



## The practicality of Malaysia dengue outbreak forecasting model as an early warning system



Suzilah Ismail <sup>a,\*</sup>, Robert Fildes <sup>b</sup>, Rohani Ahmad <sup>c</sup>, Wan Najdah Wan Mohamad Ali <sup>c</sup>, Topek Omar <sup>d</sup>

<sup>a</sup> School of Quantitative Sciences, Universiti Utara Malaysia UUM Sintok, Kedah, Malaysia

<sup>b</sup> Centre for Marketing Analytics and Forecasting, Lancaster University, United Kingdom

<sup>c</sup> Medical Entomology Unit, Institute for Medical Research, National Institute of Health, Ministry of Health Malaysia, Malaysia

<sup>d</sup> Vector Borne Disease Sector, Disease Control Division, Ministry of Health Malaysia, Malaysia

### ARTICLE INFO

#### Article history:

Received 27 April 2022

Received in revised form 7 July 2022

Accepted 30 July 2022

Available online 8 August 2022

Handling Editor: Dr Lou Yijun

#### Keywords:

dengue

rainfall

Machine learning

Early warning system

IoT

### ABSTRACT

Dengue is a harmful tropical disease that causes death to many people. Currently, the dengue vaccine development is still at an early stage, and only intervention methods exist after dengue cases increase. Thus, previously, two scientific experimental field studies were conducted in producing a dengue outbreak forecasting model as an early warning system. Successfully, an Autoregressive Distributed Lag (ADL) Model was developed using three factors: the epidemiological, entomological, and environmental with an accuracy of 85%; but a higher percentage is required in minimizing the error for the model to be useful. Hence, this study aimed to develop a practical and cost-effective dengue outbreak forecasting model with at least 90% accuracy to be embedded in an early warning computer system using the Internet of Things (IoT) approach. Eighty-one weeks of time series data of the three factors were used in six forecasting models, which were Autoregressive Distributed Lag (ADL), Hierarchical Forecasting (Bottom-up and Optimal combination) and three Machine Learning methods: (Artificial Neural Network (ANN), Support Vector Machine (SVM) and Random Forest). Five error measures were used to evaluate the consistency performance of the models in order to ensure model performance. The findings indicated Random Forest outperformed the other models with an accuracy of 95% when including all three factors. But practically, collecting mosquito related data (the entomological factor) was very costly and time consuming. Thus, it was removed from the model, and the accuracy dropped to 92% but still high enough to be of practical use, i.e., beyond 90%. However, the practical ground operationalization of the early warning system also requires several rain gauges to be located at the dengue hot spots due to localized rainfall. Hence, further analysis was conducted in determining the location of the rain gauges. This has led to the recommendation that the rain gauges should be located about 3–4 km apart at the dengue hot spots to ensure the accuracy of the rainfall data to be included in the dengue outbreak forecasting model so that it can be embedded in the early warning system. Therefore, this early warning system can save lives, and prevention is better than cure.

© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\* Corresponding author.

E-mail address: [halizus@uum.edu.my](mailto:halizus@uum.edu.my) (S. Ismail).

Peer review under responsibility of KeAi Communications Co., Ltd.

## 1. Introduction

Dengue is a harmful disease caused by a virus commonly found in *Aedes aegypti* and *Aedes albopictus* mosquito (Johari et al., 2019; Nor Aliza et al., 2019; Rohani et al., 2014). Every year, an estimated 100–400 million people are affected by dengue (WHO, 2021), where about 5% of cases die (World Mosquito Program, 2021). At present, a dengue vaccine is still at an early stage of development (Lim & Poh, 2018; Sheng-Qun et al., 2020). Thus, only an intervention method exists to combat dengue. In many countries, including Malaysia, the current intervention around a dengue outbreak is usually conducted after dengue cases have occurred, such as fogging the dengue cases area with insecticide (MaHTAS, 2019; Song-Quan, 2016), but this conventional approach causes harm to the community.

Consequently, many studies have been conducted to predict dengue outbreaks using mathematical models (either statistical or artificial intelligence or both) in attempts to create an early warning system to prevent dengue. Sanchez-Gendriz et al. (2022) developed a data-driven computational intelligence approach to forecasting the dengue outbreak at Natal, Brazil, using Regression and Neural Network models. Liu, Yin, et al. (2021) implemented Machine Learning Models (Artificial Neural Network and Support Vector Machine) in Guangzhou, China, by integrating them with environmental features. Patil and Pandya (2021) employed Regression, Machine Learning (Random Forest, Decision Trees, Support Vector Machine), and Time Series Models (Moving average, Exponential Smoothing, and ARIMA) in Maharashtra State, India, in forecasting dengue hotspots using meteorological parameters. Colon-Gonzalez et al. (2021) applied Bayesian spatiotemporal models in predicting dengue cases in Vietnam. Johansson et al. (2019) used probabilistic forecasting for dengue epidemics in Peru and Puerto Rico. Ong et al. (2018) used Random Forest to predict risk dengue transmission based on dengue, population, entomological and environmental data in Singapore. The outcome of these studies highlighted the significance of weather data as the key driver and machine learning methods as the best performing dengue outbreak forecasting model.

In Malaysia, the search for a useful dengue outbreak forecasting model started 15 years ago by conducting scientific experimental field studies to establish appropriate impactful factors because many past studies had failed to establish sufficiently accurate models to be of practical value (Ang & Li, 2002; Cheah et al., 2006; Cheng, 2006; Fatimah et al., 2005; Luz et al., 2003; Seng et al., 2005, pp. 109–123; Usman, 2003). One reason was relating a seasonal pattern to the epidemiological factor (dengue cases) where the impact of climate change has amplified the inconsistency of the rainfall season in Malaysia (Anyamba et al., 2006; Bartley et al., 2002; Cheah et al., 2006; Hales et al., 2002; Hartley et al., 2002; Sulaiman et al., 1996; Woodruff et al., 2006). Another reason, although there were studies using two factors: epidemiological (dengue cases) and environmental (weather data) but relying on weather data from Meteorological Department (secondary data) jeopardized the accuracy of forecasting model (Barbazan et al., 2002; Cheah et al., 2006; Kumarasamy, 2006; Ram et al., 1998; Rosa-Freitas et al., 2006) because of the localized rainfall in Malaysia.

Thus, previously, two scientific studies were implemented using an experimental field design to establish factors determining dengue outbreaks in Malaysia, which were epidemiological (dengue cases data), entomological (mosquito-related data), and environmental (weather data) factors. These studies aimed to produce a dengue outbreak forecasting model suitable to be incorporated into an early warning system. Experimental field design was used to collect the primary data of the entomological and environmental factors to guarantee the accuracy of the obtained dengue outbreak forecasting model. Past studies have shown that using secondary weather data in the dengue prediction model has lowered accuracy, such as in Salim et al. (2021), who only obtained 70% and Jain et al. (2019), attaining 73%. The first experimental study was implemented from 2007 until 2009 in four high dengue cases housing areas in Malaysia (Rohani et al., 2011). But due to the small areas selected, only two factors (entomological and environmental) were significant. However, the study contributed important findings regarding the relationship between entomological (mosquito related data) and environmental (rainfall, temperature, and humidity); where it revealed that the previous week's rainfall (which was collected primarily by placing a mobile weather station in the housing areas) influenced the dengue mosquito population's increment. But the rainfall data (secondary data) obtained from the Meteorological Department based on their weather stations showed insignificant results. This is because Malaysia's rainfall pattern is very localized (Muhammad et al., 2020; Singh et al., 2022), meaning there are rains in certain areas but no rain just a few blocks away. The localized rain can jeopardize rainfall data quality (obtained from Meteorological Department). But rainfall is an important indicator in dengue outbreak forecasting model (Hii et al., 2012; Liu, Yin, et al., 2021; Ong et al., 2018; Patil & Pandya, 2021; Singh et al., 2022) because rainfall influence the *Aedes* mosquito population (the vector of dengue virus) to breed. Therefore, a practical solution is by placing several rain gauges (mobile weather stations) in the dengue hot spot areas, thereby enhancing the quality of the rainfall data, and potentially increasing the accuracy of the dengue outbreak forecasting model.

The first study's findings and limitations were considered when implementing the second experimental field study in two dengue-prone areas in Malaysia (Rohani et al., 2018). Two larger areas were selected based on five consecutive years of high dengue cases: ovitraps were used to collect the *Aedes* larvae and several mobile weather stations comprising rain gauges, temperature and humidity data loggers were allocated at both areas. Similar findings were obtained regarding the relationship between entomological and environmental as in Rohani et al. (2011). The findings also indicated the three factors were significant, and in-depth relationships among them were established. An Autoregressive Distributed Lag (ADL) Model was developed to forecast dengue outbreaks with an accuracy of 85%.

Hence, this study aims to develop a dengue outbreak forecasting model with at least 90% accuracy to be embedded in an early warning computer system using the Internet of Things (IoT) approach that is practical and cost-effective for ground

operationalization of the system. High accuracy of 90% is essential to minimize the costs and ensure the benefits when implemented in the dengue hot spot area.

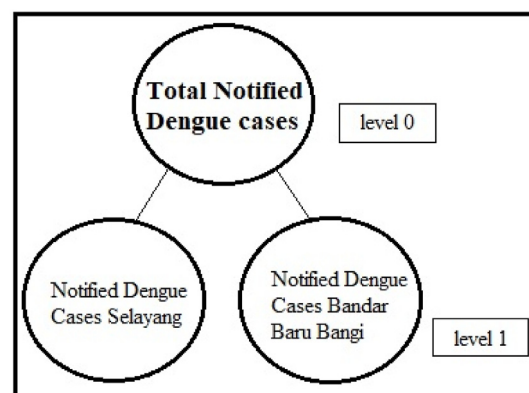
The last few years have seen considerable interest and recent success in developing new forecasting methods. Two areas that have seen greatest success is by applying machine learning (ML) methods to time series forecasting problems such as the ones outlined here (see [Makridakis et al., 2018](#)), and second, the use of combinations of forecasts through hierarchical models. We, therefore, consider three standard ML methods and analyze their performance compared to a state-of-the-art statistical approach, whilst also examining the value that taking a hierarchical approach brings. Thus, the contribution of this study is twofold: to develop a forecasting method sufficiently accurate to be of value to those working on Dengue fever, and second, to examine the benefits of introducing more advanced forecasting methods into the field.

## 2. Methodology

We used [Rohani et al. \(2018\)](#) time series data which was collected weekly for eighty-one (81) weeks involving three factors (epidemiological, entomological, and environmental) for two large areas in Malaysia (Selayang and Bandar Baru Bangi). The two areas were selected based on five consecutive years of high dengue cases. Epidemiological data were notified dengue cases, onset cases and the number of interventions was obtained from the e-Dengue system developed by the Malaysia Ministry of Health. Notified dengue cases were defined as clinical description dengue cases and notified to the nearest health office, while the onset cases were denoted as the first day of having high fever related to dengue symptoms. The number of interventions was based on interventions conducted in the dengue cases areas. Entomological data were the number of *Aedes* larvae collected weekly using 50 ovitraps in Selayang and 55 ovitraps in Bandar Baru Bangi; and screened using reverse transcriptase-polymerase chain reaction (RT-PCR) method to detect the dengue virus (positive PCR). This process was time consuming and very costly to conduct because every week, all ovitraps (105) had to be collected and replaced with fresh ovitraps. Then the ovitraps were brought to the laboratory and the eggs/larvae were allowed to further develop up to 5 days and 10 days respectively for the *Aedes* identification which requires a total of 15 days for each batch each week. Next, the larvae were screened using RT-PCR to detect the type of dengue virus with each RT-PCR costing around MYR100, a total cost of MYR200,000 (USD50,000) for 81 weeks. But this huge cost was worth it because of the success in establishing deep understanding regarding the relationships among the three factors as the parameters in the dengue outbreak forecasting model. The environmental data were collected using ten and eleven mobile weather stations allocated at both areas, respectively, which consisted of rain gauges, temperature, and humidity data loggers. The Air Pollution Index (API) was obtained from Malaysia Environment Department. Refer to [Rohani et al. \(2018\)](#) for the details and description of the data collections.

Based on [Fig. 1](#), all the data were aggregated for level 0 described as the pooled data because the aim was to develop one dengue outbreak forecasting model for Malaysia. Total was calculated to pool the epidemiological data (notified dengue cases, onset cases and number of interventions) in both areas as well as the entomological data (*Aedes* larvae and number of positive PCR) and rainfall data. Next, the data were partitioned for model estimation/training (week 1 until week 70) and model evaluation/testing (week 71 until week 81) in measuring the model's performance.

Six forecasting methods were used, which were Autoregressive Distributed Lag (ADL) model, Hierarchical Forecasting (Bottom-up and Optimal combination), and Machine Learning (ML) methods (Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest). The ADL and ML forecasting methods were used because based on past studies ([Chiung et al., 2018](#); [Guo et al., 2017](#); [Liu, Yin, et al., 2021](#); [Ong et al., 2018](#); [Patil & Pandya, 2021](#); [Sanchez-Gendriz et al., 2022](#)) showed their relevance in predicting dengue outbreaks. They had also proved their value in various forecasting evaluation studies (see e.g., [Makridakis et al., 2018](#)). Hierarchical Forecasting approaches (Bottom up and Optimal combination) have also been adopted because while the aim is to produce only one generic dengue outbreak forecasting model, by pooling the two areas data improved accuracy may be expected, as several studies have shown that pooling the data can increase the accuracy



**Fig. 1.** Hierarchical structure.

of the resulting model (Matthias & Feng, 2022; Petropoulos et al., 2022; Siriayasatien et al., 2018; Spiliotis et al., 2019; Zhao et al., 2020).

2.1. a. Autoregressive Distributed Lag (ADL)

The following equation representing the ADL model was used in this study,

$$Y_t = \alpha_0 + \sum_{k=1}^{10} \sum_{j=0}^4 \phi_{kj} X_{k(t-j)} + \epsilon_t \tag{1}$$

where  $j$  is the lag length,  $t = 1, 2, \dots, T$  (time periods) and  $Y_t$  is the target variable which is the notified dengue cases.  $\{X_{kt}\}$  represent the ten predictors, which are epidemiological, entomological, and environmental variables.  $\epsilon_t$  are independent identically distributed random errors with mean zero and variance  $\sigma_{\epsilon_t}^2$ ,  $\alpha$  and  $\phi$  are unknown parameters to be estimated using Ordinary Least Squares (Mohd Alias, 2013).

Based on the Autoregressive Distributed Lags (ADL) models obtained in Rohani et al. (2018), the following variables contributed to the 85% accuracy, supported by the theoretical explanation of the dengue experts; these have been based on the *Aedes* mosquito life cycle related to weather, virus incubation period, dengue symptoms, and intervention taken by the authority when there were reported cases.

Dependent (Target Variable):

Notified Dengue Cases ( $Y_t$ )

Independent Variables (Predictors):

Onset dengue cases (Onset)

Last week Onset dengue cases (Onset1)

Last 3 weeks Larvae (Larvae3)

Last 3 weeks Positive PCR (PCR3)

Last 4 weeks Rainfall (Rainfall4)

Last 3 weeks Minimum Temperature (MinTemp3)

Last 3 weeks Maximum Temperature (MaxTemp3)

Last 3 weeks Maximum Humidity (MaxHumid3)

Last 4 weeks of Air Pollution Index (API4)

Next week Intervention (InterventionLD1)

In this study, the pooled data of these variables were used in the ADL model and also the following forecasting methods in determining the best dengue outbreak forecasting model that can produce at least 90% accuracy.

2.2. b. hierarchical forecasting

There are four hierarchical forecasting methods: Top-down, Bottom-up, Middle out, and Optimal combination (Hyndman & Athanasopoulos, 2021). However, in this study, only Bottom-up and Optimal combinations were used because the aim was to produce one general model for Malaysia by pooling the data as explained and illustrated in Fig. 1 previously. Top-down was excluded because it would generate a separate individual model for the two areas. Middle out was also excluded because it required at least two hierarchical levels, and, in this study, there was only one level. In the Bottom-up, two ADL models were estimated at the bottom level of the hierarchical structure (level 1) as in Fig. 1. Then the forecast values of the two models were added together to produce the level 0 forecast values.

In the optimal combination, forecast values were produced at level 1, and level 0 using three ADL models (i.e., two for the individual areas data and one using the pooled data) and named as the base forecast values. Next, the three sets of base forecast values were used in the following equation to produce coherent forecast (Hyndman & Athanasopoulos, 2021)

$$\tilde{y}_h = \mathbf{S}\mathbf{G}\hat{y}_h \tag{2}$$

$$\mathbf{G} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}$$

Therefore, the optimal reconciled forecasts are given by,

$$\tilde{y}_h = \mathbf{S}(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}\hat{y}_h \tag{3}$$

$\mathbf{W}_h = k_h \text{diag}(\widehat{\mathbf{W}}_1)$  for all  $h$ , where  $k_h > 0$

$$\widehat{\mathbf{W}}_1 = \frac{1}{T} \sum_{t=1}^T \mathbf{e}_t \mathbf{e}_t'$$

where  $\widehat{\mathbf{y}}_h$  is coherent forecasts,  $\widehat{\mathbf{y}}_h$  is base forecasts,  $\mathbf{G}$  is a matrix that maps the base forecasts into the bottom-level, and the summing matrix  $\mathbf{S}$  sums these up using the aggregation structure to produce a set of coherent forecasts,  $\mathbf{W}_h$  is the forecast error variance of the  $h$ -step-ahead base forecasts and  $\mathbf{e}_t$  is vector of residuals of the models that generated the base forecasts stacked in the same order as the data.

2.3. c. machine learning methods

Three methods were used in this study which are Artificial Neural Network (ANN), Support Vector Machine (SVM) and Random Forest. ANN adopts the Multi-Layer Perceptron algorithms together with the back propagation method (Blokdyk, 2020, p. 305). Support Vector Machine (SVM) for regression with a polynomial kernel (Kowalczyk, 2017). Random Forest constructs multiple decision trees with bootstrap aggregation (Geneur & Poggi, 2020). All these ML methods were conducted using WEKA (Brownlee, 2020).

After obtaining 6 estimated models from the forecasting methods, forecast errors were generated based on *out-of-sample* data (week 71–81 as stated in earlier section for model evaluation) and the following error measures (Koutsandreas et al., 2022; Chen et al., 2017) were used to evaluate the performances of the models. In this application, several error measures were required in order to determine the performance consistency of the forecasting models (Davydenko & Fildes, 2013) because there are multiple potential users and applications of the final forecasting model under construction who do not necessarily share the same loss functions.

a. Geometric Mean Relative Absolute Error,  $GMRAE = \left( \prod_{t=1}^n \left| \frac{e_t}{e_t^*} \right| \right)^{\frac{1}{n}}$  Where  $e_t^*$  is the *out-of-sample* forecast error obtained

from Random Walk Model  $Y_t = Y_{t-1} + Z_t$ , and  $Z_t$  is a white noise variable with zero mean and constant variance  $\sigma^2$ .

b. Relative Mean Absolute Error,  $ReMAE = \frac{MAE}{MAE^*}$

Where  $MAE = \frac{1}{n} \sum_{t=1}^n |e_t|$  and  $MAE^*$  is the *out-of-sample* MAE obtained from Random Walk Model  $Y_t = Y_{t-1} + Z_t$ , and  $Z_t$  is a white noise variable with zero mean and constant variance  $\sigma^2$ .

c. Mean Absolute Scaled Error,  $MASE = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{MAE^{**}}$  Where  $MAE^{**}$  is the *in-sample* MAE obtained from Random Walk Model  $Y_t = Y_{t-1} + Z_t$ , and  $Z_t$  is a white noise variable with zero mean and constant variance  $\sigma^2$ .

d. Unscaled Mean Bounded Relative Absolute Error,  $UMBRAE = \frac{MBRAE}{1-MBRAE}$  Where  $MBRAE = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{|e_t| + |e_t^*|}$ ,  $e_t^*$  is the *out-of-sample* forecast error obtained from Random Walk Model  $Y_t = Y_{t-1} + Z_t$ , and  $Z_t$  is a white noise variable with zero mean and constant variance  $\sigma^2$ .

e. Mean Absolute Percentage Error,  $MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|e_t|}{|Y_t|}$

MAPE was used to calculate the percentage accuracy of the model (100 – MAPE)

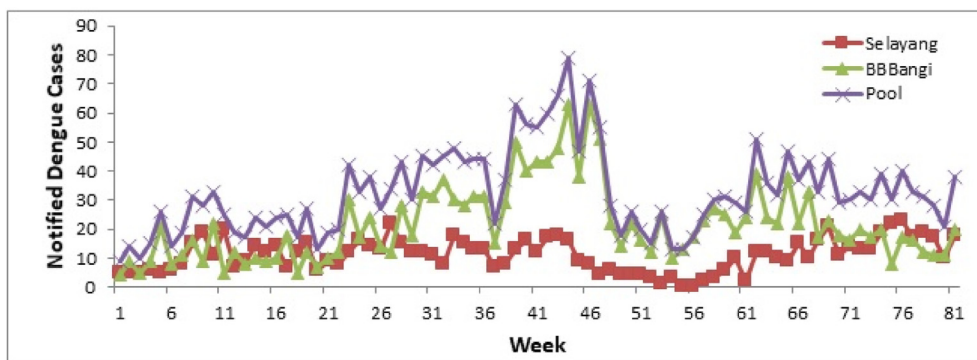


Fig. 2. Notified dengue cases for Selayang, Bandar Baru Bangi and pool data.

These forecast errors were calculated based on 4 weeks ahead (Lead 1, 2, 3 and 4); while median was used to summarize the overall 4 leads for each of the above error measures. Next, the median was ranked from 1 to 6 to determine the best forecasting model.

### 3. Results and discussion

Fig. 2 shows 81 weeks of notified dengue cases of Selayang, Bandar Baru Bangi, and aggregated (pool) cases for both areas. Selayang and Bandar Baru Bangi have quite similar patterns, although the numbers of cases were more in Bandar Baru Bangi as compared to Selayang. Rohani et al. (2018) has established a relationship between notified dengue cases and three factors which were the epidemiological, entomological, and environmental using correlation and an Autoregressive Distributed Lag (ADL) model.

Table 1 displays the correlation of the notified dengue cases (target variable) and predictors of Selayang, Bandar Baru Bangi, and the pooled data. There were significant relationships in the pooled data between notified dengue cases with all the predictors, where the notified cases related to this week and last week onset cases; the last 3 weeks larvae; the last 3 weeks positive PCR; the last 3 weeks minimum and maximum temperature; the last 3 weeks maximum humidity; the last 4 weeks rainfall and the last 4 weeks air pollution index (API), respectively. Notified cases were also related with next week intervention.

Based on the correlations obtained in Table 1, Fig. 3 displays the conceptual relationship between the target variable and predictors according to the interconnection of the 3 factors epidemiological, entomological, and environmental. The relationship was established by dengue experts in connecting the variables with the *Aedes* mosquito life cycle related to weather, virus incubation period, dengue symptoms, and intervention taken by the authority when there were reported cases. Rainfall plays an important role in *Aedes* mosquito life cycle by providing potential breeding sites for the *Aedes* mosquito to lay eggs and develop to the adult stage (Banulata et al., 2021; Malinda et al., 2012; Masnita et al., 2018; Seidahmed & Eltahir, 2016) but long-term air pollution (API) due to haze may interfere with the life cycle of the *Aedes* mosquito thus reduce their population (Ahmad Khairul Hakim et al., 2020; Carneiro et al., 2017; Rohani et al., 2018; Wilder-Smith et al., 2010). Next, the eggs hatch into larvae within 2–4 days, depending on temperature and humidity. Lower minimum temperature (21–22 °C), higher minimum temperature (26 °C and above) and higher maximum temperature (36 °C and above) decreased the number of larvae, but higher maximum humidity (90% and above) increased the number of larvae. However, the temperature and humidity must be supported with the amount of rain to influence the multiplication of larvae (Drakou et al., 2020; Masnita et al., 2018; Tran et al., 2020). According to Tran et al. (2020), the effects of precipitation and relative humidity on Breteau Index (BI) were more obvious when the average temperature exceeded the threshold. The larger number of larvae also contributed to higher chances of obtaining more positive PCR (Fansiri et al., 2021; Peña-García et al., 2016). Then after 5–10 days, adult *Aedes* mosquito (male and female) emerge and mate within 3–5 days. Later, the infected adult female mosquito bites human for blood in producing eggs (Harrison et al., 2021; Clements, 1992). *Aedes* mosquito acquires the virus while feeding on the blood of an infected (viraemic) human. It takes 3–8 days before the mosquito become capable of transmitting the virus (Alto & Bettinardi, 2013; Carrington & Simmons, 2014) either vertically (transovarial) or horizontally. Vertical or transovarial transmission involves the passage of dengue viruses to the offspring through the eggs of infected female mosquitoes (Wan Najdah et al., 2021; Windyaraini et al., 2019; da Costa et al., 2017; Joshi et al., 2002) while horizontal transmission involves the passage of dengue viruses to a vertebrate host via a vector. Typically, 4 days after being bitten by an infected *Aedes* mosquito, a person will develop viremia, a condition in which there is a high level of the dengue virus in the blood. Viremia lasts for approximately 5–12 days depending on the virus incubation period in the human. On the first day of viremia, the person generally shows no symptoms of dengue fever. Five to 7 days after being bitten by the infected mosquito, the person develops symptoms of dengue fever, starting with high fever, defined as onset dengue cases. Typically, 70% of the dengue cases have been notified to the nearest health office, 2–6 days after onset (termed as notified dengue cases) because in the early-stage people usually misdiagnose themselves (MaHTAS, 2015). Once the cases have been notified, the Health Department conducts an intervention at the dengue hot spot (MaHTAS, 2015). For larva control, the strategies conducted would be environment management, source reduction, use of larvicides such as temephos (Abate), house inspection and enforcement of Destruction of Disease-Bearing Insect Act 1975; while for adult control, fogging will be carried out based on

**Table 1**  
Correlations between notified dengue cases (target variable) and predictors.

| Area     | Notified Dengue Cases (Target Variable) |                    |                     |                    |                    |                    |                     |                     |                    |                    |
|----------|-----------------------------------------|--------------------|---------------------|--------------------|--------------------|--------------------|---------------------|---------------------|--------------------|--------------------|
|          | Predictors                              |                    |                     |                    |                    |                    |                     |                     |                    |                    |
|          | Onset                                   | Onset1             | API4                | Larvae3            | PCR3               | Rainfall4          | Min Temp3           | Max Temp3           | Max Humid3         | InterventionLD1    |
| Selayang | 0.782 <sup>a</sup>                      | 0.814 <sup>a</sup> | -0.637 <sup>a</sup> | 0.847 <sup>a</sup> | 0.537 <sup>a</sup> | 0.678 <sup>a</sup> | -0.453 <sup>a</sup> | -0.452 <sup>a</sup> | 0.674 <sup>a</sup> | 0.533 <sup>a</sup> |
| BBBangi  | 0.819 <sup>a</sup>                      | 0.811 <sup>a</sup> | -0.393 <sup>a</sup> | 0.747 <sup>a</sup> | 0.407 <sup>a</sup> | 0.678 <sup>a</sup> | -0.379 <sup>a</sup> | -0.474 <sup>a</sup> | 0.656 <sup>a</sup> | 0.304 <sup>a</sup> |
| Pool     | 0.837 <sup>a</sup>                      | 0.871 <sup>a</sup> | -0.415 <sup>a</sup> | 0.797 <sup>a</sup> | 0.465 <sup>a</sup> | 0.665 <sup>a</sup> | -0.485 <sup>a</sup> | -0.518 <sup>a</sup> | 0.649 <sup>a</sup> | 0.260*             |

<sup>a</sup> Significant at 1% \* Significant at 5%.

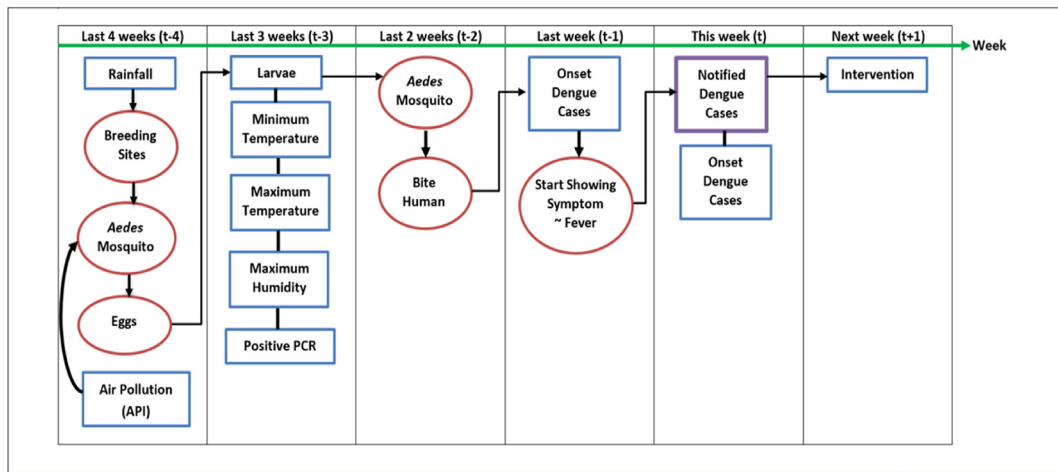


Fig. 3. Conceptual relationship (target variable & predictors): Epidemiological, entomological and environmental factors based on weeks.

the viral cases reported (Kumarasamy, 2006; Tham, 2000, pp. 15–23). All the fogging activities are employed by the trained practitioners of Ministry of Health and Ministry of Housing and Local Government. Since the intervention was implemented after 4 weeks that larvae have existed, there is ample time for dengue transmission. We then can expect a subsequent initial increase in dengue cases being notified from the locality 3–4 weeks after the existence of larvae (Sanchez-Gendriz et al., 2022; Rohani et al., 2018). The reason is these intervention methods cannot prevent dengue outbreaks by themselves as they are not sustainable. An important reason for this is the low vector threshold for dengue transmission (Shamsul et al., 2012). As low as 2–3 adult female mosquitoes emerging every day in a locality of 100 people is sufficient to start an outbreak.

Table 2 presents the estimated ADL model for Selayang, Bandar Baru Bangi and the Pooled data. Five variables were significant in the pooled model, which were Onset, Onset1, Larvae3, Rainfall4 and InterventionLD1. Although there were other variables that were not significant in the model but since these variables were important in explaining the notified dengue cases and the relationship have been verified by experts and established in previous study Rohani et al. (2018). Thus, the same set of variables were used in the other forecasting methods considered.

Table 3 show the error measures results based on RelMAE, MASE, GMRAE and UMBRAE for 4 weeks ahead forecast (lead 1, 2, 3 and 4). All error measures indicated Random Forest as the best forecasting model, followed by Support Vector Machine (SVM) and Artificial Neural Network (ANN). These demonstrated machine learning (ML) methods performed better than the Autoregressive Distributed Lag (ADL) and Hierarchical Forecasting when applied to an ADL base forecaster. Since the ADL model is the basis of Hierarchical Forecasting (Bottom-up and Optimal combination), three error measures (RelMAE, MASE and GMRAE) indicated Optimal combination performs better than ADL and Bottom-up. But ADL on pooled data outperformed Bottom-up, which showed that pooling the data initially was better than estimating it individually as in the Bottom-up approach. However, UMBRAE rank slightly different from RelMAE, MASE, and GMRAE but the same with MAPE as in Table 4; where ADL outperformed Hierarchical Forecasting, but still optimal combination outperformed bottom up. These findings aligned with Zhao et al. (2020) where pooling the data and using Random Forest enhanced the dengue prediction. Other studies also have highlighted the significant of data pooling (Matthias & Feng, 2022; Petropoulos et al., 2022; Siriysatien et al., 2018; Spiliotis et al., 2019). In this study pooling the data increased the number of dengue cases as well as other predictors, thus contributing to the sufficient numbers of data points needed in the dengue modelling process: previously Rohani et al. (2011), only established the model using 2 factors (epidemiological and environmental) due to insufficient number of dengue cases at small areas.

Table 4 also displays percentage accuracy calculated using MAPE. The ADL model based on the pooled data has 86% accuracy compared to individual area Selayang (85%) and Bandar Baru Bangi (84%) as in Table 2. Pooling the data has increased the accuracy by 1–2%. Although Artificial Neural Network (ANN) is a machine learning method, its percentage accuracy was not much different from ADL and Hierarchical Forecasting, ranging from 85% to 86%. However, other machine learning methods, Support Vector Machine (SVM) and Random Forest, performed very well with 90% and above. The percentage accuracy of Random Forest was the highest (95%). Past dengue forecasting studies also has highlighted Random Forest as the best model (Carvajal et al., 2018; Ong et al., 2018; Silitonga et al., 2020; Zhao et al., 2020). Random forest performed well here as in other comparisons because of using the bootstrap method (i.e., resampling with replacement) to produce multiple decision trees (Cutler et al., 2011; Fawagreh et al., 2014; Schonlau & Yuyan Zou, 2020; Khaled) where it managed to capture different patterns (i.e., behaviour) in the relationship between notified cases, larvae and rainfall. The first pattern showed that the amount of the last 4 weeks rainfall would increase the last 3 weeks larvae (thus increasing the last 2 weeks mosquito population), finally increasing this week's notified cases. However, another pattern (the second pattern) existed in the relationship due to rainfall. Both previous studies by Rohani et al. (2011 and 2018) have indicated that the rainfall data

**Table 2**  
Estimated autoregressive distributed lag (ADL) model.

| Variables (Predictors)  | Selayang           | Bandar Baru Bangi  | Pool               |
|-------------------------|--------------------|--------------------|--------------------|
| Intercept               | -18.400            | 8.138              | -80.4545           |
| Onset                   | 0.372 <sup>a</sup> | 0.393 <sup>a</sup> | 0.490 <sup>a</sup> |
| Onset1                  | 0.381 <sup>a</sup> | 0.399 <sup>a</sup> | 0.137 <sup>b</sup> |
| Larvae3                 | 0.007 <sup>a</sup> | 0.005 <sup>b</sup> | 0.011 <sup>a</sup> |
| Rainfall4               | -0.021             | 0.017              | 0.055 <sup>a</sup> |
| MinTemp3                | -0.272             | 1.352              | 1.282              |
| MaxTemp3                | 0.705              | -1.323             | 1.305              |
| MaxHumidity3            | 0.039              | 0.051              | -0.181             |
| PCR3                    | 0.624 <sup>b</sup> | 0.727              | 0.262              |
| API4                    | -0.016             | 0.007              | 0.002              |
| InterventionLD1         | 0.061              | 0.058              | 0.113 <sup>a</sup> |
| Adjusted R <sup>2</sup> | 0.8465             | 0.8388             | 0.8506             |
| Information Criterion   | 135.039            | 268.466            | 206.447            |
| Accuracy                | 84.9% (-85%)       | 84.1% (-84%)       | 85.7% (-86%)       |

Target Variable: Notified Dengue Cases.

<sup>a</sup> Significant at 1%.

<sup>b</sup> Significant at 5%.

**Table 3**  
Error measures based on ReMAE, MASE, GMRAE and UMBRAE.

| Forecasting Methods                                        | Lead   | ReMAE      | MASE       | GMRAE      | UMBRAE     |
|------------------------------------------------------------|--------|------------|------------|------------|------------|
| Autoregressive Distributed Lag (ADL)                       | 1      | 0.9195     | 0.5672     | 0.9916     | 0.7497     |
|                                                            | 2      | 0.7535     | 0.5142     | 0.9690     | 0.6261     |
|                                                            | 3      | 0.7473     | 0.548      | 0.9642     | 0.5938     |
|                                                            | 4      | 0.5636     | 0.4496     | 0.9214     | 0.4330     |
|                                                            | Median | 0.7504 (5) | 0.5311 (5) | 0.9666 (5) | 0.6100 (4) |
| Hierarchical Forecasting:<br>Bottom Up using ADL           | 1      | 0.9434     | 0.5819     | 0.9942     | 0.9333     |
|                                                            | 2      | 0.7748     | 0.5288     | 0.9721     | 0.7618     |
|                                                            | 3      | 0.7338     | 0.5381     | 0.962      | 0.6955     |
|                                                            | 4      | 0.5824     | 0.4645     | 0.9257     | 0.5834     |
|                                                            | Median | 0.7543 (6) | 0.5335 (6) | 0.9670 (6) | 0.7287 (6) |
| Hierarchical Forecasting:<br>Optimal Combination using ADL | 1      | 0.912      | 0.5626     | 0.9908     | 0.8221     |
|                                                            | 2      | 0.7442     | 0.5079     | 0.9677     | 0.6559     |
|                                                            | 3      | 0.7247     | 0.5314     | 0.9605     | 0.6024     |
|                                                            | 4      | 0.565      | 0.4507     | 0.9217     | 0.5049     |
|                                                            | Median | 0.7344 (4) | 0.5197 (4) | 0.9641(4)  | 0.5317 (5) |
| Artificial Neural Network (ANN)                            | 1      | 0.6798     | 0.4194     | 0.4208     | 0.6339     |
|                                                            | 2      | 0.6876     | 0.4693     | 0.5841     | 0.5735     |
|                                                            | 3      | 0.6693     | 0.4908     | 0.5093     | 0.6956     |
|                                                            | 4      | 0.866      | 0.6907     | 0.6316     | 0.6037     |
|                                                            | Median | 0.6837 (3) | 0.4800 (3) | 0.5467 (3) | 0.5317 (3) |
| Support Vector Machine (SVM)                               | 1      | 0.5066     | 0.3125     | 0.3331     | 0.4947     |
|                                                            | 2      | 0.4423     | 0.3019     | 0.2557     | 0.394      |
|                                                            | 3      | 0.3579     | 0.2625     | 0.2215     | 0.3282     |
|                                                            | 4      | 0.303      | 0.2417     | 0.1819     | 0.2641     |
|                                                            | Median | 0.4001 (2) | 0.2822 (2) | 0.2386 (2) | 0.3611 (2) |
| Random Forest                                              | 1      | 0.3881     | 0.2394     | 0.2070     | 0.4048     |
|                                                            | 2      | 0.3376     | 0.2304     | 0.2463     | 0.3191     |
|                                                            | 3      | 0.2428     | 0.1781     | 0.1600     | 0.2233     |
|                                                            | 4      | 0.2315     | 0.1847     | 0.1351     | 0.2118     |
|                                                            | Median | 0.2902 (1) | 0.2076 (1) | 0.1835 (1) | 0.2712 (1) |

( ) - Rank.

**Table 4**  
Model accuracy based on MAPE.

| MAPE                                                       | Lead 1  | Lead 2  | Lead 3  | Lead 4  | Median  | Rank | Accuracy (%) |
|------------------------------------------------------------|---------|---------|---------|---------|---------|------|--------------|
| Autoregressive Distributed Lag (ADL)                       | 15.9573 | 13.8988 | 14.7043 | 11.2449 | 14.3016 | 4    | 85.70        |
| Hierarchical Forecasting:<br>Bottom Up using ADL           | 17.0057 | 15.0344 | 15.2610 | 12.7246 | 15.1477 | 6    | 84.85        |
| Hierarchical Forecasting:<br>Optimal Combination using ADL | 16.1123 | 14.0466 | 14.6186 | 11.7883 | 14.3326 | 5    | 85.67        |
| Artificial Neural Network (ANN)                            | 11.6450 | 12.5917 | 15.1991 | 20.7552 | 13.8954 | 3    | 86.10        |
| Support Vector Machine (SVM)                               | 8.5833  | 8.1001  | 6.9631  | 6.1788  | 7.5316  | 2    | 92.47        |
| Random Forest                                              | 6.7609  | 6.2222  | 4.6470  | 4.6981  | 5.4602  | 1    | 94.54        |
| Random Forest Less 2 variables                             | 7.8956  | 8.7415  | 5.4089  | 9.1792  | 8.3186  |      | 91.68        |



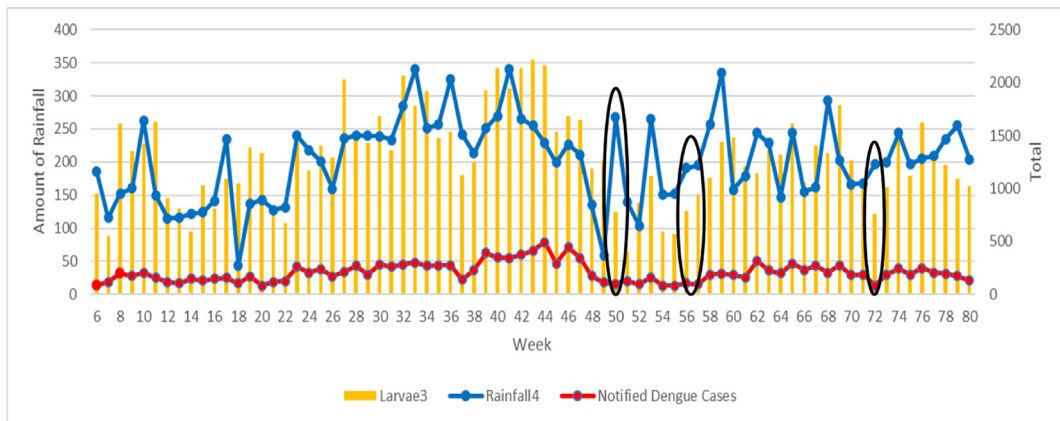


Fig. 4. Total larvae, amount of rainfall and notified dengue cases.

collected was using the amount and not the intensity. Heavy rainfall (i.e., intensity) can contribute to flush out or kill the larvae in the ovitrap (Promprous et al., 2005). As a result, although there was a high amount of rainfall but less larvae, thus less notified cases, which was the reason for the moderate correlation between notified cases and last 4 weeks rainfall (rainfall4: 0.665) as in Table 1. However, based on our data, this situation only occurred less than 10% of the time, as indicated in Fig. 4(week 50, 56, 57 and 72).

Although Random Forest was the best model with 95% accuracy, it is practically very costly and time-consuming to collect mosquito-related data (Larvae3) and do virus screening using RT-PCR (PCR3) as explained in the methodology section. Thus, we removed the two variables (Larvae3 and PCR3) of the entomological factor and implemented Random Forest. Table 4

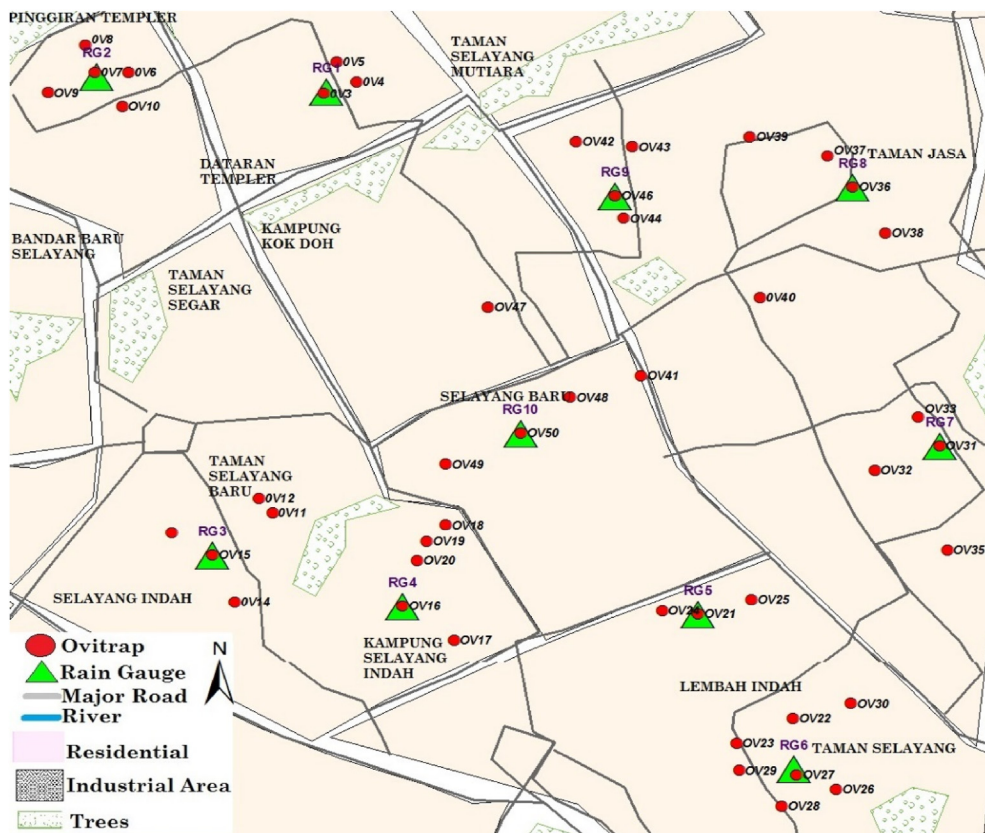


Fig. 5. Location of Rain gauge (Mobile Weather Station) at Selayang.

indicated that the percentage accuracy of Random Forest with these two less variables was 92%, reducing the full model accuracy by 3% but still acceptable to be embedded in the early warning system, which can rely only on just two factors (i.e., epidemiological and environmental). The epidemiological data (notified dengue cases, onset cases, and the number of interventions) can be retrieved automatically by the early warning system through integrating with the e-Dengue system developed by Malaysia Ministry of Health. Meanwhile, the environmental data of rainfall, temperature, humidity, and Air Pollution Index can be collected by the mobile weather station and link automatically with the system through the Internet of Things (IoT) approach. Furthermore, this Random Forest model can forecast up to 4 weeks in advance (based on leads 1,2,3 and 4), giving time for the authority to act before dengue cases occur where presently intervention only happens after notified dengue cases increase. However, the real-time operationalization of this system also requires several rain gauges to be located at the dengue hot spot because the rainfall pattern is very localized as explained in the previous section. Thus, further analysis was conducted in determining the distances of the rain gauges.

Figs. 5 and 6 show the location of the 10 and 11 rain gauges at Selayang (RG1 to RG10) and Bandar Baru Bangi (RG11 to RG21), respectively (Rohani et al., 2018). The rain gauges were spread across the two large areas to measure the localize amount of rainfall. Tables 5 and 6 display the distance of the rain gauges (upper diagonal) and correlation of the amount of rainfall (lower diagonal) for both areas. In Selayang, there was a strong correlation (0.893) between RG1 and RG2 when the distance was 0.97 km but a weak correlation (0.162) between RG1 and RG6 when the distance was 6.14 km. A similar finding was obtained for Bandar Baru Bangi, where strong correlation (0.871) between RG11 and RG13 when the distance was 1.13 km but a weak correlation (0.078) between RG11 and RG15 when the distance was 4.73 km. These correlations revealed the localized pattern of rainfall amount at the two areas based on the distance of the rain gauges, as shown in Fig. 7 and Fig. 8. There was an obvious pattern when the rain gauges were near to each other, the amounts of rainfall were similar, but when they were far away, the amounts of rainfall were different. Other rain gauges also showed similar findings. Thus, indicating the existence of localized rainfall in both areas, which often occurs in Malaysia (Muhammad et al., 2020; Singh et al., 2022). Next, we extracted the moderate correlation and respective distances as presented in Table 7 for both areas. The mean rain gauge distance was 3.34 km with a 95% confidence interval of 2.82–3.85 km. Based on this, it is recommended that the rain

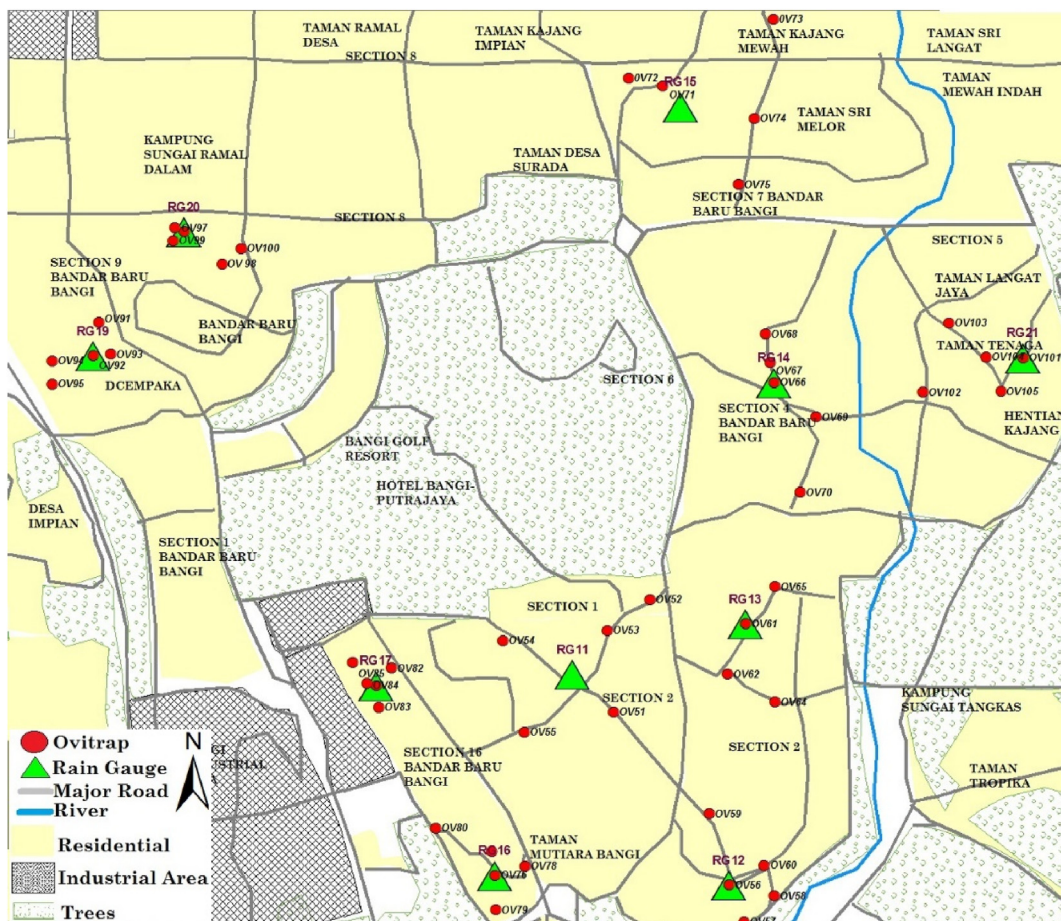


Fig. 6. Location of Rain gauge (Mobile Weather Station) at Bandar Baru Bangi.

**Table 5**  
Rain gauge distance in kilometre (KM) and correlation (Selayang).

| Distance (KM) | RG1                  | RG2               | RG3               | RG4               | RG5               | RG6               | RG7               | RG8               | RG9               | RG10              |      |
|---------------|----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------|
| Correlation   | <b>Distance (KM)</b> |                   |                   |                   |                   |                   |                   |                   |                   |                   |      |
| <b>RG1</b>    | Correlation          | –                 | 0.97              | 3.88              | 4.37              | 4.85              | 6.14              | 4.04              | 2.59              | 1.62              | 2.91 |
| <b>RG2</b>    |                      | .893 <sup>a</sup> | –                 | 4.20              | 4.69              | 5.34              | 6.79              | 5.01              | 3.56              | 2.59              | 3.72 |
| <b>RG3</b>    |                      | .483 <sup>a</sup> | .484 <sup>a</sup> | –                 | 0.89              | 2.26              | 3.23              | 3.49              | 4.38              | 3.56              | 1.78 |
| <b>RG4</b>    |                      | .481 <sup>a</sup> | .485 <sup>a</sup> | .773 <sup>a</sup> | –                 | 1.46              | 2.26              | 2.91              | 4.21              | 3.57              | 1.63 |
| <b>RG5</b>    |                      | .411 <sup>a</sup> | .423 <sup>a</sup> | .477 <sup>a</sup> | .476 <sup>a</sup> | –                 | 1.45              | 1.78              | 3.72              | 3.55              | 1.77 |
| <b>RG6</b>    |                      | .162*             | .188*             | .269 <sup>a</sup> | .306 <sup>a</sup> | .375 <sup>a</sup> | –                 | 2.91              | 5.02              | 5.01              | 3.22 |
| <b>RG7</b>    |                      | .314 <sup>a</sup> | .321 <sup>a</sup> | .351 <sup>a</sup> | .388 <sup>a</sup> | .267 <sup>a</sup> | .185*             | –                 | 2.27              | 2.59              | 1.94 |
| <b>RG8</b>    |                      | .508 <sup>a</sup> | .512 <sup>a</sup> | .492 <sup>a</sup> | .519 <sup>a</sup> | .363 <sup>a</sup> | .286 <sup>a</sup> | .431 <sup>a</sup> | –                 | 1.13              | 2.58 |
| <b>RG9</b>    |                      | .521 <sup>a</sup> | .532 <sup>a</sup> | .493 <sup>a</sup> | .535 <sup>a</sup> | .396 <sup>a</sup> | .311 <sup>a</sup> | .408 <sup>a</sup> | .886 <sup>a</sup> | –                 | 2.1  |
| <b>RG10</b>   |                      | .392 <sup>a</sup> | .390 <sup>a</sup> | .608 <sup>a</sup> | .634 <sup>a</sup> | .410 <sup>a</sup> | .221 <sup>a</sup> | .344 <sup>a</sup> | .520 <sup>a</sup> | .685 <sup>a</sup> | –    |

<sup>a</sup> Sig. at 1%, \* Sig. at 5%.

gauges should be located about 3–4 km apart at the dengue hot spot in ensuring the accuracy of the amount of rainfall collected for the dengue outbreak forecasting model to be embedded in the early warning system.

Fig. 9 displays the early warning system conceptual framework using the Internet of Things (IoT) approach by integrating the epidemiological data readily available in the e-Dengue system with the environmental data through automated mobile weather stations (around USD500 each) located at the dengue hot spots. In order to implement a real-time warning system, the Random Forest forecasting model needs to be transformed into an algorithm and embedded in the e-Dengue system by including the parameters as listed in Fig. 9. Real time notification of a potential dengue outbreak will then be alerted to the Health Department and COMBI, a community engagement group set up by Ministry of Health (My Health, 2017) to prevent dengue by destroying the *Aedes* breeding sites. Thus, the proposed e-Dengue early warning system can save lives because prevention is better than cure. According to Lee et al. (2010), the cost of dengue patient hospitalization (dengue illness cost) in Malaysia are 11 times the amount of government spending on *Aedes* vector control. This relationship indicates that increased investment on prevention could potentially reduce these illness costs. The study also found that dengue may also adversely impact tourism and create emotional and long-term burdens on families affected by illness and deaths. Md Shahin et al. (2016) stated that the mean total cost per case of dengue infection was USD365.16 and average duration of dengue illness was 9.69 days. All of these negative impacts due to dengue can be avoided through the proposed e-Dengue early warning system which is very practical and cost effective using the Internet of Things (IoT) approach.

#### 4. Conclusion

The journey in searching of Malaysia dengue outbreak forecasting model by integrating 3 important factors (epidemiological, entomological, and environmental) using scientific experimental field studies started 15 years ago due to the many past studies that have failed to establish a practical model to be embedded in an early warning system. One of the reasons was using only the epidemiological factor (dengue cases) related to seasonal patterns where the impact of climate change has affected the consistency of the rainfall season in Malaysia. Another reason for failure, even when, using just two factors: epidemiological (dengue cases) and environmental (weather data), arose due to using weather data from Meteorological Department (secondary data), jeopardized the accuracy of forecasting model because of the localized rainfall in Malaysia. Hence, to overcome all these limitations, two scientific experimental field studies were conducted to establish the relationship of the 3 factors. However, the first study (Rohani et al., 2011), only managed to relate 2 factors (entomological (mosquito related data) and environmental (weather data)) because of reliance on small localities contributed to insufficient

**Table 6**  
Rain gauge distance in kilometre (KM) and correlation (Bandar Baru Bangi).

| Distance (KM) | RG11                 | RG12              | RG13              | RG14              | RG15              | RG16              | RG17              | RG18              | RG19              | RG20              | RG21              |      |
|---------------|----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------|
| Correlation   | <b>Distance (KM)</b> |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |      |
| <b>RG11</b>   | Correlation          | –                 | 1.96              | 1.31              | 2.44              | 4.73              | 1.79              | 1.41              | 2.76              | 3.25              | 3.57              | 3.26 |
| <b>RG12</b>   |                      | .797 <sup>a</sup> | –                 | 2.12              | 3.41              | 5.03              | 1.47              | 2.44              | 1.79              | 4.54              | 5.03              | 3.91 |
| <b>RG13</b>   |                      | .871 <sup>a</sup> | .756 <sup>a</sup> | –                 | 1.95              | 4.58              | 2.46              | 2.11              | 3.09              | 3.73              | 3.91              | 2.45 |
| <b>RG14</b>   |                      | .744 <sup>a</sup> | .720 <sup>a</sup> | .763 <sup>a</sup> | –                 | 3.28              | 3.71              | 2.93              | 4.55              | 3.43              | 3.11              | 1.63 |
| <b>RG15</b>   |                      | 0.078             | 0.108             | 0.074             | 0.149             | –                 | 5.05              | 4.82              | 5.84              | 3.62              | 3.43              | 3.94 |
| <b>RG16</b>   |                      | .776 <sup>a</sup> | .741 <sup>a</sup> | .770 <sup>a</sup> | .720 <sup>a</sup> | 0.078             | –                 | 1.78              | 1.30              | 4.06              | 4.54              | 4.21 |
| <b>RG17</b>   |                      | .810 <sup>a</sup> | .746 <sup>a</sup> | .804 <sup>a</sup> | .711 <sup>a</sup> | 0.067             | .815 <sup>a</sup> | –                 | 2.62              | 2.95              | 3.24              | 3.74 |
| <b>RG18</b>   |                      | .723 <sup>a</sup> | .654 <sup>a</sup> | .721 <sup>a</sup> | .697 <sup>a</sup> | 0.038             | .820 <sup>a</sup> | .754 <sup>a</sup> | –                 | 4.70              | 5.35              | 5.02 |
| <b>RG19</b>   |                      | .415 <sup>a</sup> | .395 <sup>a</sup> | .433 <sup>a</sup> | .515 <sup>a</sup> | .228 <sup>a</sup> | .465 <sup>a</sup> | .461 <sup>a</sup> | .433 <sup>a</sup> | –                 | 1.32              | 4.38 |
| <b>RG20</b>   |                      | .372 <sup>a</sup> | .352 <sup>a</sup> | .398 <sup>a</sup> | .488 <sup>a</sup> | .238 <sup>a</sup> | .441 <sup>a</sup> | .431 <sup>a</sup> | .422 <sup>a</sup> | .867 <sup>a</sup> | –                 | 4.05 |
| <b>RG21</b>   |                      | .542 <sup>a</sup> | .537 <sup>a</sup> | .583 <sup>a</sup> | .680 <sup>a</sup> | 0.119             | .576 <sup>a</sup> | .538 <sup>a</sup> | .573 <sup>a</sup> | .436 <sup>a</sup> | .430 <sup>a</sup> | –    |

<sup>a</sup> Sig. at 1%, \* Sig. at 5%.

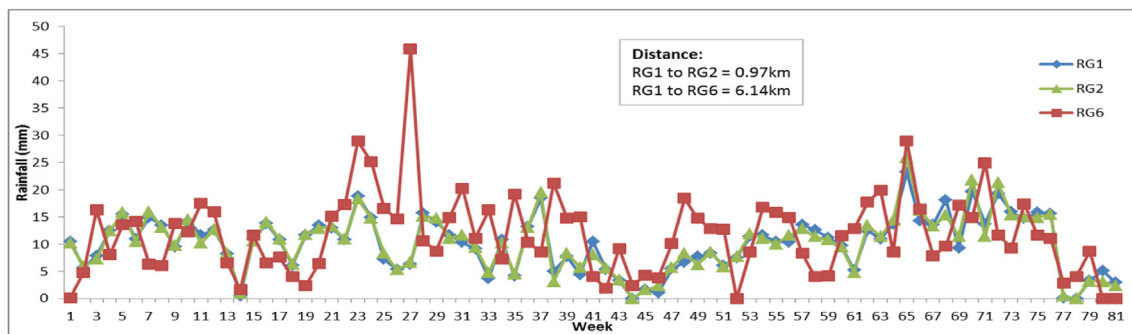


Fig. 7. Rainfall amount pattern of near and far rain gauge (Selayang).

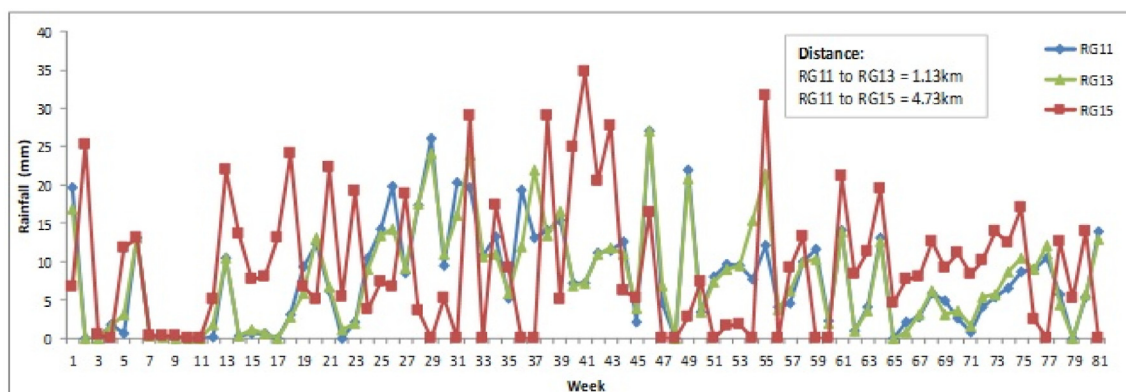


Fig. 8. Rainfall amount pattern of near and far rain gauge (Bandar Baru Bangi).

number of dengue cases (epidemiological). The second study was conducted at two large high dengue cases areas and successfully established specific relationships of the 3 factors and obtained 85% of accuracy using an ADL model (Rohani et al., 2018). But higher accuracy of at least 90% and above were required for embedding the forecasting model into dengue outbreak early warning system (e-Dengue).

Thus, the search continued in this study using Rohani et al. (2018) data and six forecasting methods (Autoregressive Distributed Lag (ADL), Hierarchical Forecasting (Bottom-up and Optimal combination) and Machine Learning methods (Artificial Neural Network (ANN), Support Vector Machine (SVM) and Random Forest) with the aimed of at least 90% accuracy.

**Table 7**  
Moderate correlation & distance (KM).

| Area              | Rain Gauge                                             | Moderate Correlation | Distance (KM)              |
|-------------------|--------------------------------------------------------|----------------------|----------------------------|
| Selayang          | RG1 & RG8                                              | 0.508                | 2.59                       |
|                   | RG1 & RG9                                              | 0.521                | 1.62                       |
|                   | RG2 & RG8                                              | 0.512                | 3.56                       |
|                   | RG2 & RG9                                              | 0.532                | 2.59                       |
|                   | RG4 & RG8                                              | 0.519                | 4.21                       |
|                   | RG4 & RG9                                              | 0.535                | 3.57                       |
|                   | RG8 & RG10                                             | 0.520                | 2.58                       |
| Bandar Baru Bangi | RG19 & RG14                                            | 0.515                | 3.43                       |
|                   | RG21 & RG11                                            | 0.542                | 3.91                       |
|                   | RG21 & RG12                                            | 0.537                | 2.45                       |
|                   | RG21 & RG13                                            | 0.583                | 4.21                       |
|                   | RG21 & RG16                                            | 0.576                | 3.74                       |
|                   | RG21 & RG17                                            | 0.538                | 5.02                       |
|                   | RG21 & RG18                                            | 0.573                | 3.26                       |
| Overall           | Mean                                                   |                      | 3.34                       |
|                   | Standard Deviation                                     |                      | 0.89                       |
|                   | 95% Confidence Interval for Mean (Lower & Upper Bound) |                      | <b>2.82</b><br><b>3.85</b> |

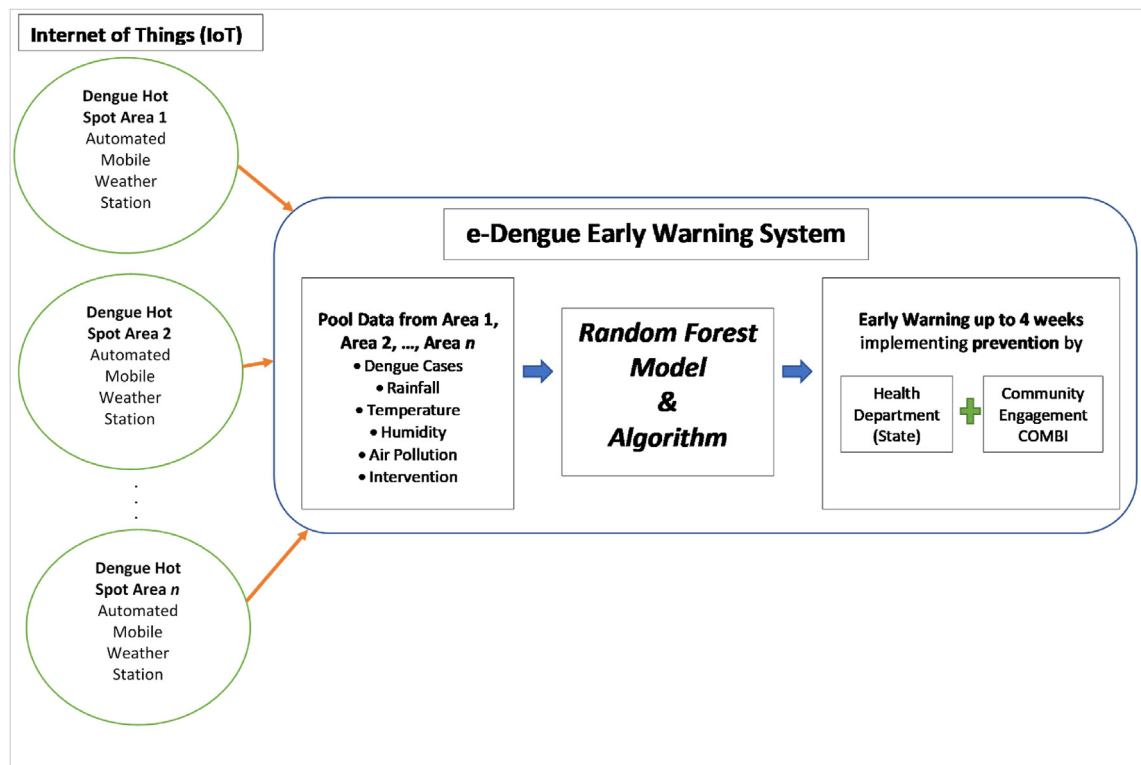


Fig. 9. e-dengue early warning system conceptual framework.

Successfully, Random Forest proved the most appropriate for forecasting dengue outbreak in Malaysia with 95% accuracy when using the 3 factors, but it was very costly and time consuming to collect mosquito related data (entomological). Thus, only 2 factors (epidemiological and environmental) were used in the model, and while the accuracy dropped to 92%, this was still acceptable. However, the localized rainfall requires practically several rain gauges to be located at the dengue hot spot areas about 3–4 km apart; in ensuring the accuracy of the forecasting model when embedded in the early warning system using an Internet of Things (IoT) approach. The e-Dengue early warning system integrating the epidemiological data readily available in the system with the environmental data through automated mobile weather stations (as the IoT objects/sensors) located at the dengue hot spots and using Random Forest to forecast up to 4 weeks the dengue cases. Real time notification of potential dengue outbreak will then be alerted to Health Department and COMBI, a community engagement group set up by Ministry of Health to prevent dengue by destroying the *Aedes* breeding sites. Thus, the proposed practical and cost-effective e-Dengue early warning system can save lives because prevention is better than cure. The research contribution here is to show how recent advances in forecasting methods when combined to the careful and practice-based simplification of the included drivers of dengue outbreaks has led to a successful implementable warning system.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This study is a collaboration between Universiti Utara Malaysia (UUM), Centre for Marketing Analytics and Forecasting (CMAF), Lancaster University, Institute for Medical Research (IMR), and Disease Control Division, Ministry of Health Malaysia. We are very grateful to UUM for providing the funding, CMAF for forecasting expertise, IMR for the data and entomological expertise, and Disease Control Division for epidemiological and ground operationalization expertise.

## References

- Ahmad Khairul Hakim, M., Nazri, C. D., Siti Nazrina, C., & Sharifah Norkhadajah, S. I. (2020). Correlational analysis of air pollution index levels on dengue surveillance data: A retrospective study in melaka, Malaysia. *Journal of Sustainability Science and Management*, 15(8), 114–121. <https://jssm.umt.edu.my/wp-content/uploads/sites/51/2020/12/10-15.8.pdf>.
- Alto, B. W., & Bettinardi, D. (2013). Temperature and dengue virus infection in mosquitoes: Independent effects on the immature and adult stages. *The American Journal of Tropical Medicine and Hygiene*, 88(3), 497–505. <https://doi.org/10.4269/ajtmh.12-0421>
- Ang, K. C., & Li, Z. (2002). Modeling the spread of dengue in Singapore. In , Vol. 2. *Conference proceedings for the international congress on modeling and simulation 1999, Hamilton, New Zealand, december 1999* (pp. 555–560).
- Anyamba, A., Chretien, J. P., Small, J., Tucker, C. J., & Linthicum, K. J. (2006). Developing global climate anomalies suggest potential disease risks for 2006–2007. *International Journal of Health Geographics*, 5(60). <https://doi.org/10.1186/1476-072X-5-60>
- Banulata, G., Latifah, S. Y., Nissa Nabila, A. R., & Mustafa, M. (2021). Association of temperature and rainfall with *Aedes* mosquito population in 17th College of Universiti Putra Malaysia. *Malaysian Journal of Medicine and Health Sciences*, 17(2), 78–84. [https://medic.upm.edu.my/upload/dokumen/2021040613061511\\_MJMjHS\\_0088.pdf](https://medic.upm.edu.my/upload/dokumen/2021040613061511_MJMjHS_0088.pdf).
- Barbazan, P., Yoksan, S., & Gonzalez, J. P. (2002). Dengue hemorrhagic fever epidemiology in Thailand: Description and forecasting of epidemics. *Microbes and Infection*, 4(7), 699–705. [https://doi.org/10.1016/s1286-4579\(02\)01589-7](https://doi.org/10.1016/s1286-4579(02)01589-7)
- Bartley, L. M., Donnelly, C. A., & Garnett, G. P. (2002). The seasonal pattern of dengue in endemic areas: Mathematical models of mechanisms. *Trans R Soc Trop Med Hyg Jul-Aug*, 96(4), 387–397.
- Blokdyk, G. (2020). *Artificial neural network: A complete guide* (2020 Ed.). Brendale, Australia: 5STARCOoks.
- Brownlee, J. (2020). *Machine learning mastery with WEKA*. eBook <https://machinelearningmastery.com/machine-learning-mastery-weka>.
- Carneiro, M., Alves, B., Gehrke, F. S., Domingues, J. N., Sá, N., Paixão, S., Figueiredo, J., Ferreira, A., Almeida, C., Machi, A., Savóia, E., Nascimento, V., & Fonseca, F. (2017). Environmental factors can influence dengue reported cases. *Revista da Associação Médica Brasileira*, 63(11), 957–961. <https://doi.org/10.1590/1806-9282.63.11.957>, 1992.
- Carrington, L. B., & Simmons, C. P. (2014). Human to mosquito transmission of dengue viruses. *Frontiers in Immunology*, 5, 290. <https://doi.org/10.3389/fimmu.2014.00290>
- Carvajal, T. M., Viacrusis, K. M., Hernandez, L. F. T., Ho, H. T., Amalin, D. M., & Watanabe, K. (2018). Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. *BMC Infectious Diseases*, 18, 183. <https://doi.org/10.1186/s12879-018-3066-0>
- Cheah, W. L., Chang, M. S., & Wang, Y. C. (2006). Spatial, environmental and entomological risk factors analysis on a rural dengue outbreak in Lundu District in Sarawak, Malaysia. *Tropical Biomedicine*, 23(1), 85–96. PMID: 17041556.
- Cheng, K. Y. (2006). *Quantitative forecasting trend of dengue fever and dengue haemorrhagic fever cases in Peninsular Malaysia*. Universiti Sains Malaysia: Masters thesis.
- Chen, C., Twycross, J., & Garibaldi, J. M. (2017). A new accuracy measure based on bounded relative error for time series forecasting. *PLoS One*, 12(3), Article e0174202. <https://doi.org/10.1371/journal.pone.0174202>
- Chiung, C. H., Choo-Yee, T., & Dhesi, B. R. (2018). Using public open data to predict dengue epidemic: Assessment of weather variability, population density, and land use as predictor variables for dengue outbreak prediction using Support vector machine. *Indian Journal of Science and Technology*, 11(4). <https://doi.org/10.17485/ijst/2018/v11i4/115405>
- Clements, A. N. (1992). *The biology of mosquitoes: Development, nutrition and reproduction*. London: Chapman & Hall.
- Colon-Gonzalez, F. J., Soares Bastos, L., Hofmann, B., Hopkin, A., Harpham, Q., Crocker, T., et al. (2021). Probabilistic seasonal dengue forecasting in vietnam: A modelling study using super-ensembles. *PLoS Medicine*, 18(3), Article e1003542. <https://doi.org/10.1371/journal.pmed.1003542>
- da Costa, C. F., Dos Passos, R. A., Lima, J., Roque, R. A., de Souza Sampaio, V., Campolina, T. B., Secundino, N., & Pimenta, P. (2017). Transovarial transmission of DENV in *Aedes aegypti* in the amazon basin: A local model of xenomonitoring. *Parasites & Vectors*, 10(1), 249. <https://doi.org/10.1186/s13071-017-2194-5>
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2011). Random forests. In C. Zhang, & Y. Ma (Eds.), *Ensemble machine learning* (pp. 157–176). New York: Springer.
- Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29, 510–522.
- Drakou, K., Nikolaou, T., Vasquez, M., Petric, D., Michaelakis, A., Kapranas, A., Papatheodoulou, A., & Koliou, M. (2020). The effect of weather variables on mosquito activity: A snapshot of the main point of entry of Cyprus. *International Journal of Environmental Research and Public Health*, 17(4), 1403. <https://doi.org/10.3390/ijerph17041403>
- Fansiri, T., Buddhari, D., Pathawang, N., Pongsiri, A., Klungthong, C., Iamsirithaworn, S., Jones, A. R., Fernandez, S., Srikiatkachorn, A., Rothman, A. L., Anderson, K. B., Thomas, S. J., Endy, T. P., & Ponlawat, A. (2021). Entomological risk assessment for dengue virus transmission during 2016–2020 in kamphaeng phet, Thailand. *Pathogens*, 10(10), 1234. <https://doi.org/10.3390/pathogens10101234>
- Fatimah, I., Mohd Nasir, T., Wan Abu Bakar, W. A., Guan, C. C., & Saadiah, S. (2005). A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN). *Computer Methods and Programs in Biomedicine*, 79(3), 273–281. <https://doi.org/10.1016/j.cmpb.2005.04.002>
- Fawagreh, K., Mohamed Medhat, G., & Eyad, E. (2014). Random forests: From early developments to recent advancements. *Systems Science & Control Engineering*, 2(1), 602–609. <https://doi.org/10.1080/21642583.2014.956265>
- Genuer, R., & Poggi, J. M. (2020). *Random forests with R* (p. 108). Switzerland: Springer Cham.
- Guo, P., Liu, T., Zhang, Q., Wang, L., Xiao, J., Zhang, Q., Luo, G., Li, Z., He, J., Zhang, Y., & Ma, W. (2017). Developing a dengue forecast model using machine learning: A case study in China. *PLoS Neglected Tropical Diseases*, 11(10), Article e0005973. <https://doi.org/10.1371/journal.pntd.0005973>
- Hales, S., de Wet, N., Maingon, J., & Woodward, A. (2002). Potential effect of population and climate changes on global distribution of dengue fever: An empirical model. *Lancet*, 360(9336), 830–834. [https://doi.org/10.1016/S0140-6736\(02\)09964-6](https://doi.org/10.1016/S0140-6736(02)09964-6)
- Harrison, R. E., Brown, M. R., & Strand, M. R. (2021). Whole blood and blood components from vertebrates differentially affect egg formation in three species of anautogenous mosquitoes. *Parasites & Vectors*, 14, 119. <https://doi.org/10.1186/s13071-021-04594-9>
- Hartley, L. M., Donnelly, C. A., & Garnett, G. P. (2002). The seasonal pattern of dengue in endemic areas: Mathematical models of mechanisms. *Transactions of the Royal Society of Tropical Medicine & Hygiene*, 96(4), 387–397. [https://doi.org/10.1016/S0035-9203\(02\)90371-8](https://doi.org/10.1016/S0035-9203(02)90371-8)
- Hii, Y. L., Zhu, H., Ng, N., Ng, L. C., & Rocklöv, J. (2012). Forecast of dengue incidence using temperature and rainfall. *PLoS Neglected Tropical Diseases*, 6(11), Article e1908. <https://doi.org/10.1371/journal.pntd.0001908>
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed). OTexts: Melbourne, Australia. OTexts.com/fpp3.
- Jain, R., Sontisirikit, S., Iamsirithaworn, S., & Prendinger, H. (2019). Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. *BMC Infectious Diseases*, 19(1), 272. <https://doi.org/10.1186/s12879-019-3874-x>
- Johansson, M. A., Apfeldorf, K. A., Dobson, S., Devita, J., Buczak, A. L., Baugher, B., et al. (2019). An open challenge to advance probabilistic forecasting for dengue epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 116(48). <https://doi.org/10.1073/pnas.1909865116>
- Johari, N. A., Voon, K., Toh, S. Y., Sulaiman, L. H., Yap, I. K. S., & Lim, P. K. C. (2019). Sylvatic dengue virus type 4 in *Aedes aegypti* and *Aedes albopictus* mosquitoes in an urban setting in Peninsular Malaysia. *PLoS Neglected Tropical Diseases*, 13(11), Article e0007889. <https://doi.org/10.1371/journal.pntd.0007889>
- Joshi, V., Mourya, D. T., & Sharma, R. C. (2002). Persistence of dengue-3 virus through transovarial transmission passage in successive generations of *Aedes aegypti* mosquitoes. *The American Journal of Tropical Medicine and Hygiene*, 67(2), 158–161. <https://doi.org/10.4269/ajtmh.2002.67.158>
- Koutsandreas, D., Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2022). On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*, 73(5), 937–954. <https://doi.org/10.1080/01605682.2021.1892464>
- Kowalczyk, A. (2017). *Support vector machines succinctly* (p. 114pp). USA: Synfusion, Inc.

- Kumarasamy, V. (2006). Dengue fever in Malaysia: Time for review? *Medical Journal of Malaysia*, 61(10), 1–3.
- Lee, H. L., Vasanb, S. S., Birgelenc, L., Murtolae, T. M., Hong-Fei, G., Fielddet, R. W., et al. (2010). Immediate cost of dengue to Malaysia and Thailand: An estimate. *Dengue Bulletin*, 34, 65–76. <https://apps.who.int/iris/bitstream/handle/10665/170972/db2010v34p65.pdf?sequence=1&isAllowed=y>.
- Lim, C. S., & Poh, C. L. (2018). Development of dengue vaccines. *Australasian Medical Journal*, 11(6), 370–380. <https://doi.org/10.21767/AMJ.2018.3451>
- Liu, K., Yin, L., Zhang, M., Kang, M., Deng, A. P., Li, Q. L., & Song, T. (2021). Facilitating fine-grained intra-urban dengue forecasting by integrating urban environments measured from street-view images. *Infectious Diseases Poverty*, 10, 40. <https://doi.org/10.1186/s40249-021-00824-5>
- Luz, P. M., Codeco, C. T., Massad, E., & Struchiner, C. J. (2003). Uncertainties regarding dengue modeling in Rio de Janeiro, Brazil. *Memorias do Instituto Oswaldo Cruz*, 98(7), 871–878.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS One*, 13(3), Article e0194889. <https://doi.org/10.1371/journal.pone.0194889>
- Malaysian Health Technology Assessment Section (MaHTAS). (2015). CPG management of dengue infection in adults (third edition), 82 pp. Putrajaya, Malaysia: MaHTAS. Retrieved from: <https://www.moh.gov.my/moh/resources/penerbitan/CPG/CPG%20Dengue%20Infection%20PDF%20Final.pdf>.
- Malaysian Health Technology Assessment Section (MaHTAS). (2019). Integrated vector management for Aedes control. *Health Technology Assessment report*, 272. Putrajaya, Malaysia: MaHTAS. Retrieved from: <http://www.moh.gov.my/moh/resources/Penerbitan/MAHTAS/HTA/Report%20HTA%20IVM%20for%20Aedes%20Control.pdf>.
- Malinda, M., Rohani, A., Noor Azleen, M. K., Wan Najdah, W. M. A., Suzilah, I., & Lee, H. L. (2012). Climatic influences on *Aedes* mosquito larvae population. *Malaysian Journal of Science*, 31(1), 30–39. <https://doi.org/10.22452/mjs.vol31no1.4>
- Masnita, M. Y., Dom, N. C., Ismail, R., & Zainuddin, A. (2018). Assessing the temporal distribution of dengue vectors mosquitoes and its relationship with weather variables. *Serangga*, 23(1), 112–125. <http://ejournals.ukm.my/serangga/issue/view/1099/showToc>.
- Matthias, A., & Feng, L. (2022). Hierarchical forecasting with a top-down alignment of independent level forecasts. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2021.12.015>
- Md Shahin, M., Begum, R. A., Er, A. C., & Pereira, J. J. (2016). Assessing the cost burden of dengue infection to households in Seremban, Malaysia. *Southeast Asian Journal of Tropical Medicine & Public Health*, 47(6), 1167–1176. PMID: 29634177.
- Mohd Alias, L. (2013). *Introductory business forecasting. A practical approach (3rd ed)*. Shah alam: Pusat penerbitan Universiti, Universiti teknologi mara.
- Muhammad, N. S., Abdullah, J., & Julien, P. Y. (2020). Characteristics of rainfall in peninsular Malaysia. *Journal of Physics: Conference Series*, 1529(5), Article 052014. <https://doi.org/10.1088/1742-6596/1529/5/052014>
- My Health. (2017). *The importance of community involvement in COMBI Program*. <http://www.myhealth.gov.my/en/importance-community-involvement-combi-program/>.
- Nor Aliza, A. R., Harvie, S., Nur Ain, M. R., Lela, S., Siti Fairouz, I., Razitasham, S., & Marlini, O. (2019). Detection of transovarial dengue viruses in *Aedes albopictus* from selected localities in Kuching and Samarahan divisions, Sarawak, Malaysia by reverse transcription polymerase chain reaction (RT-PCR). *Serangga*, 24(2), 145–158.
- Ong, J., Liu, X., Rajarethinam, J., Kok, S. Y., Liang, S., Tang, C. S., et al. (2018). Mapping dengue risk in Singapore using Random Forest. *PLoS Neglected Tropical Diseases*, 12(6), Article e0006587. <https://doi.org/10.1371/journal.pntd.0006587>
- Patil, S., & Pandya, S. (2021). Forecasting dengue hotspots associated with variation in meteorological parameters using regression and time series models. *Frontiers in Public Health*, 9, Article 798034. <https://doi.org/10.3389/fpubh.2021.798034>
- Peña-García, V. H., Triana-Chávez, O., Mejía-Jaramillo, A. M., Díaz, F. J., Gómez-Palacio, A., & Arboleda-Sánchez, S. (2016). Infection rates by dengue virus in mosquitoes and the influence of temperature may be related to different endemicity patterns in three Colombian cities. *International Journal of Environmental Research and Public Health*, 13(7), 734. <https://doi.org/10.3390/ijerph13070734>
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., ... Ziel, F. (2022). Forecasting : Theory and practice. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2021.11.001>
- Promprous, S., Jaroensutasinee, M., & Jaroensutasinee, K. (2005). Climatic factors affecting Dengue haemorrhagic fever incidence in Southern Thailand. *Dengue Bulletin*, 29, 41–49. <https://apps.who.int/iris/handle/10665/164135>.
- Ram, S., Khurana, S., Kaushal, V., Gupta, R., & Khurana, S. B. (1998). Incidence of dengue fever in relation to climatic factors in Ludhiana, Punjab. *Indian Journal of Medical Research*, 108, 128–133.
- Rohani, A., Aidil Azahary, A. D., Malinda, M., Zurainee, M. N., Rozilawati, H., Wan Najdah, W. M. A., & Lee, H. L. (2014). Eco-virological survey of *Aedes* mosquito larvae in selected dengue outbreak areas in Malaysia. *Journal of Vector Borne Diseases*, 51, 327–332.
- Rohani, A., Suzilah, I., Malinda, M., Anuar, I., Topek, O., & Tanrang, Y. (2011). *Aedes* larval population dynamics and risk for dengue epidemics in Malaysia. *Tropical Biomedicine*, 28(2), 237–248.
- Rohani, A., Suzilah, I., Wan Najdah, W. M. A., Topek, O., Mustafakamal, I., & Lee, H. L. (2018). Factors determining dengue outbreak in Malaysia. *PLoS One*, 13(2). <https://doi.org/10.1371/journal.pone.0193326>
- Rosa-Freitas, M. G., Schreiber, K. V., Tsouris, P., Weimann, E. T., & Luitgards-Moura, J. F. (2006). Associations between dengue and combinations of weather factors in a city in the Brazilian Amazon. *Pan American Journal of Public Health*, 20(4), 256–267. <https://doi.org/10.1590/s1020-49892006000900006>
- Salim, N. A. M., Wah, Y. B., Reeves, C., Smith, M., Wan Fairros, W. Y., Rose Nani, M., et al. (2021). Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Scientific Reports*, 11, 939. <https://doi.org/10.1038/s41598-020-79193-2>
- Sanchez-Gendriz, I., de Souza, G. F., de Andrade, I., Neto, A., de Medeiros Tavares, A., Barros, D., de Moraes, A., Galvão-Lima, L. J., & de Medeiros Valentim, R. A. (2022). Data-driven computational intelligence applied to dengue outbreak forecasting: A case study at the scale of the city of natal, RN-Brazil. *Scientific Reports*, 12(1), 6550. <https://doi.org/10.1038/s41598-022-10512-5>
- Schonlau, M., & Yuyan Zou, R. (2020). The random forest algorithm for statistical learning. *STATA Journal*, 20(1), 3–29. <https://doi.org/10.1177/1536867X20909688>
- Seidahmed, O. M. E., & Eltahir, E. A. B. (2016). A sequence of flushing and drying of breeding habitats of *Aedes aegypti* (L.) prior to the low dengue season in Singapore. *PLoS Neglected Tropical Diseases*, 10(7), Article e0004842. <https://doi.org/10.1371/journal.pntd.0004842>
- Seng, S. B., Chong, A. K., & Moore, A. (2005). Geostatistical modelling, analysis and mapping of epidemiology of dengue fever in Johor State, Malaysia. Presented at the 17th Annual Colloquium of the Spatial Information Research Centre (SIRC 2005: A Spatio-temporal Workshop).
- Shamsul, A. S., Jamiatul Aida, M. S., Mohd Rohaizat, H., Nazarudin, S., Azimatun, N. A., & Rozita, H. (2012). Relationships between *Aedes* indices and dengue outbreaks in Selangor, Malaysia. *Dengue Bulletin*, 36.
- Sheng-Qun, D., Xian, Y., Yong, W., Jia-Ting, C., Xiao-Jun, W., & Hong-Juan, P. (2020). A review on dengue vaccine development. *Vaccines*, 8(1), 63. <https://doi.org/10.3390/vaccines8010063>
- Silitonga, P., Dewi, B. E., Bustamam, A., & Al-Ash, H. S. (2021). Evaluation of dengue model performances developed using artificial neural network and random forest classifiers. *Procedia Computer Science*, 179, 135–143. <https://doi.org/10.1016/j.procs.2020.12.018>
- Singh, S., Hergn, L. C., Sulaiman, L. H., Wong, S. F., Jelip, J., Mokhtar, N., Harpham, Q., Tsarouchi, G., & Gill, B. S. (2022). The effects of meteorological factors on dengue cases in Malaysia. *International Journal of Environmental Research and Public Health*, 19, 6449. <https://doi.org/10.3390/ijerph19116449>
- Siriyasatien, P., Chadsuthi, S., Jampachaisri, K., & Kesorn, K. (2018). Dengue epidemics prediction: A survey of the state-of-the-art based on data science processes. *IEEE Access*, 6, 53757–53795. <https://doi.org/10.1109/ACCESS.2018.2871241>
- Song-Quan, O. (2016). Dengue vector control in Malaysia: A review for current and alternative strategies. *Sains Malaysiana*, 45(5), 777–785, 2016.
- Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2019). Improving the forecasting performance of temporal hierarchies. *PLoS One*, 14(10), Article e0223422. <https://doi.org/10.1371/journal.pone.0223422>
- Sulaiman, S., Pawanchee, Z. A., Arifin, Z., & Wahab, A. (1996). Relationship between Breteau and House indices and cases of dengue/dengue hemorrhagic fever in Kuala Lumpur, Malaysia. *Journal of the American Mosquito Control Association*, 12(3 Pt 1), 494–496.

- Tham, A. S. (2000). *Issue and challenges in Aedes surveillance and control workshop proceeding behavior intervention in dengue control of Malaysia Centre of Drug Malaysia*. Malaysia: Universiti Sains Malaysia and Institute of Health Promotion, Ministry of Health.
- Tran, B. L., Tseng, W. C., Chen, C. C., & Liao, S. Y. (2020). Estimating the threshold effects of climate on dengue: A case study of taiwan. *International Journal of Environmental Research and Public Health*, 17(4), 1392. <https://doi.org/10.3390/ijerph17041392>
- Usman, P. (2003). *Modeling mathematics for dengue fever disease*. ITB. Bandung.
- Wan Najdah, W. M. A., Rohani, A., Zurainee, M. N., Thuraiya, A. R., & Yvonne, L. A. L. (2021). Spatial distribution of mosquito vector in dengue outbreak areas in Kuala Lumpur and Selangor, Malaysia. *Serangga*, 26(3), 110–139.
- WHO. (2021). *Dengue and severe dengue*. <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>.
- Wilder-Smith, A., Earnest, A., Tan, S. B., Ooi, E. E., & Gubler, D. J. (2010). Lack of association of dengue activity with haze. *Epidemiology and Infection*, 138(7), 962–967. <http://www.jstor.org/stable/40793021>.
- Windyaraini, D. H., Marsifah, T., Mustangin, Y., & Poerwanto, S. H. (2019). Detection of transovarial transmission of dengue virus in *Aedes* spp. (Diptera: Culicidae) from brontokusuman village, yogyakarta, Indonesia. *Biodiversitas Journal of Biological Diversity*, 20(7). <https://doi.org/10.13057/biodiv/d200737>
- Woodruff, R. E., McMichael, T., & Butler, C. (2006). Action on climate change: The health risks of procrastinating. *Australian & New Zealand Journal of Public Health*, 30(6), 567–571.
- World Mosquito Program. (2021). *Mosquito borne diseases*. <https://www.worldmosquitoprogram.org/en/learn/>.
- Zhao, N., Charland, K., Carabali, M., Nsoesie, E. O., Maheu-Giroux, M., Rees, E., et al. (2020). Machine learning and dengue forecasting: Comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia. *PLoS Neglected Tropical Diseases*, 14(9), Article e0008056. <https://doi.org/10.1371/journal.pntd.0008056>