|---------------------|----------------|-------------------|------------------|-----------|------|
|            |   |      | Min                    | Avg | Max | Relative (mean) | Relative (max) | Cumulative rainfall | Total rainfall | Max 24-h rainfall | Max 1-H rainfall | Bi-weekly |      |
| [47]       | Artificial neural networks (ANN)  | 2009 |                        | *   | *   | *               | *              | *                   |                |                   |                  | *         |      |
| [48]       | Entropy and artificial neural network   | 2008 | *                      | *   | *   | *               | *              | *                   |                |                   |                  | *         |      |
| [49]       | Kolmogorov-Sminov test/Pearson's correlation coefficient/stepwise regression techniques | 2005 | *                      | *   | *   | *               | *              | *                   |                |                   |                  | *         |      |
|            | Total   | 11   | 10                     | 15  | 3   | 17              | 1              | 1                   | 1              | 2                 | 3                |           |      |

and maximum temperature (10 studies) as inputs of prediction models. However, none of the studies focused on the detailed analysis of the factors nor investigated the detailed relationship that can exist amongst factors.

**Methods**

This section explains the methodology used for this research, including the dataset used, the analysis process, the newly identified integrated input factors, the evaluation with machine learning models and the evaluation method. Figure 1 illustrates the conceptual framework of our research.

**Dataset**

Data are retrieved from two official sources. The dengue fever incident and confirmed cases has published weekly on Ministry of Health Malaysia (MOH) portal and the report of location and number of dengue confirmed case is available and accessible on weekly basis from [12]. In order to access each of the files, the following link provides the access. However, the report number based on the weeks must be stated explicitly.

[http://www.moh.gov.my/index.php/database\\_stores/attach\\_download/337/report number](http://www.moh.gov.my/index.php/database_stores/attach_download/337/report%20number)

For example to retrieve the report no. 234 (week10, 2012):

[http://www.moh.gov.my/index.php/database\\_stores/attach\\_download/337/234](http://www.moh.gov.my/index.php/database_stores/attach_download/337/234)

Besides that, this report can also be obtained via a simple search in any search engine by using the following search terms together with required week number and year:

“SITUASI SEMASA DEMAM DENGGI DI MALAYSIA Bagi Minggu *week number/year*”

*English Translation:*

Situation of Dengue Fever in Malaysia for *week number/year*

Moreover, the climatic data are obtained from Malaysian Meteorological Department (MMD) [50]. However, the processed data is available upon reasonable request from the authors.

Data were collected from two sources. We obtained weekly data on dengue confirmed cases based on two federal territories, namely, Kuala Lumpur (Wilayah Persekutuan Kuala Lumpur) and Putrajaya, from January 2010 to December 2013. The weather data of Kuala Lumpur and Putrajaya were retrieved from Malaysian Meteorological Department (MMD) for the period of January 2010 to December 2013 [50]. Thus, 209 weeks of confirmed

dengue cases and meteorological data were evaluated in this study. However, approximately 8% of the data were missing in the MMD datasheets for the study period. Thus, we obtained the missing data for this period from the US Weather Channel Interactive (<https://weather.com>), which also provides Malaysian meteorological data. The data were fitted simultaneously with the Putrajaya–Cyberjaya Station in Malaysia. Only minimum temperature, maximum temperature, average temperature, minimum humidity and rainfall were selected because many studies have emphasised that these factors are the most important risk factors for dengue outbreak prediction models, as shown in Table 1. Figure 2 illustrates two plots of data from January 2010 to December 2013.

The data are combined and cleaned accordingly. The preprocessed data are analysed, and new detailed factors are identified. The factors are then integrated and fed as integrated inputs to different machine learning models and evaluated. The following sections provide a detailed description of each process involved in this framework.

**Analysis**

Weather data from MMD provide daily weather information, and the incidence of dengue cases is published weekly by MOH. Thus, data were normalised and classified into two levels namely, ‘low risk’ and ‘high risk’, on a weekly basis [51]. Weather and meteorological factors play important roles in the incidence of dengue fever. Thus, the dataset was analysed, and the relationship between the incidence of dengue cases and weather information was determined every week using the Pearson correlation coefficient (PCC) [52].

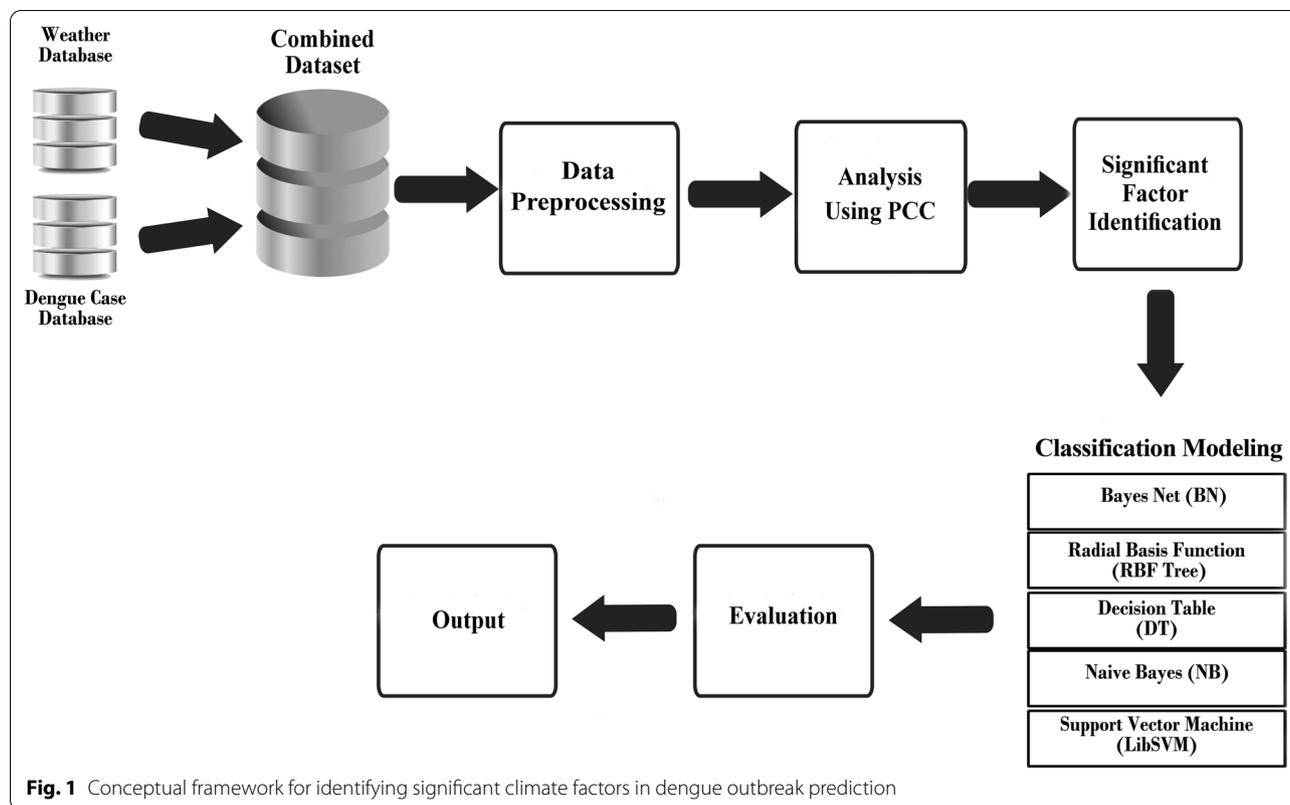
$$R = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{N}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{N}\right) \left(\sum y_i^2 - \frac{(\sum y_i)^2}{N}\right)}} \quad (1)$$

**Identification of significant factors**

The most significant climate factors were identified based on the correlation analysis of the dataset, as shown in Table 2. The analysis result indicated that the highest correlation exists between minimum temperature and dengue incidence, followed by cumulative rainfall and the incidence of dengue cases determined in different weeks.

Minimum temperature and daily rainfall are the most significant dengue weather-based risk factors [36, 53–56]. The average minimum temperature can be calculated as follows (Eq. 2):

$$\text{Average Min Temperature Week}(i) = \frac{\text{Minimum Temperature (Current Week)} + \sum_{n=0}^5 \text{Min Temperature [Week}(i - n)]}{6} \quad (2)$$



where  $i$  is the number of weeks from which the average minimum temperature and  $[Week(i - n)]$  is the minimum temperature of the prior weeks to the current week plus minimum temperature of current week  $[n=0]$ . To find average, the result divided by 6 [5 weeks before plus current week].

The cumulative rainfall for week  $i$  can be calculated using Eq. 3, as follows:

$$\text{Cumulative Rain fall Week}(i) = \sum_{n=0}^1 [\text{Total Rainfall Week}(i - n)] \tag{3}$$

where  $i$  is the desired week from which the total rainfall will be calculated, cumulative rainfall week  $(i)$  is the final calculation and week  $(i - n)$  is the week prior to week  $(n)$ .

Table 3 provides the PCCs between the weather variables and the incidence of dengue cases. The underlined and highlighted high positive numbers showed the highest correlation and coefficients between weather parameters and the incidence of dengue fever. Table 3 presents the results for 7 weeks prior to the current week and the optimum value for the average minimum temperature (0.499).

The highest value for cumulative rainfall (0.0071) was obtained for 2 weeks prior to the current week (Table 3).

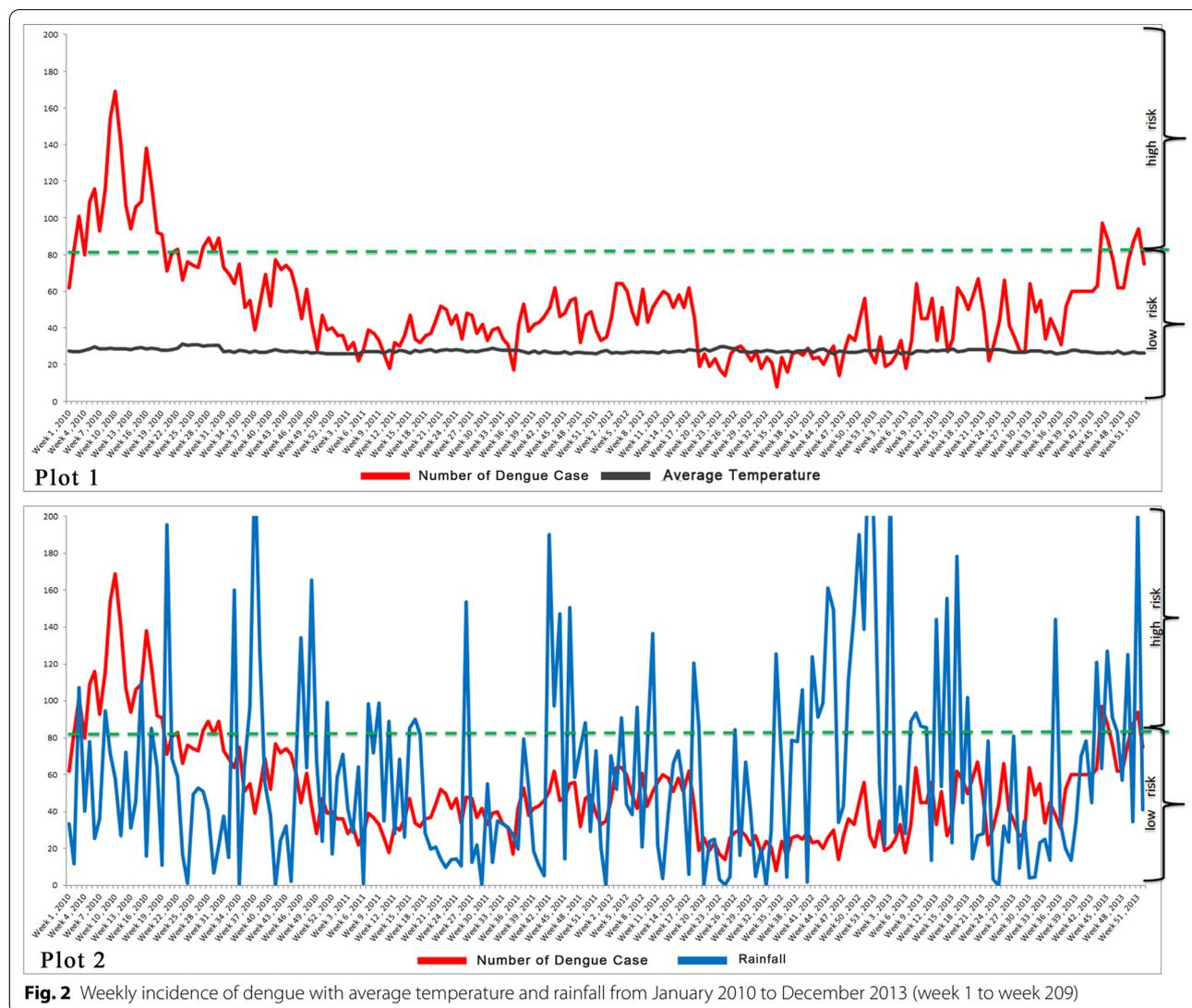
Thus, the average minimum temperature of Week 5 (plus the current week) and the cumulative rainfall for Week 2 (prior to the current week) exhibit high correlation with dengue cases in accordance with the correlation analysis. The two factors will be regarded as TRF and used as input parameters for dengue outbreak risk level

prediction. The combination of factors is shown in Fig. 3.

The cumulative rainfall for 2 weeks prior to the current week is identified as a significant factor because it coincides with the life cycle of an *Aedes aegypti* mosquito, i.e. approximately 2 weeks. Thus, this shows that there is a possibility that it may happen right after an *A. aegypti* mosquito completes its life cycle and becomes an adult [38, 53–58].

**Prediction using machine learning models**

Once significant factors have been identified, the research proceeded towards predicting the risk incidence level of dengue fever. We considered high risk



**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   |

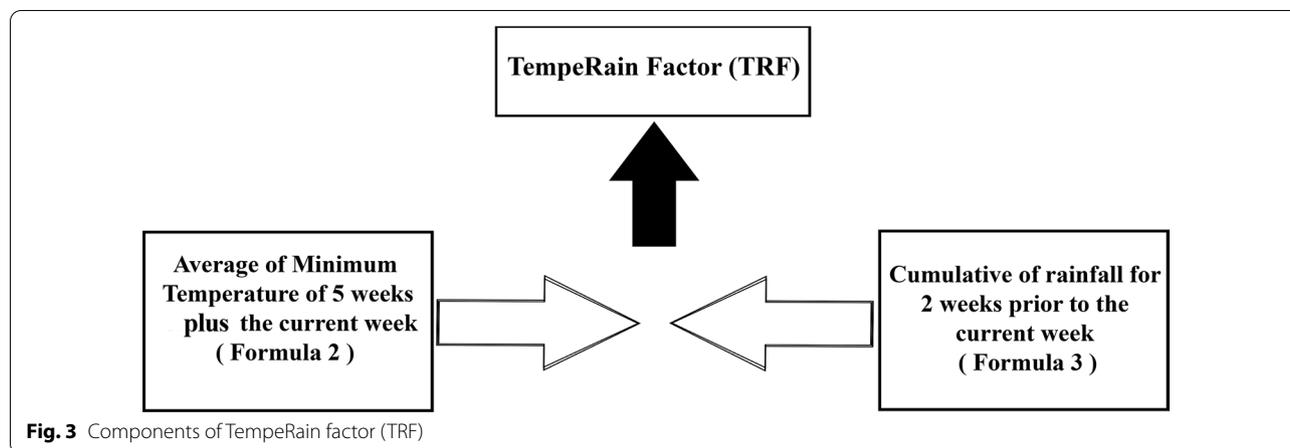
as existence of dengue outbreak and low risk as no outbreak. To predict this level, we tested five machine learning models using input factors with and without TRE. Table 4 provides the detailed input factors and descriptions.

On the basis of the high accuracies obtained [21, 59], we selected Bayes network (BN) models, support vector machine (SVM), RBF tree, decision table and naive

**Table 3** Pearson correlation coefficient (PCC) between climatic factors and 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              |

Bayes to evaluate the factors using WEKA version 3.8.0 [60]. We used the cross-validation (tenfold) technique to evaluate the models.



**Fig. 3** Components of TempeRain factor (TRF)

**Table 4** List of input factors used in prediction model with identified factors (TRF) and without TRF

| Input factors <u>without</u> TRF |                             | Input factors <u>with</u> 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 plus current week (°C)     |
|                                  |                             |                               | Cumulative of rainfall for 2 weeks prior to the current week (mm) |

**Evaluation metrics**

We can evaluate the performance of classifiers on the basis of several accuracy measures and parameters. Moreover, some accuracy and error measures are used to determine the distance between the predicted and the actual known values [61]. In this study, we used the accuracy metric to measure the performance of the classifiers [52, 62, 63]. Equation 4 shows how accuracy is calculated.

$$Accuracy = 100 * \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (4)$$

**Results**

Table 5 presents the results from five machine learning models with and without TRF inputs. Improved results and reduced errors were obtained using the weather data (as external risk factors for a dengue fever outbreak prediction model) by applying machine learning models (as data analysers) and adding newly identified factors (TRF).

**Table 5** Machine learning classifier models using cross-validation (tenfold) with TempeRain factor (TRF)

| Models         | Accuracy (%) |
|----------------|--------------|
| Bayes net      |              |
| With TRF       | <b>92.35</b> |
| Without TRF    | 91.39        |
| SVM            |              |
| With TRF       | 88.04        |
| Without TRF    | 88.00        |
| RBF tree       |              |
| With TRF       | 89.47        |
| Without TRF    | 89.47        |
| Decision table |              |
| With TRF       | 90.41        |
| Without TRF    | 89.95        |
| Naive Bayes    |              |
| With TRF       | 89.4737      |
| Without TRF    | 88.9952      |

Thus, the proposed factors and machine learning model are beneficial for predicting the dengue risk level. The results also showed that models with TRF achieved slightly higher accuracies compared with those without TRF. The highest accuracy was obtained by the BN classifier with TRF (92.35%).

Other studies exhibit different accuracies based on their own private databases, which consist of data collected from patients in hospitals, compared with our research area [20, 23, 25, 64]. Our research used accessible data for climate factors and dengue cases.

Table 6 shows the accuracy of the BN classifier with TRF compared with the other models that used climate factors. All the models compared in this study used binary classification in dengue outbreak prediction. [16, 22, 48, 63] including the proposed model in this study classified dengue outbreak to “outbreak” and “no outbreak”, [21] classified it as “less” and “no case”. The proposed model with TRF achieved the highest accuracy of 92.35% compared with the other models. Besides that, this research used higher number of data compared to [21, 22] which yield an accuracy of more than 90%.

## Discussions

In this study, the use of TRF in BN classifiers managed to outperform the accuracies obtained by other studies compared in this research. It managed to reduce the error of prediction models as well. We believed TRF is the contributing factor that enhances the accuracy. This factor is believed to retain the accuracy of the outbreak prediction model in other countries with similar geographical settings. This will definitely impact many countries such as Philippines, Indonesia, Thailand, Vietnam and Singapore that has similar geographical settings. The risk factors used by all of these countries are portrayed in Table 1. However, countries with different geographical settings, may differs in terms of lagged temperature and rainfall values. This study also supports previous studies that shows temperature and rainfall are most important risk factor that contributes to dengue outbreak.

This outbreak prediction model is expected to particularly help authorized organizations or decision makers in health organizations, governments and other concerned groups to become aware and develop improved prevention programs in the near future. An early warning system based on this model may help in surveillance and controlling the outbreak. This will ensure good reactive management intervention to be in placed effectively and efficiently to curb the epidemics. Thus, this helps communities to be prepared to face the outbreak.

Future work should explore other Malaysian dataset as well as dataset from other countries especially in using TRF by using different machine learning models. Besides that, future research should emphasize the exploration of other risk factors for predicting dengue outbreaks.

## Conclusion

We identified a new significant risk factor, called TRF, which combined the lagged average minimum temperature of 5 weeks together with the current week and lagged cumulative rainfall for 2 weeks. TRF has contributed to dengue outbreak prediction and these lagged weather variables can be useful in determining the dengue outbreak more accurately. The research managed to reveal that the use of accurate and appropriate input factors for outbreak prediction provides enhanced and precise results.

The integration of TRF into the BN model resulted in a significant accuracy of 92.35%. The results showed that using TRF in the BN model outperformed all other outbreak prediction models considered in this study. We do acknowledge although the results showed only almost 1% increase compared to without TRF, this improvement is important as it managed to predict 1 more extra outbreak in every 100 predicted outbreaks. Predicting an increase of 1% outbreak will definitely give significant impact especially for public health surveillance in dealing with infectious diseases like dengue.

**Table 6** Benchmarking and comparing accuracy of the proposed model with previous studies on dengue outbreak prediction model that uses accessible data

| References         | Year | Model  | Accuracy (%)     |
|--------------------|------|--|------------------|
| [63]               | 2018 | Correlation and autoregressive distributed lag model                       | 84.90            |
| [21]               | 2016 | C-SVC kernel and RBF   | 90.50            |
| [22]               | 2013 | Poisson multivariate regression models                                     | 90.00            |
| [16]               | 2010 | Artificial neural networks   | 82.39            |
| [48]               | 2008 | Automatic prediction system by using entropy and artificial neural network | 85.92            |
| Our proposed model |      | Bayes network model using TRF  | Accuracy = 92.35 |

Although many risk factors for dengue outbreak are available, we only focused on the detailed analysis of temperature and rain risk factors for dengue outbreaks, which have been emphasised as the most important factors due to the analysis of importance and access limitation. Future researchers should also test and explore the TRF factors in other datasets from different countries, region or different time period. This research is believed to be an eye opener for future researchers in exploring lagged variables in their outbreak prediction, which include but not limited to dengue. Besides that, the use of deep learning in dengue outbreak prediction should also be ventured.

#### Abbreviations

ANN: Artificial neural networks; BN: Bayes network; DLNM: Distributed lag non-linear model; GA: Genetic algorithm; GEE: Generalized estimating equation; GIS: Geographic information system; GLM: Generalised linear model; MCMC: Markov chain Monte Carlo; MMD: Malaysian Meteorological Department; MOH: Ministry of Health Malaysia; NBR: Negative binomial regression; PCC: Pearson correlation coefficient; PPMCC: Pearson product-moment correlation coefficient; RBF: Radial basis function; SRCC: Spearman's rank correlation coefficient; SVM: Support vector machine; TRF: TempeRain factor; WHO: World Health Organization.

#### Acknowledgements

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

#### Authors' contributions

FYN contributed on the related works, method, experiments and analysis of the studies. KDV contributed in supervision and discussions. Both authors read and approved the final manuscript.

#### Funding

Research University Grant-Faculty Program (GPF011D-2019) has funded the writing of this manuscript. The funding body played no role in the design of the study and collection, analysis, and interpretation of data.

#### Availability of data and materials

Raw Dengue confirmed cases available in portal of Ministry of Health (Malaysia): [http://www.moh.gov.my/index.php/database\\_stores/store\\_view/1](http://www.moh.gov.my/index.php/database_stores/store_view/1)  
For the meteorological data:  
1. <http://www.weather.com>  
2. Available upon reasonable request from the authors (processed data).

#### Declarations

##### Ethics approval and consent to participate

No ethics approval needed to access the data used in this study.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

Received: 25 September 2019 Accepted: 11 April 2021

Published online: 30 April 2021

#### References

- Holmes EC, Tio PH, Perera D, Muhi J, Cardosa J. Importation and co-circulation of multiple serotypes of dengue virus in Sarawak. *Malay Virus Res.* 2009;143(1):1–5. <https://doi.org/10.1016/j.virusres.2009.02.020>.
- Wongkoon S, Jaroensutasinee M, Jaroensutasinee K. Development of temporal modeling for prediction of dengue infection in Northeastern Thailand. *Asian Pac J Trop Med.* 2012;5(3):249–53.
- Chen SC, Hsieh MH. Modeling the transmission dynamics of dengue fever: Implications of temperature effects. *Sci Total Environ.* 2012;431:385–91. <https://doi.org/10.1016/j.scitotenv.2012.05.012>.
- Chinikar S, Ghiasi SM, Shah-Hosseini N, Mostafavi E, Moradi M, Khakifrouz S, Rasi Varai FS, Rafigh M, Jalali T, Goya MM, Shirzadi MR, Zainali M, Fooks AR. Preliminary study of dengue virus infection in Iran. *Travel Med Infect Dis.* 2013;5(3):166–9. <https://doi.org/10.1016/j.tmaid.2012.10.001>.
- Juanarita J, Azmi MNR, Azhary Y, Liza-Sharmini AT. Dengue related maculopathy and foveolitis. *Asian Pac J Trop Biomed.* 2012;2(9):755–6. [https://doi.org/10.1016/S2221-1691\(12\)60223-8](https://doi.org/10.1016/S2221-1691(12)60223-8).
- WHO/TDR. Dengue: guidelines for diagnosis, treatment, prevention and control – New edition. Geneva: World Health Organization; 2009.
- Abeyrathna MPAR, Abeygunawardane DA, Wijesundara RAAV, Mudalige VB, Danaja M, Kaushalya M, Sriganesh L, Madushi B, Shehan P. Dengue propagation prediction using human mobility. *Moratuwa engineering research conference (MERCon).* 2016. p. 156–61.
- World Health Organization (WHO). Weekly epidemiological record. Factsheet117. 2016;30(91):349–64.
- Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature.* 2013;496:504–5077. <https://doi.org/10.1038/nature12060>.
- Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl Trop Dis.* 2012;6(8):e1760. <https://doi.org/10.1371/journal.pntd.0001760>.
- Messina JP, Brady OJ, Golding N, et al. The current and future global distribution and population at risk of dengue. *Nat Microbiol.* 2019;4:1508–15. <https://doi.org/10.1038/s41564-019-0476-8>.
- Ministry of Health Malaysia (MOH). Dengue fever and chikungunya situation. Retrieved from [http://www.moh.gov.my/index.php/database\\_stores/store\\_view/17](http://www.moh.gov.my/index.php/database_stores/store_view/17). Accessed March 2018.
- World Health Organization (WHO). Distribution of dengue, worldwide. Average number of suspected or confirmed dengue cases reported to WHO, 2010–2016. 2018. Retrieved from [www.who.int/denguecontrol/epidemiology/en](http://www.who.int/denguecontrol/epidemiology/en). Accessed March 2019.
- Husin NA, Mustapha N, Sulaiman MN, Yaakob R. A hybrid model using genetic algorithm and neural network for predicting dengue outbreak. 4th conference on. 2012. <https://doi.org/10.1109/DMO.2012.6329793>.
- Andrick B, Clark B, Nygaard K, Logar A, Penaloza M. Infectious disease and climate change: detecting contributing factors and predicting future outbreaks. *Geoscience and Remote Sensing, 1997. IGARSS'97.* 1997. <https://doi.org/10.1109/IGARSS.1997.609159>.
- Aburas HM, Cetiner BG, Sari M. Dengue confirmed-cases prediction: a neural network model. *Expert Syst Appl.* 2010;37(6):4256–60. <https://doi.org/10.1016/j.eswa.2009.11.077>.
- Mathulamuthu SM, Asirvadam VS, Dass SC, Gill BS, Loshini T. Predicting dengue incidences using cluster based regression on climate data. *Control system, computing and engineering (ICCSCE), 2016 6th IEEE international.* 2016. p. 245–250. <https://doi.org/10.1109/ICCSCE.2016.7893579>.
- Soemsap T, Wongthanavasu S, Satimai W. Forecasting number of dengue patients using cellular automata model. *Proc Int Electric Eng Cong.* 2014. <https://doi.org/10.1109/IEECON.2014.6925876>.
- Hii YL, Zaki RA, Aghamohammadi N, Rocklöv J. Research on Climate and Dengue in Malaysia: a systematic review. *Curr Environ Health Rep.* 2016;3(1):81–90. <https://doi.org/10.1007/s40572-016-0078-z>.
- Tazkia RAK, Narita V, Nugroho AS. Dengue outbreak prediction for GIS based early warning system. *International conference on science in information technology (ICSITech).* 2016. <https://doi.org/10.1109/ICSITech.2015.7407789>.
- Rahmawati D, Huang YP. Using C-support vector classification to forecast dengue fever epidemics in Taiwan. *International conference on system science and engineering (ICSSE) National Chi Nan University, Taiwan; July 7–9.* 2016. 978-1-4673-8966-2/16.

22. Hii YL. Climate and dengue fever: early warning based on temperature and rainfall. Umeå University Medical Dissertations. New Series No 1554. 2013. ISSN 0346-6612, ISBN 978-91-7459-589-5.
23. Thitiprayoonwongse D, Suriyaphol P, Soonthornphisaj N. Data mining of dengue infection using decision tree. *Latest Adv Inf Sci Appl Entropy*. 2012;2(2):154–9. <https://doi.org/10.1109/CSITech.2015.7407789>.
24. Tanner L, Schreiber M, Low JGH, Ong A, Tolfvenstam T, et al. Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. *PLoS Negl Trop Dis*. 2008;2(3):e196. <https://doi.org/10.1371/journal.pntd.0000196>.
25. Ibrahim F, Faisal T, Mohamad Salim MI, Taib MN. Non-invasive diagnosis of risk in dengue patients using bioelectrical impedance analysis and artificial neural network. *Med Biol Eng Comput*. 2010;48(11):1141–8. <https://doi.org/10.1007/s11517-010-0669-z>.
26. Korstanje M, George B. Media constructions of fear in the outbreak of an epidemic disease: the case of dengue fever in Argentina. *Int J Emerg Serv*. 2016;5(1):95–104. <https://doi.org/10.1108/IJES-01-2016-0001>.
27. Pinto E, Coelho M, Oliver L, Massad E. The influence of climate variables on dengue in Singapore. *Int J Environ Health Res*. 2011;21(6):415–26. <https://doi.org/10.1080/09603123.2011.572279>.
28. Mathur N, Asirvadam VS, Sarat C. Generating vulnerability maps of dengue incidences for petaling district in Malaysia. 12th international colloquium on signal processing & its applications (CSPA2016). 2016. <https://doi.org/10.1109/CSPA.2016.7515836>.
29. Loshini T, Asirvadam VS, Dass SC, Gill BS. Predicting localized dengue incidences using ensemble system identification. In: International conference on computer, control, informatics and its applications (IC3INA). 2015. pp. 6–11. <https://doi.org/10.1109/IC3INA.2015.7377737>.
30. Burattini MN, Chen M, Chow A, Coutinho FAB, Goh KT, Lopez LF, Ma S, Massad E. Modelling the control strategies against dengue in Singapore. *Epidemiol Infect*. 2008;136(3):309–19. <https://doi.org/10.1017/S0950268807008667>.
31. Mochammad CR, Achmad B, Tri H. Comparison of Montecarlo linear and dynamic polynomial regression in predicting dengue fever case. *Knowl Creat Intell Comput*. 2016. <https://doi.org/10.1109/KCIC.2016.7883649>.
32. Iguchi JA, Seposo XT, Honda Y. Meteorological factors affecting dengue incidence in Davao, Philippines. *BMC Public Health*. 2018;18:629. <https://doi.org/10.1186/s12889-018-5532-4>.
33. Paul KK, Dhar-Chowdhury P, Haque CE, Al-Amin HM, Goswami DR, Kafi MAH, et al. Risk factors for the presence of dengue vector mosquitoes, and determinants of their prevalence and larval site selection in Dhaka, Bangladesh. *PLoS ONE*. 2018;13(6):e0199457. <https://doi.org/10.1371/journal.pone.0199457>.
34. Lee HS, Nguyen-Viet H, Nam VS, Lee M, Won S, Duc PP, Grace D. Seasonal patterns of dengue fever and associated climate factors in 4 provinces in Vietnam from 1994 to 2013. *BMC Infect Dis*. 2017;17:218. <https://doi.org/10.1186/s12879-017-2326-8>.
35. Datoc HI, Caparas R, Caro J. Forecasting and data visualization of dengue spread in the Philippine Visayas Island group. In: 7th international conference on information, intelligence, systems & applications (IISA). 2016. <https://doi.org/10.1109/IISA.2016.7785420>.
36. Xiang J, Hansen A, Liu Q, Liu X, Tong MX, Sun Y, Weinstein P. Association between dengue fever incidence and meteorological factors in Guangzhou, China, 2005–2014. *Environ Res*. 2016;153:17–26. <https://doi.org/10.1016/j.envres.2016.11.009>.
37. Hai-Yan Xu, Fu X, Lee LKH, Ma S, Goh KT, Wong J, Lim CL. Statistical modeling reveals the effect of absolute humidity on dengue in Singapore. *PLoS Negl Trop Dis*. 2014;8(5):e2805. <https://doi.org/10.1371/journal.pntd.0002805>.
38. Lung CC, Hwa LY. Impact of meteorological factors on the spatiotemporal patterns of dengue fever incidence. *Environ Int*. 2014;73:46–56.
39. Bouzid M, Colón-González FJ, Lung T, Lake IR, Hunter PR. Climate change and the emergence of vector-borne diseases in Europe: case study of dengue fever. *BMC Public Health*. 2014;2014(14):781. <https://doi.org/10.1186/1471-2458-14-781>.
40. Colón-González FJ, Fezzi C, Lake IR, Hunter PR. The effect of weather and climate change on dengue. *PLoS Negl Trop Dis*. 2013;7(11):e2503. <https://doi.org/10.1371/journal.pntd.0002503>.
41. Cheong YL, Burkart K, Leitão PJ, Lakes T. Assessing weather effects on dengue disease in Malaysia. *Int J Environ Res Public Health*. 2013;10(12):6319–34. <https://doi.org/10.3390/ijerph10126319>.
42. Dom NC, Hassan AA, Latif ZA, Ismail R. Generating temporal model using climate variables for the prediction of dengue cases in Subang Jaya, Malaysia. *Asian Pac J Trop Dis*. 2013;3(5):352–61. [https://doi.org/10.1016/S2222-1808\(13\)60084-5](https://doi.org/10.1016/S2222-1808(13)60084-5).
43. Hii YL, Zhu H, Ng N, Ng LC, Rocklöv J. Forecast of dengue incidence using temperature and rainfall. *PLoS Negl Trop Dis*. 2012;6(11):e1908. <https://doi.org/10.1371/journal.pntd.0001908>.
44. Wang Z, Chan HM, Hibberd ML, Lee GKK. Delayed effects of climate variables on incidence of dengue in Singapore during 2000–2010. *APCBEE Proc*. 2012;1:22–6. <https://doi.org/10.1016/j.apcbee.2012.03.005>.
45. Rachel L, Bailey TC, Stephenson DB, Graham RJ, Coelho CAS, Carvalho MS, Barcellos C. Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Comput Geosci*. 2011;37(3):371–81. <https://doi.org/10.1016/j.cageo.2010.01.008>.
46. Halmar H. Assessing quality and value of predictive models for dengue hemorrhagic fever epidemics. New York: Nova Publisher; 2010.
47. Cetiner BG, Sari M, Aburas HM. Recognition of dengue disease patterns using artificial neural networks. In: 5th international advanced technologies symposium (IATS'09). 2009. p. 359–362.
48. Rachata N, Charoenkwan P, Yooyativong T, Chamnongthai K, Lursinsap C, Higuchi K. Automatic prediction system of dengue haemorrhagic-fever outbreak risk by using entropy and artificial neural network. *Communications and information technologies, 2008 (ISCIT)*. 2008. p. 210–214. <https://doi.org/10.1109/ISCIT.2008.4700184>.
49. Promprou S, Jaroensutasinee M, Jaroensutasinee K. Climatic factors affecting dengue haemorrhagic fever incidence in Southern Thailand. 2005.
50. Malaysian Meteorological Department (MMD). Climatic and weather information, request from <http://www.met.gov.my>. Accessed Feb 2019.
51. Saranya C, Manikandan G. A study on normalization techniques for privacy preserving data mining. *IJET*. 2013;5:2701–4.
52. Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans Asabe*. 2007;50(3):885–900. <https://doi.org/10.13031/2013.23153>.
53. Christophers SR. *Aedes aegypti* (L.) the yellow fever mosquito. Its life history. In: *Bionomics and structure*. Cambridge Univ. Press, Cambridge. 1960;133(3463):1473–1474. <https://doi.org/10.1126/science.133.3463.1473-a>.
54. Yang HM, Macoris MLG, Galvani KC, Andrighetti MTM, Wanderley DMV. Assessing the effects of temperature on the population of *Aedes aegypti*, the vector of dengue. *Epidemiol Infect*. 2009;137:1188–202. <https://doi.org/10.1017/S0950268809002040>.
55. Ahmad R, Wong YC, Zamre I, Lee HL, Zurainee MN. 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*. 2009;40(5):942–650.
56. Ibrahim A, Zin NAM, Ashaari NS. Simulation model for predicting dengue fever outbreak. *World Acad Sci Eng Technol Int J Comput Inf Sci Eng* 2011;5(11).
57. Watts DM, Burke DS, Harrison BA, Whitmire RE, Nisalak A. Effect of temperature on the vector efficiency of *Aedes aegypti* for dengue 2 virus. *Am J Trop Med Hyg*. 1987;36(1):143–52. <https://doi.org/10.4269/ajtmh.1987.36.143>.
58. Chan M, Johansson MA. The incubation periods of dengue viruses. *PLoS ONE*. 2012;7(11):e50972. <https://doi.org/10.1371/journal.pone.0050972>.
59. Fathima S, Hundewale N. Comparison of classification techniques-SVM and Naives Bayes to predict the arboviral disease-dengue. In: International conference on bioinformatics and biomedicine workshops. 2011. <https://doi.org/10.1109/BIBMW.2011.6112426>.
60. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explor*. 2009;11(1):10–8.
61. Nasa C, Suman S. Evaluation of different classification techniques for WEB data. *Int J Comput Appl*. 2012;52(9):34–40.
62. Boyle DP, Gupta HV, Sorooshian S. Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. *Water Resour Res*. 2000;36(12):3663–74. <https://doi.org/10.1029/2000WR900207>.
63. Ahmad R, Suzilah I, Wan Najdah WMA, Topek O, Mustafakamal I, Lee HL. Factors determining dengue outbreak in Malaysia. *PLoS ONE*. 2018;13(2):e0193326. <https://doi.org/10.1371/journal.pone.0193326>.

64. Saha S. Combined committee machine for classifying dengue fever. In: Microelectronics, computing and communications (MicroCom), 2016 international conference on. 2016. p. 1–6. <https://doi.org/10.1109/MicroCom.2016.7522585>.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

