

Forecasting Dengue Outbreaks in Kalutara District using Hybrid Machine Learning Techniques

E.J.K.P. Nandani* and A.M.D.P. Bandara

Department of Mathematics, University of Ruhuna, Matara, Sri Lanka.

*Corresponding author: nandani@maths.ruh.ac.lk

 ORCID ID: <https://orcid.org/0000-0002-1255-5252>

Received: 24.12.2024

Revised: 07.03.2025

Accepted: 17.11.2025

Online: 15.12.2025

Abstract Dengue is the most rapidly expanding mosquito-borne viral disease that has emerged as a significant global public health concern in tropical and subtropical regions. Almost all districts in Sri Lanka, one of the dominant countries affected by the dengue epidemic, are announcing dengue cases, and Colombo, Gampaha and Kalutara districts in the western province have recorded the highest case rate. Ongoing prevention efforts aim to reduce dengue risk, and accurate forecasting is vital for effective surveillance and control. This study demonstrates that a hybrid machine learning approach improves vector-borne disease forecasting accuracy. The spread of dengue depends on factors such as weather, urbanization, population density, water storage containers and tanks, etc. This study aims to design an Artificial Neural Network (ANN) model utilizing the Backpropagation (BP) algorithm and Levenberg-Marquardt (LM) algorithm independently, to forecast the average number of Dengue cases for next week in Kalutara District where one of the highest dengue-reported districts in Sri Lanka, based on various weather factors, and enhance the accuracy of the models by applying gradient boosting method. The model's input and output were selected as the current week's climate factors and the number of Dengue cases for the following week in the Kalutara district, respectively. Weekly data from 2022 January to 2023 June, a period with minimal external impact due to COVID-19, were collected from District Secretariat Kalutara, and NASA POWER Project which provides solar and meteorological data sets from NASA research. Parameters and accuracy of the models were measured based on minimum values of mean square error (MSE), root mean square error (RMSE) and mean absolute error (MAE). While the ANN model with the LM algorithm initially performed well (MSE=0.012166, RMSE=0.113006, and MAE=0.069093), applying gradient boosting improved model accuracies, and the BP algorithm later achieved high accuracy (MSE=0.000613, RMSE=0.025006 and MAE=0.017425). These results show that hybrid machine learning algorithms effectively enhance the accuracy of vector-borne disease forecasting.

Keywords: Artificial Neural Network, Dengue, Forecasting, Gradient boosting, Learning-algorithms

Introduction

Mosquitoes play a crucial role in the spread of epidemic arbovirus diseases. Worldwide, *Aedes aegypti* and *Aedes albopictus* mosquitoes are the main vectors responsible for dengue transmission (Li et al., 2020; Li et al., 2023). Dengue has emerged as a significant global public health concern, and Sri Lanka is one of the dominant countries affected by the dengue epidemic in recent years due to a tropical island in the Indian Ocean (Li et al., 2023). Dengue has a seasonal transmission in Sri Lanka, with two peaks occurring with the monsoon rains in June-July and October-December. Most cases occur during the summer monsoon in June-July. Almost all districts in Sri

Lanka are announcing dengue cases, and Colombo, Gampaha and Kalutara districts in the western province have recorded the highest case rate every year (Walczak & Cerpa, 2003).

The disease has a wide spectrum of clinical presentations, from undifferentiated dengue fever (DF) to dengue hemorrhagic fever (DHF), to life-threatening dengue shock syndrome (DSS), creating significant health, economic and social burdens in endemic areas (Ministry of Health, n.d.). The spread of dengue depends on factors such as weather, urbanization, population density, and water storage containers and tanks, etc. (Tissera et al., 2020). Cities are rapidly becoming more



urbanized, and poor water sanitation management makes it easier to reproduce and spread dengue. Government and community-related vector control techniques like fogging, larvicide, and environmental management are used to reduce mosquito reproduction (Tsai et al., 2013). Unfortunately, dengue prevention and control efforts are proving less than successful in reducing the spread of Dengue transmitted vectors (Murugesan & Manoharan, 2020) because of insufficient budgetary support and intensified efforts at both the national and community levels (Tissera et al., 2020). However, the World Health Organization (WHO) recommends the use of two or more vector control approaches such as environmental control, Mechanical control, biological control, and chemical control (Tissera et al., 2020). As well it describes a promising, low-cost, year-round vector control measure that is feasible to implement, is acceptable and safe to the public, and, once established, has minimal recurring costs (Tissera et al., 2020).

Accurately forecasting Dengue incidence is essential for effective surveillance and disease control. Therefore large number of studies around the world have addressed about the early warning system under different techniques such as time series approaches, (Autoregressive Integrated Moving Average (ARIMA) model (World Health Organization, 2013; Withanage et al., 2018), Seasonal ARIMA (Withanage et al., 2018), Multiple time series regression (Karasinghe et al., 2024), Poisson multivariate regression model (Martheswaran et al., 2022), stochastic SIR model (Othman et al., 2022), Machine Learning Techniques (Artificial Neural Network (ANN) (Yusof & Mustaffa, 2011; Aburas et al., 2010; Herath & Perera, 2014), Least Square-support Vector Machines (LS-SVM) (Yusof & Mustaffa, 2011), Numerous studies have revealed the influence of weather variables on the magnitude of dengue distribution (Martheswaran et al., 2022; Yusof & Mustaffa, 2011; Cho et al., 2022; Herath & Perera, 2014). Some creative methods for dengue prevention and control in Sri Lanka include ongoing population analysis and container classifications to pinpoint high-risk breeding places (Sri Lanka Journal of Infectious Diseases, 2016).

The dynamics of dengue transmission are significantly shaped by environmental elements such as temperature, precipitation, and humidity, as

well as vector control measures and community involvement. A few crucial variables are age distribution, population density, socioeconomic status, and educational background. Creating a more reliable and precise forecasting model that can track the complex dynamics of dengue outbreaks would have the capacity to produce accurate forecasts (World Health Organization, 2019)

Several studies conducted in Sri Lanka (Munasinghe, Premaratne, & Fernando, 2013) paid attention to developing an early warning ANN system by utilizing data from the Western and Southern provinces.

Although there are various studies, an accurate early warning system is an essential tool for pre-epidemic preparedness and the effectiveness of dengue control. Therefore, this study focuses on designing an Artificial Neural Network (ANN) model with Levenberg Marquardt and Back-propagation optimization learning algorithm independently to forecast the following week's Dengue outbreak in Kalutara district, Sri Lanka based on present week weather details, and enhance the accuracy of the model by applying gradient boosting method. The input and output for the ANN model were selected as present-week climate factors and a number of reported Dengue cases in the following week in the Kalutara district respectively. To ensure data consistency and reduce the influence of external factors, weekly data from 2022 January to 2023 June, a period with COVID-19, were collected from district secretariat Kalutara and NASA POWER Project which provides solar and meteorological data sets from NASA research (<http://power.larc.nasa.gov/data-access-viewer>).

Methodology

Data Collection

The Kalutara district, identified as one of high-risk areas by analyzing the dengue spreading in Sri Lanka from 2020-2023 -- a period with minimal external influence from COVID-19, was selected for the study. This study was done over 13 Medical Offices of Health (MOH) divisions which covered 14 divisional secretary areas, in Kalutara district. For the period January 2022-June 2023, the number of weekly reported dengue cases was collected from the regional epidemiology unit Kalutara, Sri Lanka, and weather variables---average minimum and

maximum temperature, wind speed, UV Index, total rainfall, and average relative humidity were extracted from the online resource. (<https://power.larc.nasa.gov/data-access-viewer/>), data access viewer for NASA Prediction of Worldwide Energy Resource (POWER) project.

Designing the ANN models

Train and test data sets consist without repetition by randomly selected 80% and 20% of the collected data respectively. The input-output data pattern for the ANN model was selected as present week climate factors ---

There were a total 716 data records, and each variable normalized to [0,1] by using the eq. (1).

$$x'_i = \frac{x_i}{\max(x)} \quad (1)$$

where x'_i is the normalized value of the i^{th} data x_i in vector x .

average minimum and maximum temperature, wind speed, UV Index, total rainfall, average relative humidity, and the number of reported Dengue cases for following week in Kalutara district.

All neurons in ANN model was activated under sigmoid function given in eq. (2).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

Relevant parameter of the ANN model were selected based on the minimum MSE. The training process of the ANN model was conducted using two different algorithms: Backpropagation (Ministry of Health, n.d.) and Levenberg-Marquardt (Fathima & Manimeglai, 2015). The Backpropagation algorithm utilizes the first derivative of the loss function, while the Levenberg-Marquardt algorithm incorporates both

first and second derivatives for more efficient optimization. The weights updates rules of Backpropagation and Levenberg Marquardt algorithms are given in eq. (3) and (4) respectively. Weights update rules, eq. (3) and (4), are built up by minimizing the considering the loss function (L) given in eq (5), and curvature of the loss surface (S) given in eq. (6), respectively.

$$W_{i+1} = W_i - \alpha \frac{dL}{dW} \quad (3)$$

$$W_{i+1} = W_i - \left[(J_i^T \cdot J_i + \lambda_i \cdot I)^{-1} \cdot (2 \cdot J_i^T \cdot e_i) \right] \quad (4)$$

$$L = \frac{1}{N} \sum_{i=1}^N (e_i^2) \quad (5)$$

$$S = \frac{1}{2N} \sum_{i=1}^N (e_i^2) \quad (6)$$

where W_i - weight matrix in the i^{th} iteration, α -learning rate, λ -step size, I -identity matrix, J - Jacoby matrix and e is the error that is the difference between actual and network output.

The suitable values for the parameters such as the number of nodes for the hidden layer, learning rate, and number of iterations were selected by comparing each variable verses MSE separately. Test set with final updated weights was used to

measure the accuracy of the developed ANN model for each learning algorithm separately.

In the next step, the gradient boosting technique was applied to the outputs of the ANN model to enhance forecasting performance by focusing on the errors from the previous step. This gradient

boosting model was trained using the dataset along with the predictions from the ANN model. Update

the output of the m^{th} data record $F_m(x)$ using the eq. (7),

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x), \quad (7)$$

where γ_m and $h_m(x)$ are learning rate and residue of the m^{th} data record. Number of trees for the gradient boosting technique and the accuracy comparison between forecasting models were measured using MSE and RMSE.

Results and Discussion

Figure (1) shows the weekly reported dengue cases in Sri Lanka from 2020 to 2023. Reported dengue cases has been reduced in 2021 due to the Covid 19 pandemic, people have limited outdoor activities and started adopting new and healthy lifestyle behaviors. As well reported dengue cases have been reduced during weeks 7 - 17, and 35-45. Week 7-17 occurs after the northeast monsoon but before the inter-monsoon season, resulting in dry conditions that limit standing water resources where Aedes mosquitoes breed. Similarly, weeks 35-45, which fall after the southwest monsoon and before the northeast monsoon, see reduced rainfall, further limiting mosquito breeding opportunities.

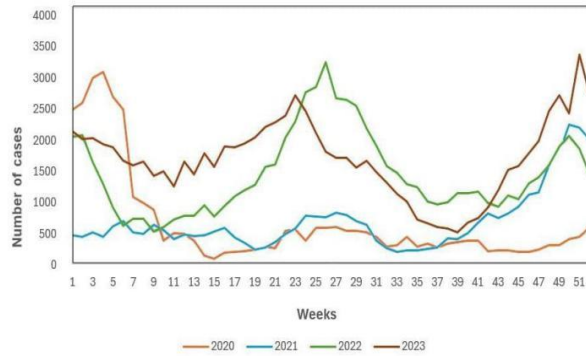


Figure 1: Dengue Distribution in Sri Lanka from 2020 to 2023.

Table (1) presents the province-wise percentage distribution of annually reported dengue cases in Sri Lanka for the period 2020-2023. The Western Province, which includes the Kalutara, Gampaha, and Colombo districts in Sri Lanka, has reported a higher number of dengue cases compared to other districts each year.

Table 1. Province-wise Dengue percentage distribution

Province Name	2020	2021	2022	2023
Western	28.0	58.9	48.5	46.0
Central	13.5	5.4	10.6	12.7
Southern	8.3	5.5	9.2	7.9
Northern	8.9	1.7	5.1	4.2
Eastern	23.5	12.1	6.2	8.1
North Western	4.7	7.4	8.3	8.7
North Central	2.2	1.4	1.3	1.6
Uwa	1.7	2.7	2.3	2.8
Sabaragamuwa	9.2	4.8	8.4	8.0

Province-wise percentage distribution of annually reported dengue cases in Sri Lanka for the period 2020-2023 (Ministry of Health, n.d.).

Based on this analysis, the study is focused on the Kalutara district, one of the higher dengue-reported districts in Sri Lanka, for the period 2022 January -2023 June. There are 13 MOH divisions and a higher quantity of reported dengue cases was observed in high population density and low hygienic maintenance areas according to the details given in Figure 2.

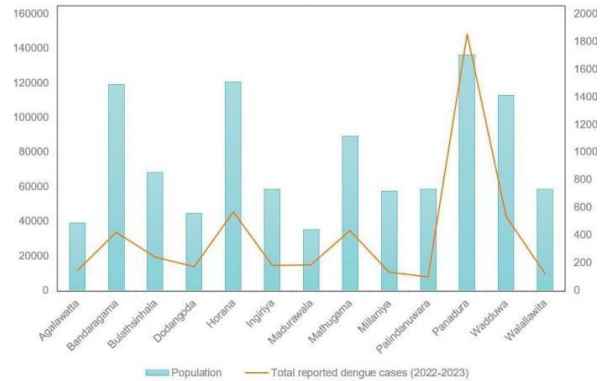


Figure 2: Bar graph and line graph show the average population and reported dengue cases in Kalutara district, separately, for 2022-2023. Left and right y-axis correspond to the bar graph and line graph respectively.

Pearson's correlation coefficient between each pair of variables are shown in Figure (3). There is no significant correlation between any pair of input variables and the input-output variable.

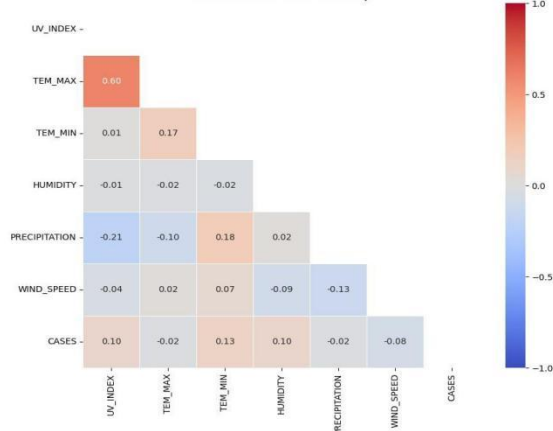


Figure 3: Heat map for Pearson's correlation coefficient between pair of input variables and input-output variable.

Based on the minimum MSE, evidenced in Figure (4) number of hidden nodes for both learning algorithms was selected as 4.

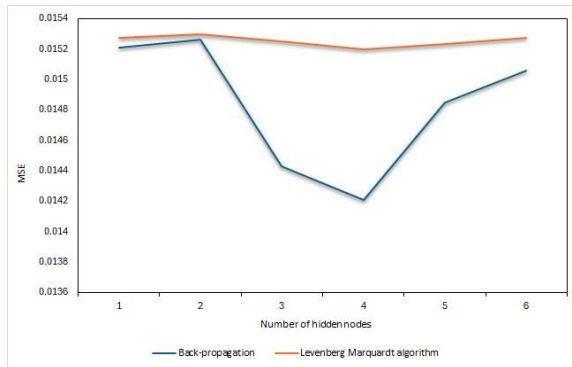


Figure 4: Measure the MSE against the different number of hidden nodes.

Figure (5) illustrates the constructed ANN architecture for both learning algorithms. The input layer consists of 6 neurons, the hidden layer contains 4 neurons, and the output layer has 1

neuron.

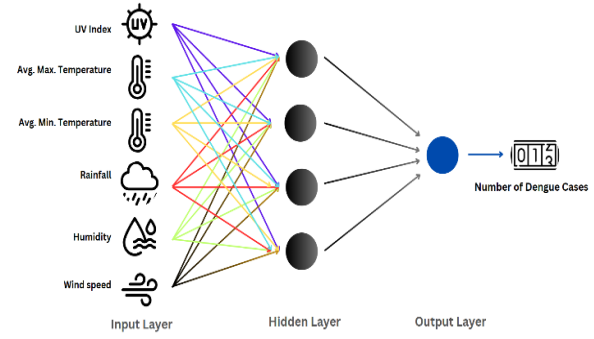


Figure 5: Built-up 6-4-1 ANN architecture for both learning algorithms.

The learning rate was set to 0.01 for both learning algorithms based on the results shown in Figure (6). As evidenced in Figures (7) and (8), the number of iterations for the training process was chosen as 9000 for back-propagation and 250 for the Levenberg Marquardt algorithm. While the Levenberg Marquardt algorithm reaches an MSE of 0.01 in 150 iterations, back-propagation takes 9000 iterations to reach the same MSE. This is because the Levenberg-Marquardt algorithm is a second-order optimization method, which typically leads to faster and more accurate convergence, whereas the back-propagation algorithm relies solely on the first derivative of the error function during optimization.

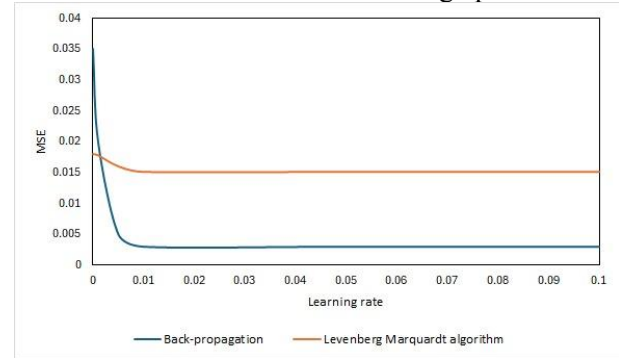


Figure 6: Measure the MSE against the learning rate.

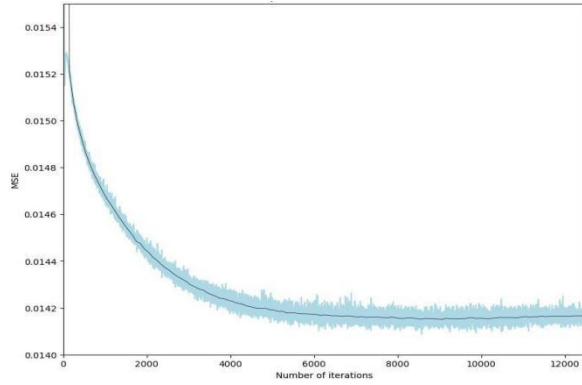


Figure 7: Measure the MSE against the number of iterations by the ANN model with a back-propagation learning algorithm.

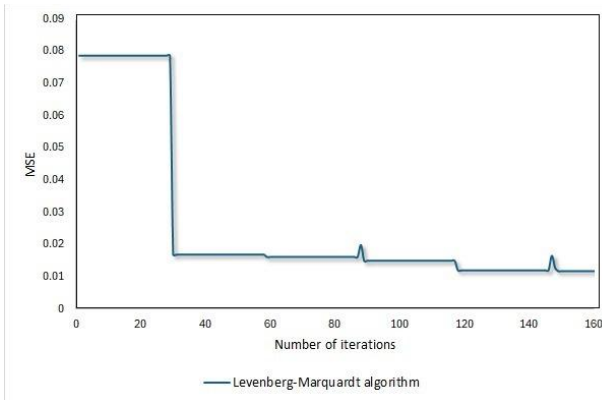


Figure 8: Measure the MSE against the number of iterations by ANN model Levenberg Marquardt algorithm.

As the second step of the study, gradient boosting was used to enhance the accuracy of the developed ANN model. The loss function value converges to zero when the number of iterations (number of trees) are greater than 300 with a learning rate of 0.1. After applying the gradient boosting technique to ANN output values, prediction accuracy was measured by using a test set. Figures 4 and 5 display the actual values and the predicted values before and after the application of the gradient boosting technique to ANN model, and Table 2 shows the error measurements of each model.

According to the values provided in Table 2, the ANN model using the Levenberg-Marquardt algorithm outperforms the back-propagation algorithm in the first step of the calculation. This result demonstrates that the Levenberg-Marquardt algorithm is best suited for small and medium-sized models.

The prediction performance of ANN model I and model II was increased after applying the gradient boosting technique. After the second step calculations, the ANN model with back-propagation performs better than Levenberg-Marquardt since gradient boosting strengthens for a weaker model.

Table 2. Prediction performance

Model	MSE	RMSE	MAE
Model 1	0.014106	0.119210	0.069204
Model 2	0.012166	0.113006	0.069093
Model 3	0.000613	0.025006	0.017425
Model 4	0.014203	0.117456	0.068770

Prediction performance of Model I (ANN with Back-Propagation algorithm), Model II (ANN with Levenberg Marquardt algorithm), Model III (Results of Model I were enhanced by Gradient Boosting technique) and Model IV (Results of Model II were enhanced by Gradient Boosting technique).

There are other factors such as population density, number of water collected places, and patient-based data that cause to dengue transmission. This study could not use the above factors due to the unavailability of sufficient data.

The accuracy of the model will be able to increase by using other factors as input variables. This type of results will help to make necessary decisions and actions for relevant authorities.

Conclusion

ANN model trained with the Levenberg-Marquardt algorithm outperformed back-propagation ANN, achieving lower prediction errors across all metrics MSE = 0.012166, RMSE=0.113006, and MAE=0.069093, compared to MSE=0.014106, RMSE=0.119210, and MAE=0.069204 for backpropagation algorithm. Furthermore, gradient boosting technique with the back-propagation improved accuracy of the ANN, reducing RMSE from 0.119 to 0.025 by iteratively correcting residual errors. After applying gradient boosting improved the model accuracies, and the BP algorithm later achieved high accuracy (MSE=0.000613, RMSE=0.025006 and MAE=0.017425). The accuracy of the model will be able to increase by including other factors such

as weather data, economic indicators, or demographic factors, as input variables.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Author contributions:

Bandara A.M.D.P. conducted the data collection, calculation, writing the initial draft of the manuscript, preparing all preliminary figures and tables, and Nandani E.J.K.P. provided overall project direction and scientific insight, guided the manuscript structure and content development, critically reviewed and corrected the manuscript.

References

- Aburas, H. M., Cetiner, B. G., & Sari, M. (2010). Dengue confirmed-cases prediction: A neural network model. *Expert Systems with Applications*, 37(6), 4256–4260.
- Analytics Vidhya. (2021). *Gradient boosting algorithm: A complete guide for beginners*. <https://www.analyticsvidhya.com/blog/2021/09/gradient-boosting-algorithm-a-complete-guide-for-beginners/>
- Caicedo-Torres, W., Montes-Grajales, D., Miranda-Castro, W., Fennix-Agudelo, M., & Agudelo-Herrera, N. (2017). Kernel-based machine learning models for the prediction of dengue and chikungunya morbidity in Colombia. In *Advances in Computing* (pp. 472–484). Springer.
- Cho, C., Kwon, K., & Wu, C. (2022). On weather data-based prediction of gamma exposure rates using gradient boosting learning for environmental radiation monitoring. *Sensors*, 22, 7062. <https://doi.org/10.3390/s22017062>
- de Jesús Rubio, J. (2020). Stability analysis of the modified Levenberg–Marquardt algorithm for the artificial neural network training. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8), 3510–3524.
- Fathima, A. S., & Manimeglai, D. (2015). Analysis of significant factors for dengue infection prognosis using the random forest classifier. *International Journal of Advanced Computer Science and Applications*, 6(2), 240–245.
- Hamid, P. H., Prastowo, J., Nasrin, N., Toriba, M., Hirayama, K., & Kusumawati, R. L. (2018). Species composition and blood meal sources of *Anopheles malaria* vectors in South Halmahera District, North Maluku Province, Indonesia. *Parasites & Vectors*, 11, 316. <https://doi.org/10.1186/s13071-018-3005-2>
- Herath, H. P. H. M. N., & Perera, A. A. I. (2014). Prediction of dengue outbreaks in silence using artificial neural networks. *International Journal of Computer Applications*, 101.
- 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), e1908. <https://doi.org/10.1371/journal.pntd.0001908>
- Karasinghe, N., Peiris, S., Jayathilaka, R., & Dharmasena, T. (2024). Forecasting weekly dengue incidence in Sri Lanka: Modified autoregressive integrated moving average modeling approach. *PLoS ONE*, 19(3), e0299953. <https://doi.org/10.1371/journal.pone.0299953>
- Li, Y., An, Q., Sun, Z., et al. (2023). Distribution areas and monthly dynamic distribution changes of three *Aedes* species in China: *Aedes aegypti*, *Aedes albopictus*, and *Aedes vexans*. *Parasites & Vectors*, 16, 297. <https://doi.org/10.1186/s13071-023-05924-9>
- Li, Y., Feng, Z., Chen, S., Zhao, Z., & Wang, F. (2020). Application of the artificial neural network and support vector machines in forest fire prediction in the Guangxi Autonomous Region, China. *Research Article*, 2020.
- Martheswaran, T. K., Hamdi, H., Al-Barty, A., et al. (2022). Prediction of dengue fever outbreaks using climate variability and Markov chain Monte Carlo techniques in a stochastic susceptible-infected-removed model. *Scientific Reports*, 12, 5459. <https://doi.org/10.1038/s41598-022-09489-y>
- Medical Research Institute & Dengue Coordination Unit of Ministry of Healthcare and Nutrition of Sri Lanka. (2011). *The practical manual and guidelines for dengue vector surveillance*.
- Ministry of Health. (n.d.-a). *Dengue in Sri Lanka: Situation update*. <https://www.dengue.health.gov.lk/web/index.php/en/information/sri-lanka-situation>

- Ministry of Health. (n.d.-b). *National Dengue Control Program*. <https://www.dengue.health.gov.lk/>
- Munasinghe, A., Premaratne, H., & Fernando, M. (2013). Towards an early warning system to combat dengue. *International Journal of Computer Science and Electronic Engineering*, 1(2), 252–256.
- Murugesan, A., & Manoharan, M. (2020). Dengue virus. In *Emerging and reemerging viral pathogens* (pp. 281–359). <https://doi.org/10.1016/B978-0-12-819400-3.00016-8>
- Othman, M., Indawati, R., Suleiman, A. A., Qomaruddin, M. B., & Sokkalingam, R. (2022). Model forecasting development for dengue fever incidence in Surabaya City using time series analysis. *Processes*, 10(11), 2454. <https://doi.org/10.3390/pr10112454>
- ResearchGate. (n.d.). *Backpropagation algorithm*. https://www.researchgate.net/figure/Backpropagation-Algorithm-This-research-uses-training-in-the-form-of-a-backpropagation_fig2_376002876
- Rudnick, A., & Lim, T. W. (1986). *Dengue fever studies in Malaysia*. Bulletin from the Institute for Medical Research in Malaysia.
- Sri Lanka Journal of Infectious Diseases. (2016). Dengue transmission and vector dynamics in a tropical urban setting. *Sri Lanka Journal of Infectious Diseases*, 6(1), 44–53. <https://sljid.sljol.info/articles/8107/files/submission/proof/8107-1-28941-1-10-20160428.pdf>
- Suratgar, A. A., Thawaloki, N. B., & Hosienibadi, M. (2005). Modifying Levenberg–Marquardt algorithm for neural network. *World Academy of Science, Engineering and Technology*.
- Tissera, H. A., Jayamanne, B. D. W., Raut, R., Janaki, S. M. D., Tozan, Y., Samaraweera, P. C., et al. (2020). Severe dengue epidemic, Sri Lanka, 2017. *Emerging Infectious Diseases*, 26(4), 682–691. <https://doi.org/10.3201/eid2604.190435>
- Tsai, J. J., Chokephaibulkit, K., Chen, P. C., Liu, L. T., Hsiao, H. M., Lo, Y. C., et al. (2013). Role of cognitive parameters in dengue hemorrhagic fever and dengue shock syndrome. *Journal of Biomedical Science*, 20(1), 88. <https://doi.org/10.1186/1423-0127-20-88>
- Walczak, S., & Cerpa, N. (2003). Artificial neural networks. In R. A. Meyers (Ed.), *Encyclopedia of Physical Science and Technology* (3rd ed., pp. 631–645). Academic Press. <https://doi.org/10.1016/B0-12-227410-5/00837-1>
- Withanage, G. P., Viswakula, S. D., Gunawardena, Y. I. S., et al. (2018). A forecasting model for dengue incidence in the District of Gampaha, Sri Lanka. *Parasites & Vectors*, 11, 262. <https://doi.org/10.1186/s13071-018-2828-2>
- World Health Organization. (2013). *Managing regional public goods for health – Community-based dengue vector control*.
- World Health Organization. (2019, July 8). *Preventive action is vital to curtail dengue outbreaks in Sri Lanka*. <https://www.who.int/srilanka/news/detail/08-07-2019-preventive-action-is-vital-to-curtail-dengue-outbreaks-in-sri-lanka>
- Yusof, Y., & Mustaffa, Z. (2011a). Dengue outbreak prediction: A least squares support vector machines approach. *International Journal of Computer Theory and Engineering*, 3(4), 489.
- Yusof, Y., & Mustaffa, Z. (2011b). Dengue outbreak prediction: A least squares support vector machines approach. *International Journal of Computer Theory and Engineering*, 3(4), 489.