Geospatial Health 2023; volume 18:1176

Prediction of dengue cases using the attention-based long short-term memory (LSTM) approach

Mokhalad A. Majeed,[1] Helmi Z. M. Shafri,[1,2] Aimrun Wayayok,[3] Zed Zulkafli[1]

[1]Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM); [2]Geospatial Information Science Research Centre (GISRC), Faculty of Engineering, Universiti Putra Malaysia (UPM); [3]Department of Biological and Agricultural Engineering, Faculty of Engineering, University Putra Malaysia Serdang, Selangor, Malaysia

Abstract

This research proposes a 'temporal attention' addition for long-short term memory (LSTM) models for dengue prediction. The number of monthly dengue cases was collected for each of five Malaysian states *i.e.* Selangor, Kelantan, Johor, Pulau Pinang, and Melaka from 2011 to 2016. Climatic, demographic, geographic and temporal attributes were used as covariates. The proposed LSTM models with temporal attention was compared with several benchmark models including a linear support vector machine (LSVM), a radial basis function support vector machine (RBFSVM), a decision tree (DT), a shallow neural network (S-ANN) and a deep neural network (D-ANN). In addition, experiments were conducted to analyze the impact of look-back settings on each model performance. The results showed that the attention LSTM (A-LSTM) model performed best, with the stacked, attention LSTM (SA-LSTM) one in second place. The LSTM and stacked LSTM (S-LSTM) models performed almost identically but with the accuracy improved by the attention mechanism was added. Indeed, they were both found to be superior to the benchmark models mentioned above. The best results were obtained when all attributes were included in the model. The four models (LSTM, S-LSTM, A-LSTM and SA-LSTM) were able to accurately predict dengue presence 1-6 months ahead. Our findings provide a more accurate dengue prediction model than previously used, with the prospect of also applying this approach in other geographic areas.

Correspondence: Helmi Z. M. Shafri, Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM), 43400, Selangor, Malaysia.
E-mail: helmi@upm.edu.my

Key words: dengue fever; LSTM; attention, deep learning; Malaysia.

Conflict of interest: the Authors declare no conflict of interest.

Availability of data and materials: all data generated or analyzed during this study are included in this published article.

Received for publication: 28 November 2022.
Accepted for publication: 19 April 2023.

Geospatial Health 2023; 18:1176
doi:10.4081/gh.2023.1176

Introduction

Dengue, a common tropical disease affecting millions of people, is heavily influenced by rainfall, temperature, relative humidity and rapid urbanization. The virus that causes infection is spread by various mosquito species, with the main vector being *Aedes aegypti*. According to the World Health Organization (WHO), dengue infections have increased globally, with 50-100 million new infections occurring each year in more than 70 countries (WHO, 2023). Malaysia experienced an unprecedented outbreak of dengue infections from 2014 to 2016, with a large increase in the number of cases (Suppiah *et al.*, 2018).

Previous research (Appice *et al.*, 2020; Bogado *et al.,* 2020; Ferdousi *et al.*, 2021; Mussumeci and Coelho, 2020; Nayak and Narayan, 2019) used statistical and machine learning models, with varying results for time-series prediction. In general, dengue fever prediction models make use of covariates to account for the number of cases forecast. Climate has both direct and indirect effects on dengue transmission, distribution, vector breeding and establishment (Gubler *et al.*, 2001; McMichael *et al.*, 2006). Temperature and rainfall work interdependently influencing vector dynamics. In addition, temperature influences the relative humidity indirectly by regulating evaporation. These factors affect the availability of breeding sites for the *Aedes* vector. The relationship between rainfall and dengue outbreaks has been shown in previous studies (Kolivras, 2010; Polwiang, 2020). Many factors, such as climatology, demography, socioeconomics, vector ecology and geography have been studied as covariates in dengue prediction systems (Altassan, 2020; Anno *et al*., 2019; Jain *et al.,* 2019, Jayaraj *et al.*, 2019; Nan *et al.*, 2018).

Various statistical and machine learning models have been used for the prediction of dengue fever outbreaks. Linear regression (LR) (Anggraeni *et al.* 2020) and autoregressive integrated moving average (ARIMA) (Promprou *et al.* 2006) are commonly used, but nonlinear regression models and machine-learning models, such as Naive Bayes (Fathima *et al*., 2011), XGBoost (Methiyothin & Ahn 2022), generalized additive models (Baquero *et al.,* 2018), random forests (Zhao *et al.,* 2020), neural networks (Zhao *et al.,* 2020), manifold learning (Souza *et al.*, 2022) and fuzzy logic (Idris *et al.,* 2018) have also been developed and tested. These models use different predictors such as climate data, historical dengue cases, vegetation, rainfall, air temperature, population counts, income inequality and education level. Researchers have demonstrated the accuracy and success of modelling for the prediction of dengue fever outbreaks in different regions, e.g., Cousien *et al.* (2019), Yuan *et al.* (2019), Siregar and Makmur (2019), Jain *et al*. (2019), and Zhao *et al*. (2020).

Deep learning has emerged as a powerful machine learning technique for predicting dengue fever outbreaks. Specifically, the long short-term memory (LSTM) model, a type of recurrent neural network (RNN), has shown promising results in several studies. For instance, Bogado *et al.* (2020), trained LSTM models using synthetic data generated from model predictions showing that LSTM models may provide accurate predictions of dengue cases. However, the authors suggested that additional parameters such as demographic, geographic, and environmental variables should be included to better characterize potential epidemic's behaviour across regions. Similarly, Xu *et al.* (2020) developed a LSTM-based dengue fever prediction model using monthly dengue cases and climate data and showed that the LSTM model outperformed previously published forecast models, while Mussumeci and Coelho (2020) compared different machine-learning models, including LSTM, for predicting weekly dengue incidence in 790 Brazilian cities revealing that LSTM outperformed other models, such as the least absolute shrinkage and selection operator (LASSO) and random forest. However, Xu *et al.* (2019) developed an LSTM-based dengue fever prediction model for 20 important Chinese cities using dengue cases and local meteorological data from 2005 to 2018 that showed that the LSTM model with only local data performed worse than the use of susceptible infected recovered (SIR) data.

Transfer learning techniques can be used to improve the LSTM model for monthly dengue case prediction. Although deep learning, particularly the LSTM model, has shown promise in predicting dengue fever outbreaks, incorporating additional parameters such as demographic, geographic, and environmental variables could further improve the accuracy of these models. Importantly, the underlying processes of dengue fever occurrences are neither purely linear nor purely nonlinear in nature. They typically include both linear and nonlinear patterns. As a result, a robust prediction model is needed to accurately model such complex structures. Furthermore, predicting dengue cases is difficult due to the multi-faceted interplay of epidemiological and environmental determinants. This intricate, causal scenario manifests itself in the variation of incidence patterns across geographic areas. The frequent absence of long-term historical records of disease incidence, as well as environmental risk factors, further complicates statistical analysis. Existing models for dengue prediction based on time-series are still limited. Although most of the models (both statistical and deep learning ones) are 'user friendly' and have good interpretability by limited data and low training cost, they lack the ability to process long sequences (Mussumeci & Coelho, 2020; Men *et al.*, 2021). While no single solution exists to address all of these issues, deep learning models have recently proven effective for computer vision tasks (convolutional neural networks) and time-series data (such as RNN). However, deep learning models with standard architectures generally fail to remember information in long time-series data. Deep learning with attention, according to studies (Lim & Zohren 2021), can provide a compelling case for time-series prediction. Attention is a method of selecting variables that are useful for forecasting. The aim of attention is to select useful information from among the various feature time-series data in order to predict the target time-series.

This research aimed to design, develop and validate a dengue prediction model based on LSTM with a novel attention mechanism. The specific objectives were to: i) develop a model based on LSTM for the prediction of monthly dengue cases; ii) design an attention mechanism for the LSTM that can boost the performance of dengue prediction; iii) validate the models and evaluate their predictive capability for 1-6 months ahead; and iv) evaluate the performance of the prediction models with different attribute groups *i.e.* climatic, geographic and temporal ones.

Materials and Methods

This research developed a temporal attention mechanism for LSTM models to predict dengue cases in five Malaysian states. The attention mechanism was tested using a single LSTM layer as well as stacked LSTM layers. The models were compared to other benchmark models such as LSVM, RBFSVM, DT, S-ANN, and D-ANN. The effects of attribute selection were also tested for the proposed models in the sensitivity analysis. Various look-back values were used to demonstrate the models' ability to predict dengue cases several months in advance and the findings validated the proposed models' ability to predict dengue cases.

Study area

Malaysia is a Southeast Asian country located between 1° to 7° N and 99° to 105° E. It has a total area of 131,587 km^2 and is divided into two regions by the South China Sea: Peninsular Malaysia and East Malaysia (Sabah, Sarawak and Labuan or Malaysian Borneo). Figure 1 shows the study area that included the five Malaysian states Selangor, Kelantan, Johor, Pulau Pinang and Melaka.

Malaysia's geographical location makes it prone to tropical diseases. Its floodplains, hills and coastline zones have a humid tropical environment with temperatures ranging from 21°C to 32°C. The two rainy seasons are caused by the Northeast Monsoon (NEM) from October to March (Moten *et al.*, 2014) and the Southwest Monsoon (SWM) from May to September (Diong *et al.,* 2015). The month of April is a transitional time with significant rains (Wong *et al*., 2009). Dengue fever is more common in Selangor, Kelantan, Johor, Pulau Pinang and Melaka than in other states. According to the Malaysian Ministry of Health (MoH), Selangor is responsible for 90% of the national number of dengue cases (MoH, 2015).

The first cases of dengue fever in the country were reported in 1902. In the 1970s, dengue became a public health concern with the first major outbreak in 1973 (Mohd-Zaki *et al.,* 2014; Shepard *et al.,* 2013). Between 2000 and 2014, the number of cases per 100,000 increased from 32 to 361. The majority of dengue patients are between the ages of 15 and 49 and 80 % occur in urban areas (MoH, 2015, Figure 1).

Data sources

Dengue cases

The MoH, Malaysia is the original source of the dengue data but the weekly, state-level data from January 2010 to December 2016 were acquired from the Malaysia Open Data (http://www.data.gov.my) for the five states under study (Figure 2). We focused on the number of deaths due to dengue. The total mortality within each state, given by week, were aggregated into monthly numbers.

Meteorological and geographic attributes

In this research, the explanatory variables were extracted from three sources: climate data (rainfall and land surface temperature), demographic data (population density) and geographic data (digital elevation model (DEM), vegetation index, road network, water bodies and type of land cover) (Table 1). These variables have been used in various dengue prediction studies around the world and shown to have a positive or negative impact on the number of dengue cases within a geographic region.

Data pre-processing

The data were acquired in different formats and needed preparation that included transformation from one format to another, filtering and cleaning, pre-processing for tabular, raster and vector datasets. The preparation also included data registration to match the datasets geographically and temporally.

Prediction of the number of dengue cases

LSTM

Traditional feed-forward neural networks are effective with regard to learning (Lim & Zohren, 2021). However, they are less useful for modelling. RNNs were thus built with feedback connections to explicitly include model sequences (Hochreiter and Schmidhuber, 1997) and are indeed more biologically plausible than feed-forward neural networks. The advantage of using feedback connections in RNNs is that they provide the model with a memory of previous activations. As a result, such models can learn the temporal dynamics of sequential data, but they continue to have drawbacks, such as vanishing or exploding gradients. To address this problem, Hochreiter and Schmidhuber (1997) proposed LSTM. Memory blocks, which consist of self-connected memory cells and three multiplicative units replace hidden units in LSTM (input, output and forget gates). The gates allow reading writing and resetting operations in the memory block and control its behaviour. Figure 3 depicts a diagram of a single LSTM unit.

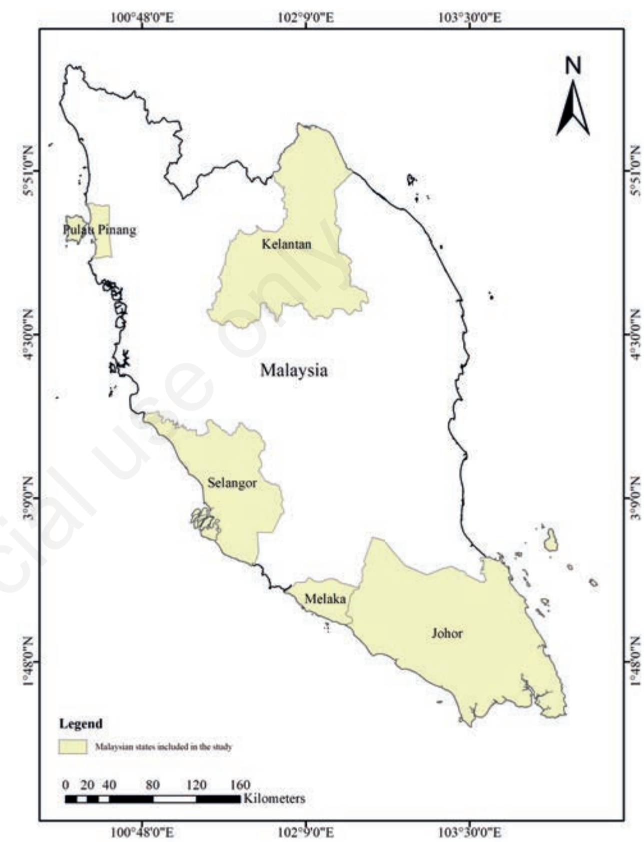

Figure 1. Geographic location of the study area in Peninsular Malaysia.

Table 1. Data sources and details used to extract attributes for the dengue prediction models.

Variable	Data	Period	Spatial resolution	Temporal resolution	Source
Dependent	Dengue (no. cases)	2011-2016	State-level	Monthly	Malaysia Open Data (http://www.data.gov.my)
Climate	Rainfall (mm)	2011-2016	0.25 degree	Daily	CMORPH Climate Data Record*
	LST (°C)	2011-2016	0.05 degree	Monthly	MODIS (https://www.modis.com)
Geographic	Vegetation index	2011-2016	0.05 degree	16-days	MODIS (https://www.modis.com)
	DEM	na	30 arc-seconds	na	Diva GIS** (CGIAR SRTM)
	Land cover	na	30 arc-seconds	na	Diva GIS** (GLC2000)
	Road networks	na	na	na	Diva GIS** (Digital Chart of the World)
	Water bodies	na	na	na	Diva GIS** (Digital Chart of the World)
Demographic	Population (no.)	2011-2016	State-level	Yearly	Malaysia Open Data (http://www.data.gov.my)

LST, land surface temperature; DEM, digital elevation model; na, not applicable; *Satellite-generated precipitation data from the U.S. National Oceanic and Atmospheric Administration (NOAA) (https://www.ncei.noaa.gov/products/climate-data-records/precipitation-cmorph); **Diva GIS (https://diva-gis.org/).

Let c_t be the sum of inputs at time step t and its previous time step activations, the LSTM updates for time step i given inputs x_t, h_t-1 and c_t-1 (Donahue *et al.*, 2015):

$$i_t = \sigma(W_{xi.}\ x_t + W_{hi.}\ h_{t-1} + Wci.\ c_{t-1} + b_i) \quad \text{(Eq. 1)}$$

$$f_t = \sigma(W_{xf.}\ x_t + W_{hf.}\ h_{t-1} + Wcf.\ c_{t-1} + b_f) \quad \text{(Eq. 2)}$$

$$c_t = i_{t.}\ tanh(W_{xc.}\ x_t + W_{hc.}\ h_{t-1} + b_c) + f_{t.}\ c_{t-1} \quad \text{(Eq. 3)}$$

$$o_t = \sigma(W_{xo.}\ x_t + W_{ho.}\ h_{t-1} + W_{co.}\ c_t + b_o) \quad \text{(Eq. 4)}$$

$$h_t = o_{t.}\ tanh(c_t) \quad \text{(Eq. 5)}$$

where i signifies the input gate; f the forget gate; c the sum of inputs; o the output gate; σ is an element-wised nonlinearity such as a sigmoid function; W the weight matrix; b_i the input bias vector; x_t the input (all attributes) at time step t; h_t-1 the hidden state vector of the previous time step (target variable - dengue cases); and *tanh* the activation function based on the hyperbolic tangent function.

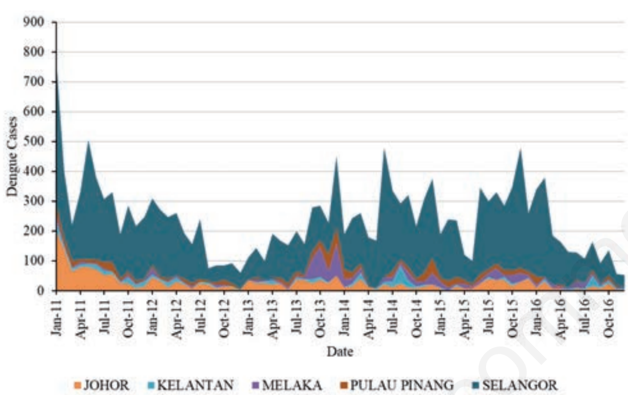

Figure 2. Number of dengue cases observed in five Malaysian states studied during January 2011 and December 2016.

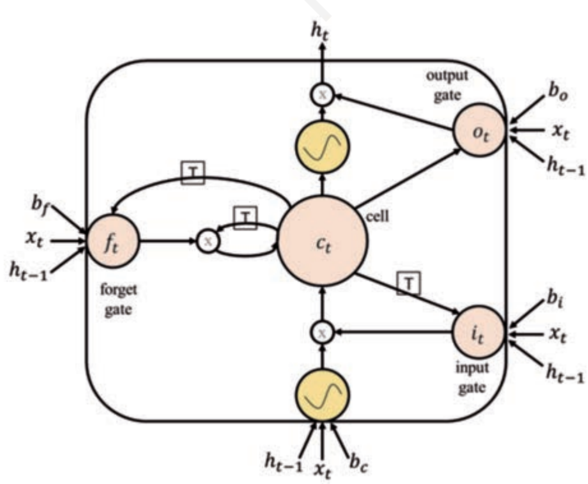

Figure 3. Basic single-cell LSTM.

In these equations, x_t is the input at time t; h_{t-1} the hidden state at time t-1; and c_{t-1} the cell state at time t-1. The LSTM cell has several gates, which are responsible for controlling the flow of information. The input gate i_t controls how much of the input is let into the cell state; the forget gate f_t controls how much of the previous cell state is retained; output gate o_t controls how much of the cell state is used to compute the output; and the cell state c_t is updated based on the input and previous cell state. Finally, the hidden state h_t is computed using the updated cell state and the output gate.

The 'attention' mechanism

Attention is a relatively new approach to improve deep learning models for better modelling of long-term dependencies. Attention mechanisms allow for a more direct dependence between the state of the model at different points in time (Raffel & Ellis, 2015). Figure 4 illustrates the concept of the temporal attention used in this research. Let a model produces a hidden state $\boldsymbol{h_t}$ at each time step, attention models compute a "context" vector $\boldsymbol{c_t}$ as the weighted mean of the state sequence $\boldsymbol{h}$ by:

$$c_t = \sum_{j=1}^{T} \alpha_{tj} h_j \quad \text{(Eq. 6)}$$

where $\boldsymbol{T}$ is the total number of time steps in the input sequence and $\boldsymbol{a_{tj}}$ a weight computed at each time step $\boldsymbol{t}$ for each state $\boldsymbol{h_j}$. These context vectors are then used to compute a new state sequence $\boldsymbol{s}$, where $\boldsymbol{s_t}$ depends on $\boldsymbol{s_{t-1}}$, c_t and the model's output at $\boldsymbol{t}$**-1**. The weightings a_{tj} are then computed by:

$$e_{tj} = a(s_{t-1}, h_j), \alpha_{tj} = \frac{\exp(e_{tj})}{\sum_{k=1}^{T} \exp(e_{tk})} \quad \text{(Eq. 7)}$$

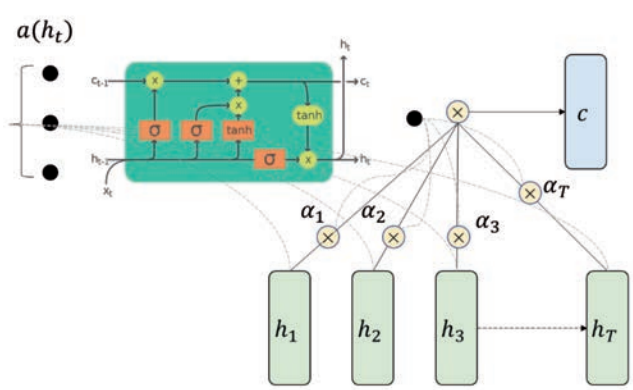

Figure 4. The proposed temporal attention: the basic working concept. The module represented by the green box is an LSTM cell; x is the multiplication operator, x_t the input vector; c_t the cell state vector at time t, b_t hidden state vector at time t. c_{t-1} the cell state vector at time t-1, h_{t-1} the hidden state vector at time t-1; tanh the activation function based on the hyperbolic tangent function; σ the LSTM gates including input, forget and output gates; α_t the attention-focused hidden state representation at time t, T the number of lookback data points; and C the computed attention-based cell state vector.

where $\boldsymbol{a}$ is a learned function which can be thought of as computing a scalar importance value for $\boldsymbol{h_j}$ given the value of $\boldsymbol{h_j}$ and the previous state $\boldsymbol{s_{t-1}}$. This formulation allows the new state sequence $\boldsymbol{s}$ to have more direct access to the entire state sequence $\boldsymbol{h}$. Attention-based RNNs have proven effective in a variety of sequence transduction tasks, including disease prediction (Men *et al*., 2021), flood forecasting (Ding *et al*., 2020), traffic flow prediction (Zheng *et al*., 2020) and time series prediction (Li *et al.,* 2019).

Proposed LSTM models

In this study, four LSTM models were developed and tested for dengue fever forecasting including one-layer LSTM, stacked LSTM (S-LSTM), attention-based LSTM (A-LSTM) and stacked, attention-based LSTM (SA-LSTM). The one-layer LSTM lacks the temporal attention module as does S-LSTM that is constructed by stacking two LSTM layers in a single model. A-LSTM and SA-LSTM models, on the other hand, are constructed by including a temporal attention mechanism in each of a one-layer LSTM model and a two-layer S-LSTM model. Naturally, stacking multiple LSTM layers in an attempt to improve the model's capability of learning hierarchical feature representation increases its complexity. However, the inclusion of a temporal attention module in a LSTM model makes it model the temporal structures of the input data more effectively.

Benchmark models

A number of machine learning models including linear support vector machine (LSVM), radial basis function support vector machine (RBFSVM), Decision Tree (DT), Shallow artificial neuronal network (S-ANN) and deep artificial neuronal network (D-ANN) were used for dengue prediction. The models were implemented in Python (https://www.python.org/) using the scikit-learn package (Kramer, 2016). Hyperparameters were optimized using the grid search method to avoid overfitting and improve the predictive performance as shown in Table 2.

Validation

The standard root mean square error (RMSE) was used to measure the performance and prediction accuracy of the proposed models. RMSE is a well-known measure used to evaluate continuous variables by measuring the differences between predicted and observed values. RMSE is estimated as follows,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_t - \underline{Y}_t\right)^2} \quad \text{(Eq. 8)}$$

where $\boldsymbol{Y_t}$ is the dengue cases of observation for time $\boldsymbol{t}$, and $\boldsymbol{Y_t}$ the number of cases predicted by the model at time $\boldsymbol{t}$. Smaller RMSE values indicate a smaller difference between the predicted and observed values and indicates a higher prediction performance**.**

The training and test splits were based on 75% and 25% of all the available samples, respectively. For example, the available samples were for one year time period and the data temporal resolution were monthly. The data split would be the first nine months for training the model, with the remaining three months kept for testing the model.

Results

LSTM performance

Four LSTM model variants were tested for the prediction of monthly dengue cases in five Malaysian states. Table 3 summarizes their average performance. To train and test the models in this experiment, entire samples from all five states were used. The results are shown with various look-back values ranging from one to six months, which indicate that A-LSTM optimally, with SA-LSTM second best. LSTM and S-LSTM, performed almost identi-

Table 2. Configuration of benchmark model hyperparameters.

Model	Hyperparameter	Description	Optimal value*
LSVM	C	Determines how much misclassification error is allowed in the SVM training process.	10
RBFSVM	C	Determines how much misclassification error is allowed in the SVM training process.	10
	Gamma	Controls the shape of the kernel function, which affects how the decision boundary is formed.	Scale
DT	Best split criterion	Criterion in decision trees typically chosen based on the measure of impurity or information gain.	MSE
	Splitter	Algorithm that determines how to split a node into child nodes based on a selected split criterion.	Best
	Maximum tree depth	A parameter that controls the maximum depth of the tree.	4
S-ANN	Hidden layer sizes	Refer to the number of neurons or nodes in each hidden layer of the neural network.	(128)
	Activation function	A mathematical function applied to the output of each neuron in a neural network, which determines whether a neuron is activated or not based on the weighted sum of inputs.	ReLU
	Optimizer	Algorithm used to update the weights of a neural network during the training process.	Adam
	Learning rate	Determines the step size at which the optimizer adjusts the weights of the network during the training process.	0.001
	Epochs	Refers to a single pass through the entire training dataset during the training process.	200
D-ANN	Hidden layer sizes	Refer to the number of neurons or nodes in each hidden layer of the neural network.	(128, 64, 32)
	Activation function	A mathematical function applied to the output of each neuron in a neural network, which determines whether a neuron is activated or not based on the weighted sum of inputs.	ReLU
	Optimizer	Algorithm used to update the weights of a neural network during the training process.	Adam
	Learning rate	Determines the step size at which the optimizer adjusts the weights of the network during the training process.	0.001
	Epochs	Refers to a single pass through the entire training dataset during the training process.	200

LVSM, linear support vector machine; RBFSVM, radial basis function support vector machine; DT, decision tree; ANN, artificial neuronal network; S-, Shallow; D-, Deep; MSE, mean square error; C, penalty parameter; ReLU, rectified linear unit; Adam, adaptive moment estimation. *by grid search.

cally. The findings show that adding the attention module to LSTM enhances its accuracy.

Comparison of LSTM and benchmark models

For the purpose of evaluating the proposed LSTM models, the five benchmark forecast models were used. Table 4 summarizes their average performance for the five Malaysian states. As can be seen in the figure, RBFSVM performed best. The D-ANN, on the other hand, outperformed both S-ANN and the DT. The fact that the RBFSVM outperformed the LSVM suggests that the dengue data were nonlinear. Figure 5 shows the comparison between the LSTM models and the benchmark ones.

Model selection and analysis

Impact on LSTM by attribute selection

For the prediction of the number of dengue cases, this study examined three types of attributes: climatic, geographic and temporal. Rainfall and land surface temperature were included in the

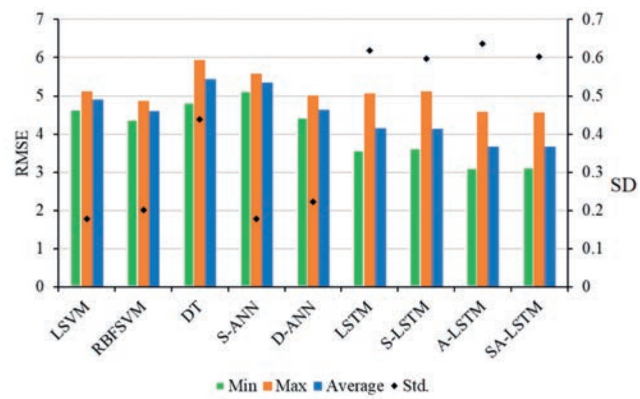

Figure 5. Comparison of LSTM models and benchmark models. LVSM, linear; RBFSVM, radial basis function support vector machine; DT, decision tree; S-ANN, shallow artificial neuronal network; D-, Deep artificial neuronal network; LSTM, long-short term memory; S-, stacked; A-, attention; SA-, stacked and attention; SD, standard deviation.

Table 3. Summary results of LSTM model performance with different look-back values based on test samples.

Look-back	Performance (RMSE)			
(no. of months)	LSTM	S-LSTM	A-LSTM	SA-LSTM
1	3.75	3.72	3.27	3.26
2	3.57	3.62	3.10	3.12
3	5.06	5.11	4.58	4.55
4	3.74	3.85	3.23	3.42
5	3.99	3.85	3.42	3.36
6	4.78	4.62	4.34	4.31
Minimum	3.57	3.62	3.10	3.12
Maximum	5.06	5.11	4.58	4.55
Average	4.15	4.13	3.66	3.67
SD	0.61	0.59	0.63	0.60

RMSE, standard root mean square error; LSTM, long-short term memory; S-, Stacked; A-, Attention; SA-, Stacked and Attention; SD, standard deviation.

Table 4. Summary results of benchmark model performance with different look-back values.

Look-back	Performance (RMSE)				
(no. of months)	LSVM	RBFSVM	DT	S-ANN	D-ANN
1	4.98	4.44	4.82	5.47	4.66
2	4.91	4.37	5.58	5.32	4.42
3	5.04	4.76	5.60	5.43	4.53
4	4.64	4.44	5.93	5.13	4.42
5	4.76	4.67	5.69	5.17	4.76
6	5.11	4.86	4.96	5.58	5.00
Minimum	4.64	4.37	4.82	5.13	4.42
Maximum	5.11	4.86	5.93	5.58	5.00
Average	4.91	4.59	5.43	5.35	4.63
SD	0.17	0.20	0.43	0.17	0.22

RMSE, standard root mean square error; VSM, support vector machine; LVSM, Linear VSM; RBFSVM, radial basis function VSM; DT, decision tree; ANN, artificial neuronal network; S-, shallow; D-, deep; SD, standard deviation.

climatic category. Vegetation index, elevation, land cover, road networks and water bodies were all part of the geography group. Population density was also included in this group, while the lag time of the attributes indicated was included in the temporal group. Table 5 compares the performance of the LSTM models with respect to the various groups of attributes. The findings show that almost all of the models performed well across all attribute groups. All of the 1–6-month look-backs yielded similar results. In addition to the climatic attributes, the results show that the models perform better with more attributes. Furthermore, the findings reveal that the geographic variables are more important the for dengue prediction than the temporal attributes (year and month).

Impact of attribute selection on the benchmark models

Benchmark models were used to examine the importance of several attribute groupings for predicting dengue cases (Table 6). When the climatic and geographic data were merged, both the LSVM and RBFSVM models performed well. With all of the attributes combined DT, S-ANN and D-ANN performed best. LSVM worked best when all attributes were paired with some look-back values.

Modelling performance with regard to look-back settings

The graphs in Figure 6 show the performance of the proposed LSTM and benchmark models for predicting dengue cases with various look-back settings. With 1, 2 and 4 months of look-back values, the LSTM models performed best for predicting dengue cases. With 6 and 3 months of look-back data, the LSTM models performed worst. With a 1-month look-back, DT performed best, whereas the other benchmark models (LSVM, RBFSVM, S-ANN, and D-ANN) performed best with a 4-month look-back. With a 6-

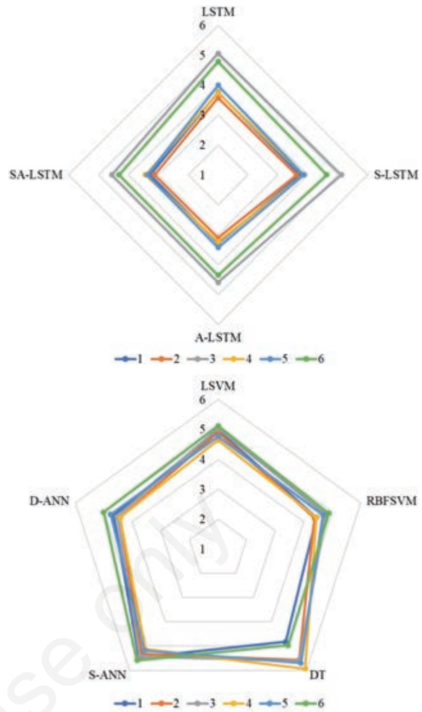

Figure 6. Radar charts of prediction performance of the proposed and benchmark models with different look-back values. LVSM, Linear support vector machine; RBFSVM, radial basis function support vector machine; DT, Decision Tree; S-ANN, Shallow artificial neuronal network; D-ANN, Deep artificial neuronal network; LSTM, long-short term memory; S-, Stacked; A-, Attention; SA-, Stacked and Attention.

Table 5. Summary results of LSTM model performance with different attribute selection and look-back values.

Look-back (no. of months)	Group attributes added	Performance (RMSE)			
		LSTM	S-LSTM	A-LSTM	SA-LSTM
1	Climate	5.478	5.721	5.589	6.766
	Climate/time	3.849	3.919	3.950	4.970
	Climate/geography	3.841	3.769	6.000	3.993
	Climate/time/geography	3.759	3.729	3.278	3.266
2	Climate	4.504	4.023	4.129	3.593
	Climate/time	5.010	4.580	4.443	4.001
	Climate/geography	4.176	4.364	4.047	3.478
	Climate/time/geography	3.575	3.620	3.100	3.124
3	Climate	5.238	5.914	5.322	5.987
	Climate/time	5.361	6.311	5.584	6.031
	Climate/geography	5.304	6.051	4.975	5.040
	Climate/time/geography	5.061	5.113	4.589	4.557
4	Climate	4.237	4.280	3.474	3.008
	Climate/time	4.546	4.878	3.857	3.736
	Climate/geography	4.328	4.574	3.414	3.495
	Climate/time/geography	3.748	3.856	3.237	3.421
5	Climate	6.684	5.646	5.385	5.971
	Climate/time	5.895	5.357	4.527	4.743
	Climate/geography	4.942	4.835	4.277	4.354
	Climate/time/geography	3.991	3.853	3.423	3.364
6	Climate	6.981	7.435	6.354	6.986
	Climate/time	5.879	5.598	5.141	6.207
	Climate/geography	5.044	5.016	4.437	5.251
	Climate/time/geography	4.789	4.629	4.349	4.313

RMSE, standard root mean square error; LSTM, long-short term memory; S-, stacked; A-, attention; SA-, stacked and attention; SD, standard deviation.

month look-back, non-LSTM models performed the worst. Thus long look-back values are important for a full understanding of the capabilities of the different models. The detailed results of this experiment are summarized in the supplementary data (Tables S1 and S2).

Modelling performance in the different states

With regard to the Malaysian states Selangor, Kelantan, Johor, Pulau Pinang and Melaka, the LSTM models performed best in Kelantan and Pulau Pinang (Figures 7 and 8). They performed moderately in Johor and Melaka, but did not do well in Selangor.

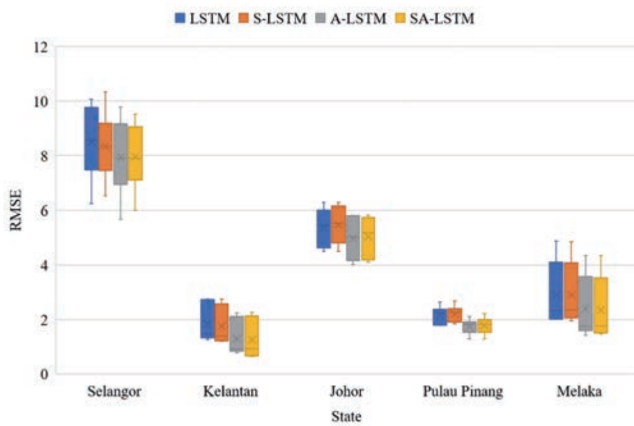

Figure 7. Comparison of LSTM model performance in different Malaysian states.

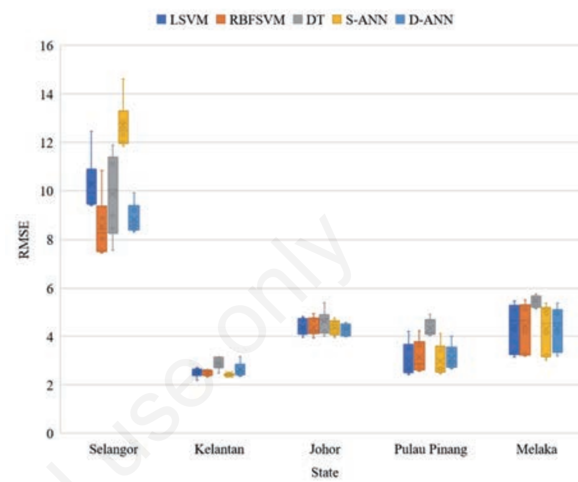

Figure 8. Performance comparison of the benchmark models in different Malaysian states. Each box represents the various look-back experiments.

Table 6. Summary results of benchmark models' performance with different attribute selection and look-back values.

Look-back (no. of months)	Group attributes added	Performance (RMSE)				
		LSVM	RBFSVM	DT	S-ANN	D-ANN
1	Climatic	5.057	4.544	5.297	5.901	4.933
	Climatic/temporal	5.431	4.708	4.854	5.801	4.900
	Climatic/geographic	4.979	4.359	5.294	5.593	4.766
	Climatic/temporal/geographic	4.984	4.448	4.824	5.474	4.661
2	Climatic	5.655	4.427	6.294	6.070	4.704
	Climatic/temporal	5.625	4.697	5.601	5.910	4.649
	Climatic/geographic	4.949	4.292	6.291	5.416	4.530
	Climatic/temporal/geographic	4.911	4.375	5.581	5.324	4.427
3	Climatic	5.938	4.855	5.910	6.569	4.902
	Climatic/temporal	5.928	5.113	5.623	6.405	4.801
	Climatic/geographic	5.049	4.657	5.883	5.599	4.741
	Climatic/temporal/geographic	5.041	4.767	5.603	5.432	4.536
4	Climatic	5.644	4.508	5.965	6.256	4.638
	Climatic/temporal	5.579	4.739	5.964	5.931	4.488
	Climatic/geographic	4.648	4.349	5.944	5.197	4.508
	Climatic/temporal/geographic	4.643	4.449	5.934	5.131	4.429
5	Climatic	5.538	4.715	5.712	6.248	4.946
	Climatic/temporal	5.492	4.946	5.709	6.059	4.747
	Climatic/geographic	4.755	4.612	5.699	5.199	4.798
	Climatic/temporal/geographic	4.763	4.674	5.699	5.177	4.762
6	Climatic	5.943	4.940	4.971	6.538	5.287
	Climatic/temporal	5.929	5.205	4.963	6.402	5.122
	Climatic/geographic	5.134	4.837	4.963	5.599	5.126
	Climatic/temporal/geographic	5.119	4.869	4.963	5.587	5.003

RMSE, standard root mean square error; LSTM, long-short term memory; S-, Stacked; A-, Attention; SA-, Stacked and Attention; SD, standard deviation.

Discussion

LSTM is one of the most commonly used models for time series prediction due to its capability in modelling long dependencies (Hochreiter & Schmidhuber, 1997). LSTM and variants show a considerable amount of success in predicting sequential data and have frequently outperformed other classical and machine learning methods (Hua *et al.,* 2019; Laptev *et al.,* 2017; Zhu & Laptev, 2017). LSTM-based models can identify periodic patterns spanning multiple time steps in non-linear time-series (Xu *et al.,* 2020). Moreover, Anno *et al.* (2019) showed that while it is possible to predict dengue outbreaks with a climate-based model, it is necessary to use models with memory capability like convolutional LSTM (ConvLSTM) to achieve accurate prediction of outbreak distributions needed to develop a proper early warning system.

In this study, a temporal attention mechanism was developed for LSTM models, with the effectiveness evaluated using both single and stacked LSTM layers. The resulting models were then compared to several benchmark models, including LSVM, RBFSVM, DT, S-ANN and D-ANN. According to the findings of this study, the attention module enhances model memory for long dependencies and improves the prediction accuracy of both LSTM and S-LSTM. However, the results also show that the LSTM model benefitted more from the attention module than the S-LSTM as the ability of the former to model complex dengue data was typically accompanied by overfitting due to the larger number of model parameters. However, our study revealed that this problem can be overcome utilizing the attention mechanism and allowing the model to focus on critical sequences of the input data.

The complexity of dengue dynamics challenges the development of predicting models. Previous studies have confirmed that a series of socioeconomic, environmental and climatic factors are closely related with dengue transmission by either facilitating virus amplification or favouring vector survival (Zhu *et al.*, 2018). Human movement was recognized as an important driver of transmission dynamics, which can introduce dengue viruses into previously low-transmission or dengue-free areas (Zhu *et al.*, 2018). This research used the most common attributes applied in previous prediction studies. In addition, an experiment was designed and tested to evaluate the performance of the prediction models with different attribute groups, *i.e.* climatic, geographic and temporal ones. Data availability and type of model were both found to be important while designing the attributes in dengue case prediction research. It is more logical to use and analyse additional attributes if such data are available. On the other hand, models perform differently according to the complexity of the input attributes. Some models can perform well with a larger number of attributes, while others cannot.

The models proposed in this work were based on LSTM as this kind of deep learning model is designed for dealing with sequential data and capable of learning spatial-temporal features automatically. These models suppress less useful features by their internal structure and discover feature trends in massive data scenarios that can be exploited to empower the accuracy of the considered prediction task (Appice *et al.*, 2020). However, the models constructed in this study only utilised temporal attention mechanisms. In contemporary models created for other prediction tasks, spatial attention is also an essential module. The model's spatial attention allows it to focus on important parts of images or array-ordered data. This helps the model to produce more accurate predictions by improving the features derived from the raw input. Thus, future research dengue fever prediction should focus on both temporal and spatial considerations in order to develop the most accurate model. In addition, future research can focus on developing models that can account for the temporal order of outbreaks and identify leading or lagging regions. These models can incorporate long-term dependencies in the data and use climate variables to predict the occurrence of dengue fever on large space-time scales.

Conclusions

Both A-LSTM and SA-LSTM were found to be superior to the benchmark models. Thus, the attention mechanism assisted the LSTM models in improving their ability to learn long dependencies and, as a result, their accuracy in predicting dengue cases. Furthermore, it was discovered that the inclusion all of attributes tested (climatic, geographic and temporal) was required for achieving the best results. The proposed models could be used to predict dengue cases on a large scale and aid the identification of high-risk areas for dengue viruses.

References

Altassan KK. 2020. Environmental and population factors influencing dengue fever emergence and spread in Saudi Arabia (Doctoral dissertation).

Anggraeni W, Sumpeno S, Yuniarno EM, Rachmadi RF, Gumelar AB, Purnomo MH, 2020. Prediction of dengue fever outbreak based on climate factors using fuzzy-logistic regression. In 2020 International Seminar on Intelligent Technology and Its Applications (ISITIA), pp. 199-204.

Anno S, Hara T, Kai H, Lee MA, Chang Y, Oyoshi K, Tadono T, 2019. Spatiotemporal dengue fever hotspots associated with climatic factors in Taiwan including outbreak predictions based on machine-learning. Geospatial Health 14:771.

Appice A Gel YR, Iliev I, Lyubchich V & Malerba D. (2020). A multi-stage machine learning approach toto predict dengue incidence: a case study in Mexico. IEEE Access 8:52713-52725.

Baquero OS, Santana LMR, Chiaravalloti-Neto F, 2018. Dengue forecasting in São Paulo City with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. PloS One 13:e0195065.

Bogado JV, Stalder D, Gómez S, Schaerer C, 2020. Deep learning-based dengue cases forecasting with synthetic data. Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, 7(1).

Cousien A, Ledien J, Souv K, Leang R, Huy R, Fontenille D, Cauchemez S, 2019. Predicting dengue outbreaks in Cambodia. Emerging Infect Dis 25:2281.

Ding Y, Zhu Y, Feng J, Zhang P, Cheng Z, 2020. Interpretable spatio-temporal attention LSTM model for flood forecasting. Neurocomputing 403:348-359.

Diong JY, Yip WS, MatAdam MK, Chang NK, Yunus F, Abdullah MH, 2015. The definitions of the southwest monsoon climatological onset and withdrawal over Malaysian region. Malaysian Meteorology Dep 3:1-30.

Donahue J, Anne HL, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, Darrell T, 2015. Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE Conference Onon Computer Vision Andand Pattern Recognition (pp. 2625-2634).

Fathima S, Hundewale N, 2011. Comparison of classification techniques-SVM and naives bayes to predict the arboviral disease-dengue. In 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW), pp. 538-9.

Ferdousi T, Cohnstaedt LW, Scoglio CM, 2021. A windowed correlation based feature selection method toto improve time series prediction of dengue fever cases. arXiv preprint arXiv:2104.10289.

Gubler DJ, Reiter P, Ebi KL, Yap W, Nasci R, Patz JA, 2001. Climate variability and change in the United States: potential impacts on vector-and rodent-borne diseases. Environ Health Perspect 109;223-33.

Hochreiter S, Schmidhuber J, 1997. Long short-term memory. Neural Computation 9:1735-1780.

Hua Y, Zhao Z, Li R, Chen X, Liu Z, Zhang H, 2019. Deep learning with long short-term memory for time series prediction. IEEE Comm Mag 57:114-9.

Idris MFIM, Abdullah A, Fauzi SSM, 2018. Prediction of dengue outbreak in Selangor using fuzzy logic. In Proceedings of the Second International Conference on the Future of ASEAN (ICoFA) 2017–Volume 2: Science and Technology (pp. 593-603). Springer Singapore.

Jain R, Sontisirikit S, Iamsirithaworn S, Prendinger H, 2019. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. BMC Infect Dis 19:272.

Jayaraj VJ, Avoi R, Gopalakrishnan N, Raja DB, Umasa Y, 2019. Developing a dengue prediction model based on climate in Tawau, Malaysia. Acta Tropica 197:105055.

Kolivras KN, 2010. Changes in dengue risk potential in Hawaii, USA, due to climate variability and change. Climate Res 42:1-11.

Kramer O, 2016. Scikit-learn. pp 45-53. In Machine learning for evolution strategies. Studies in Big Data, Vol. 20; Springer.

Laptev N, Yosinski J, Li LE, Smyl S, 2017. Time-series extreme event forecasting with neural networks at uber. In International Conference On Machine Learning 34:1-5.

Li Y, Zhu Z, Kong D, Han H, Zhao Y, 2019. EA-LSTM: Evolutionary attention-based LSTM for time series prediction. Knowledge-Based Systems 181:104785.

Lim B, Zohren S, 2021. Time-series forecasting with deep learning: a survey. Philosoph Trans Royal Soc A 379:20200209.

McMichael AJ, Woodruff RE, Hales S, 2006. Climate change and human health: present and future risks. Lancet 367:859-69.

Men L, Ilk N, Tang X, Liu Y, 2021. Multi-disease prediction using LSTM recurrent neural networks. Expert Systems Appl 177:114905.

Methiyothin T, Ahn I, 2022. Forecasting dengue fever in France and Thailand using XGBoost. In 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC); pp. 677-680.

Malaysia Ministry of Health Technology Assessment Section. Clinical practice guidelines, 2015.

Mohd-Zaki AH, Brett J, Ismail E, L'Azou M, 2014. Epidemiology of dengue disease in Malaysia (2000–2012): A systematic literature review. PLoS Neglected Trop Dis 8:e3159.

Moten S, Yunus F, Ariffin M, Burham N, Yik DJ, Adam MKM, Sang YW, 2014. Statistics of northeast monsoon onset, withdrawal and cold surges in Malaysia. Guidelines No. 1.

Mussumeci E, Coelho FC, 2020. Machine-learning forecasting for dengue epidemics-comparing LSTM, random forest and lasso regression. medRxiv.

Nan J, Liao X, Chen J, Chen X, Chen J, Dong G, Hu G, 2018. Using climate factors to predict the outbreak of dengue fever. In 2018 7th International Conference on Digital Home (ICDH) (pp. 213-218).

Nayak MSDP, Narayan KA, 2019. Forecasting dengue fever incidence using ARIMA analysis. Int J Collab Res Internal Medicine Public Health 11;924-932.

Polwiang S, 2020. The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003-2017). BMC Infect Dis 20:1-10.

Promprou S, Jaroensutasinee M, Jaroensutasinee K, 2006. Forecasting dengue haemorrhagic fever cases in southern Thailand using ARIMA models. https://apps.who.int/iris/handle/10665/170355

Raffel C, Ellis DP, 2015. Feed-forward networks with attention can solve some long-term memory problems. arXiv preprint arXiv:1512.08756.

Shepard DS, Undurraga EA, Halasa YA, 2013. Economic and disease burden of dengue in southeast Asia. PloS Neglected Trop Dis 7:e2055.

Siregar FA, Makmur T, 2019. Time series analysis of dengue hemorrhagic fever cases and climate: a model for dengue prediction. J Physics 1235:012072.

Souza C, Maia P, Stolerman LM, Rolla V, Velho L, 2022. Predicting dengue outbreaks in Brazil with manifold learning on climate data. Expert Systems Applications 192:116324.

Suppiah J, Ching S, Amin-Nordin S, Low G, 2018. Clinical manifestations of dengue in relation to dengue serotype and genotype in Malaysia: A retrospective observational study. PLoS Neglected Trop Dis 12:17.

WHO, 2023, Dengue and severe dengue, fact sheet of 17 March 2023 Accessed on 26 March 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

Wong CL, Venneker R, Uhlenbrook S, Jamil ABM, Zhou Y, 2009. Variability of rainfall in peninsular Malaysia. Hydrology Earth System Sci Disc 6:5471-5503.

Xu J, Xu K, Li Z, Meng F, Tu T, Xu L, Liu Q, 2020. Forecast of dengue cases in 20 Chinese cities based on the deep learning method. Int J Environ Res Public Health 17:453.

Xu J, Xu K, Li Z, Tu T, Xu L, Liu Q, 2019. Developing a dengue forecast model using long short term memory neural networks method. bioRxiv, 760702.

Yuan HY, Wen TH, Kung YH, Tsou HH, Chen CH, Chen LW, Lin PS, 2019. Prediction of annual dengue incidence by hydro-climatic extremes for southern Taiwan. Int J Biometeorol 63:259-268.

Zhao N, Charland K, Carabali M, Nsoesie EO, Maheu-Giroux M, Rees E, Zinszer K, 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 Trop Dis 14:e0008056.

Zheng H, Lin F, Feng X, Chen Y, 2020. A hybrid deep learning model with attention-based conv-lstm networks forfor short-term traffic flow prediction. IEEE Trans Intelligent Transport Syst 22;6910-6920.

Zhu G, Xiao J, Zhang B, Liu T, Lin H, Li X, Hao Y, 2018. The spatiotemporal transmission of dengue and its driving mechanism: a case study on the 2014 dengue outbreak in Guangdong, China. Sci Total Environment 622:252-259.

Zhu L, Laptev N, 2017. Deep and confident prediction for time series at uber. In 2017 IEEE International Conference on Data Mining Workshops, pp. 103-110.

Online supplementary material:
Table S1. Summary results of LSTM models' performance in different Malaysian states and different look-back values.
Table S2. Summary results of the benchmark model performance in different Malaysian states and different look-back values.