INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

LSTM-based Forecasting of Dengue Cases in Gujarat: A Machine Learning Approach

Amiben Maheshbhai Mehta[1], Kajal S Patel[2]*

1 Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat, India
2 Assistant Professor, VGEC, Chandkheda, Gujarat Technological University, Gujarat, India

OPEN ACCESS

Received: 11-02-2023

Accepted: 23-12-2023

Published: 14-02-2024

Citation: Mehta AM, Patel KS (2024) LSTM-based Forecasting of Dengue Cases in Gujarat: A Machine Learning Approach. Indian Journal of Science and Technology 17(7): 635-642. https://doi.org/10.17485/IJST/v17i7.2748

* **Corresponding author**.

kajalpatel@vgecg.ac.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: Dengue fever, a mosquito-borne viral disease, is particularly prevalent in tropical regions like India. Gujarat State is also one of them. Forecasting outbreaks of diseases such as dengue can prove important for public health management. The purpose of this study is to predict dengue cases in ten districts of Gujarat using the LSTM machine learning model. And if people are aware of this from the beginning, the spread of dengue can be prevented. **Methods:** This approach uses LSTM models to predict dengue cases using a total of 10 years (2010 to 2019) of data. From this data, data from 2010 to 2016 is used for training and data from 2017 to 2019 is used for testing. To predict dengue cases, population density, average temperature, average humidity, monthly rainfall, dengue cases with lag of one, two and twelve months. **Findings:** The LSTM model was applied with different parameter configurations, showing the following results: The root mean square error value is 0.04, and the R-squared (R2) score is 0.84. Many machine learning methods, like ANN, linear regression, random forest, etc., have been used to predict dengue cases in different states and countries. LSTM model gives the best results in terms of accuracy. Previously reported dengue cases, population density, and total monthly rainfall proved to be the most effective predictors of dengue in the state of Gujarat. **Novelty:** Models have been developed to predict dengue outbreaks in many other countries and states. The LSTM model is developed for the first time in this study for the state of Gujarat. 84% accuracy is obtained from the model. This model has been prepared by collecting environmental data and registered dengue cases in Gujarat state.

Keywords: Dengue Cases Predictions; Artificial Intelligence in Healthcare; LSTM Algorithm; Disease Outbreaks; Public Health Management

1 Introduction

In Thirteenth general program of work 2019-23, the world health organization (WHO) announced new 5-year strategic plan, to ensure that one billion more people in the world enjoy the benefits of better health and wellbeing (World Health Organization, 2019a). In this program the 10 most and current health issues was highlighted and the dengue was identified as one of the four main infections affecting global health (World Health Organization, 2019b). Specially for dengue disease the five key factors are needed to consider for better public health for this strategy decided is diagnosis and case management, integrated surveillance and outbreak preparedness, sustainable vector control, future vaccine implementation, and basic operational and implementation research. It is observed that DENV virus is mostly related to the weather condition and climatic conditions like temperature, rainfall, and humidity levels which considered as one of the main influences in dengue fever. It is necessary to study climate factors and to prepare appropriate analysis to detect the dengue outbreaks and define the climate factors affecting on its spread and reducing the intensity of spreading of a dengue as well as its elimination. The prediction strategies for dengue cases can play an important role in making informed decisions regarding preventive measures and tackling the disease effectively. In the year 2019, Gujarat registered a total of 18,219 dengue cases, marking the highest number of cases reported between 2010 and 2022. Similarly, in 2021, there were 10,983 reported cases. In literature study it is observed that Machine learning algorithms are widely employed for prediction purposes of various diseases in different countries and also utilizing various forecasting methods. One notable study focused on forecasting dengue incidences in Brazil, comparing multiple methods including Random Forest Regression, LASSO, and Long Short-Term Memory (LSTM). The results of this study revealed that the LSTM method outperformed the others in terms of accuracy[(1)]. Moreover, the LSTM method was utilized to predict malaria cases in different locations within Telangana, India. The research findings indicated a reasonably accurate LSTM model, achieving a forecast accuracy of 96.11% for Ventakapuram[(2)]. Consequently, the LSTM method is popular and has chosen for the present study to forecast the number of dengue cases. Furthermore, the LSTM method has been employed in forecasting various diseases, as well as phenomena like floods[(3)], air pollution[(4)], and weather conditions. The research outcomes obtained from this study can significantly assist the health agencies and health research centers for identifying preventive measures and taking appropriate actions to reduce the incidence of dengue fever in the future.

The presented research work focuses on dengue case prediction in various districts of the Gujarat state, India. Researchers from different regions have developed a machine learning-based model to predict dengue cases in specific regions. After a thorough worldwide literature review of various studies, it is observed that studies related to a particular region of Gujarat state are not highlighted for the prediction of dengue cases using machine learning techniques. It is necessary to develop a technique for predicting dengue cases in the specific Gujarat state region because during the literature review, no specific model was found for dengue case prediction in Gujarat state. The presented research work aims to develop an LSTM (Long short term memory) model to predict dengue cases in the future for some districts of Gujarat state.

2 Methodology

2.1 Data Description

The data on dengue case incidences from 2010 to 2019 was obtained from the Health and Family Welfare Department, Government of Gujarat, as part of the National Vector Borne Disease Control Program. They provided zone-wise data for six zones. Meteorological information such as temperature, humidity, and rainfall was collected from the Metrological Division under the Ministry of Earth Sciences for each district of Gujarat. However, the Meteorological Department has indicated in a letter that district-wise weather data is not available. Instead, they provided meteorological center-wise data. To reconcile and align these datasets, a total of ten districts were identified for research, which are Ahmedabad, Amreli, Bhavnagar, Bhuj, Dwarka, Porbandar, Rajkot, Surat, Vadodara, and Valsad. Furthermore, population density information was sourced from the official website of the Indian Census for the year 2011. Including this data allows for a better understanding of the demographic factors that may influence the incidence of dengue cases.

Dengue is a viral infection caused by the dengue virus (DENV), which is transmitted to humans by the bite of infected mosquitoes. So if the number of cases of dengue increases, it can spread to other people from the infected mosquito bites. Dengue cases are highly dependent on weather factors. It is more likely to happen in the monsoon period, so the number of dengue cases in the next year can be predicted from the number of dengue cases in that month. Therefore, keeping in mind the above two points, the number of cases of one month ago, two months ago and twelve months ago are considered to predict dengue cases.

Having collected data spanning from 2010 to 2019, employing city-wise data for machine learning models proved challenging due to the limited amount of information available. To address this limitation, the data was consolidated all the district-level data into a single file and conducted the analysis under this comprehensive scenario. It's important to note that the accuracy metrics derived from this approach are collective for all ten districts rather than individual accuracy values for each district.

This aggregation of data provides a more robust foundation for the machine learning model, ensuring a more reliable and meaningful assessment of the overall predictive accuracy across the entire dataset. A total of 10 years' worth of monthly data from 10 different cities was combined, resulting in a total of 1200 rows used for further implementation.

Table 1. Research Variable

Sr. No.	Name of Variable	Definitions	Measurements
1.	Number of Dengue Fever Cases	Number of dengue fever cases or number of patients	Incidents
2.	Temperature	Monthly temperature average	Celsius degree
3.	Humidity	Monthly humidity average	%
4.	Rainfall	Monthly rainfall average	Millimetre
5.	Population Density	Population per unit land area	No of people

2.2 Pre – Processing Data

This phase aimed to transform raw data into a format suitable for utilization in the formation of a forecasting model. The initial step involved processing the humidity and temperature values for each district, which were originally provided in maximum and minimum forms. The primary task was to calculate the average values for temperature and humidity. The range of values of raw data tends to have different scales. In machine learning algorithms, objective functions may not effectively perform without data normalization. Therefore, to ensure fairness and meaningful comparisons, it is necessary to normalize (scale) the range of all features, allowing each feature to have values within the same range. Min-Max scaling is a method used to rescale the range of feature values, typically transforming them to fit within the range of [0, 1] or [-1, 1]. The selection of the target range depends on the characteristics of the data being analyzed. So Min-Max normalization is used for the data normalization.

2.3 Parameter Selection

Dengue outbreaks are influenced by lagged variables, which refer to historical values of specific parameters shifted forward in time. These lagged values serve to capture the temporal relationship between a variable and its impact on dengue outbreaks. Dengue is primarily transmitted through the bite of infected mosquitoes. When a person contracts dengue, they can become a potential source of the virus for other mosquitoes. As the number of infected individuals increases, the probability of mosquito bites on infected individuals rises, leading to further transmission of the disease. Therefore, the previous registered cases serve as an indicator of the existing pool of infected individuals and the potential for continued transmission. Several studies have highlighted the relationship between lagged value of rainfall and dengue virus infection in different regions: In Thailand, a positive association was observed between two months of cumulative rainfall and dengue virus infection when the temperature exceeded 23.2 °C[(5)]. In Singapore, dengue incidence showed a linear increase in response to cumulative rainfall, with a lag period ranging from 5 to 20 weeks. A recent analysis conducted in Dhaka, Bangladesh, utilizing data from a private diagnostic facility, revealed a positive correlation between rainfall and the number of dengue cases, with a lag period of two months. In Dhaka, Bangladesh, it was found that rainfall, particularly the number of rainy days, exhibited a significant association with dengue occurrence in a high-incidence area, with a one-month lag[(6)]. In India, the transmission risk of dengue was observed to be higher between 8 and 15 weeks of lag, coinciding with precipitation patterns[(7)]. Based on previous research, the following parameters have been carefully considered for implementation. Three distinct scenarios have been chosen as the final set of parameters. Each scenario offers a unique combination of factors to be considered in the implementation process.

Table 2. Three Different Scenario of Parameter Selection

Scenario 1	Scenario 2	Scenario 3
Rainfall with a lag of one month	Rainfall with a delay of two months	Past cases of dengue
Dengue cases with a lag of one month and twelve months	Dengue cases with a lag of one month and twelve months	Population density
Average temperature	Average temperature	Month of the year
Average humidity	Average humidity	
Month of the year	Month of the year	
Population density	Population density	

2.4 LSTM Forecasting Model

The Long Short-Term Memory (LSTM) is a deep machine learning method initially developed by Hochreiter and Schmidhuber in 1997. Its primary purpose was to overcome the vanishing gradient problem that arises in Recurrent Neural Network (RNN) architecture. LSTM employs memory blocks, which are a collection of subnets connected in a recurrent manner. Each block incorporates three multiplicative gates, namely the forget gate, input gate, and output gate, enabling operations on the cells and interconnected memory cells. Over the past decade, LSTM has found widespread application in addressing various real-world challenges that demand the utilization of long-term contextual information. Such applications include protein structure prediction, speech recognition, and handwriting recognition[8]. In this research, the LSTM architecture in the univariate case, which is illustrated in Figure 1, is utilized. At this stage, the development of LSTM forecasting models was conducted in two steps. Firstly, determination of the optimal combination of data partition proportions and the number of timestamps that yielded the lowest error values for each variable's usage scenario. Once the best combination was obtained, the next step involved parameter tuning, where experiments were carried out with various values for the parameters used in the formation of the LSTM forecasting model. The parameters considered in this phase included optimizer, epochs, and batch size. It's important to note that some of these parameter values were based on previous research findings[1], while others were adopted from relevant literature[9] and underwent trial-and-error for optimization purposes.

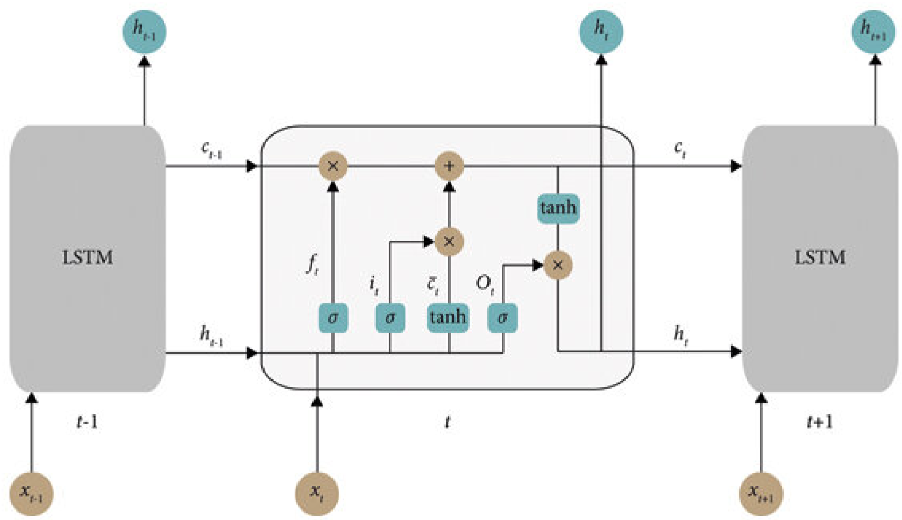

Fig 1. The LSTM architecture consisting of the input layer, a single hidden layer, and the output layer

2.4.1 Conducting tests and evaluations on the forecasting model

To evaluate a forecasting model, two commonly used metrics are the Root Mean Square Error (RMSE) and the R-squared (R2) score. These metrics provide valuable insights into the accuracy and goodness-of-fit of the model. They can be represented as follows:

Root mean square error (RMSE):

RMSE measures the average magnitude of the prediction errors made by the forecasting model. It is calculated as the square root of the average of the squared differences between the predicted values and the actual values.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \tag{1}$$

(Best value = 0; worst value = +∞)

R-squared (R2) Score:

R2 score, also known as the coefficient of determination, evaluates the proportion of the variance in the dependent variable that is predictable from the independent variables used in the model. It ranges from 0 to 1, where 0 indicates that the model does not explain any variance, and 1 indicates a perfect fit.

$$R2 = 1 - \frac{\sum_{i=1}^{m} (X_i - Y_i)^2}{\sum_{i=1}^{m} (\bar{Y} - Y_i)^2} \tag{2}$$

(Worst value = $-\infty$; best value = +1)

The evaluation phase involved testing the constructed model using an independent test set of data. This test was conducted to generate forecasted results for the number of dengue cases. Subsequently, the performance of each developed model was assessed by calculating the error values using R2 Square and RMSE metrics.

3 Results and Discussion

The results of the LSTM prediction model development are presented here. A model was created by combining various data partition proportions and exploring three different scenarios, as mentioned in Table 2 earlier. The experiments were also conducted on two different types of LSTM: bidirectional LSTM and stacked LSTM. Experimentation was performed with three different data partitions: 90:10, 70:30, and 80:20. The outcomes of each model, with the best results from the three data partitions, are tabulated.

Table 3. Best result of LSTM among three data partition proportions (80:20, 70:30 and 90:10)

Parameter Selection Scenario	RMSE		R2 Score	
	Training	Testing	Training	Testing
1	0.02	0.04	0.89	0.83
2	0.02	0.04	0.89	0.84
3	0.02	0.05	0.85	0.79

Table 4. Best result of Bidirectional LSTM among three data partition proportions (80:20, 70:30 and 90:10)

Parameter Selection Scenario	RMSE		R2 Score	
	Training	Testing	Training	Testing
1	0.01	0.05	0.91	0.80
2	0.02	0.04	0.89	0.84
3	0.02	0.05	0.85	0.79

Table 5. Best result of Stacked LSTM among three data partition proportions (80:20, 70:30 and 90:10)

Parameter Selection Scenario	RMSE		R2 Score	
	Training	Testing	Training	Testing
1	0.02	0.05	0.89	0.80
2	0.02	0.04	0.88	0.83
3	0.02	0.05	0.88	0.81

At this stage of model building, the following parameter values were used: Adam optimizer, 150 epochs, and a batch size of 4. Among the LSTM models, the one with the highest accuracy was obtained when using a data partition proportion of 70% for training data and 30% for test data. The selected parameters for this model are based on the 2nd scenario, which includes features such as month of year, population density, average temperature of month, rainfall before two months, average humidity, number of dengue incidences before one month, and number of dengue incidences in the same month of the previous year.

Furthermore, in the Bidirectional LSTM model, the highest accuracy was achieved with the same data partition proportion (70:30) and the 2nd scenario feature selection. Similarly, in the Stacked LSTM model, the best accuracy was observed when using a 70:30 data partition and the 2nd scenario features. The best-performing model exhibits an impressive Root Mean Square Error (RMSE) value of 0.02 for the training dataset and 0.04 for the testing dataset. Additionally, the R2 score, which measures the goodness of fit, reaches 0.89 for the training dataset and 0.84 for the testing dataset.

The comparison between actual data and forecasted data is presented in Figures 2 and 3. The graph clearly shows that the forecasting model performed quite well in capturing the general trend of the actual data, even though the values were not an exact match. The graph indicates that the model was particularly adept at following the actual data pattern, especially during periods of increasing actual data. However, it encountered challenges in tracking the actual data pattern when the values decreased and remained consistently low for several periods.

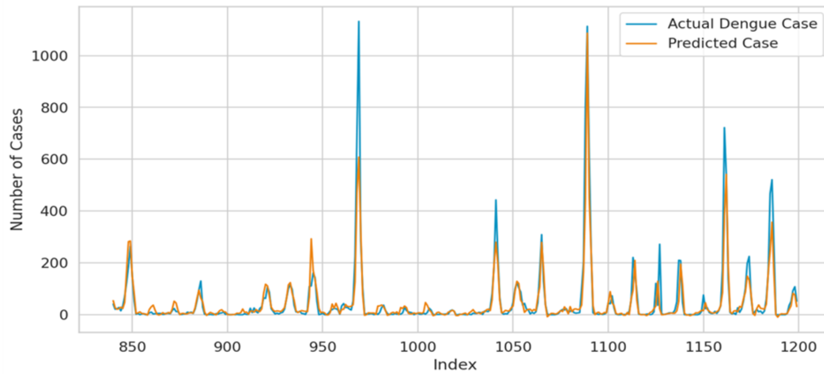

Fig 2. Comparison of actual data and forecast data on testing of LSTM Algorithm (70:30 Data Partition)

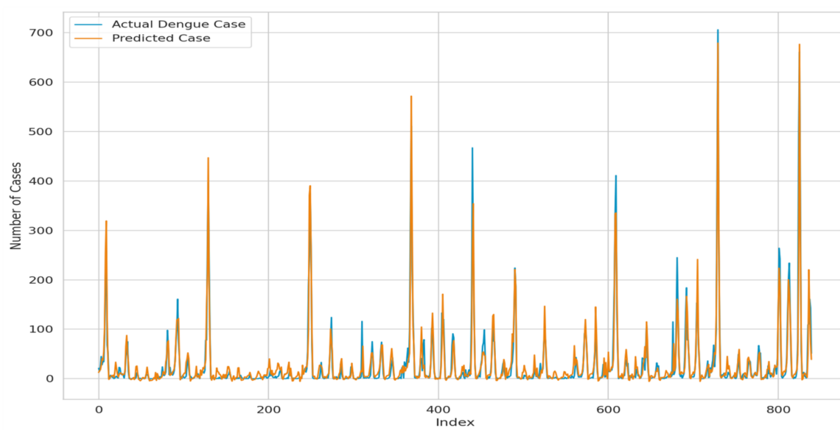

Fig 3. Comparison of actual data and forecast data on training of LSTM Algorithm (70:30 Data Partition)

The following Table 6 shows the comparison of this research work with the previous research and the results and conclusions of the previous research are shown.

Table 6. Comparison, Validation and Discussion with Relevant Literature

Sr. No.	Validation Source Description	Results and discussion	(Favouring/Contradicting with presented research)	Overview/Outcome
1	Predicted Dengue Outbreak in Rio de Janeiro, Brazil Using Gradient Boosting Decision Tree Algorithm (CatBoost). For this, the National System of Information on Notifiable Diseases (SINAN) and Brazilian socio-demographic data Used by the Institute of Geography and Statistics (IBGE).[10]	It gives an rmse value of 83.0 for three months ahead forecast in Baguna district and 81.0 in Realango.	Favouring	The result indicates that the classification matrix is more useful than the RMSE when evaluating the model's ability to predict outbreaks.

Continued on next page

Table 6 continued

2	Generalized Additive Models (GAMs) is used to predict dengue outbreaks in 50 districts of Thailand using a total of 8 years of data from 2008 to 2015.[11]	The predictive model, could explain only 73 percent of the variation in the occurrence of dengue cases. The remaining 27 percent unexplained variation could be due to the influence of other factors	Favouring	Generalized Additive Models (GAMs) has used to predict dengue but RMSE and SRMSE value for analysis of result and it shows that out of sample result is not as in sample
3	LSTM-RNN has been used to predict dengue cases in India using data from 2014 to 2019. The parameters used for forecasting are climatic conditions, temperature, rainfall data, humidity and population. This research focuses on forecasts of dengue cases for various Union Territories and States of India.[12]	Research has presented result in which LSTM training model gives 89% accuracy.	Training model gives same accuracy with presented research.	LSTM is more accurate than other machine learning approaches like BPNN, SVM, XGBoost, Random Forest and GAM.
4	This research focuses on predict dengue fever cases in Malaysia using six different LSTM models. Here monthly registered dengue cases from 2010 to 2016 were used with various climate, topographic, demographic, and land-use variables.[13]	The SSA-LSTM model, that used stacked LSTM layers and spatial attention, which performed the best, with an average root mean squared error (RMSE) of 3.17.	Favouring	For different states of Malaysia the SSA-LSTM model perform differently with RMSE value ranging from 2.91 to 4.55.
5	In this research the early prediction of dengue fever was developed in the Vietnam using attention-enhanced LSTM (LSTM-ATT) model and other different variants of lstm model. Data from year 1997 to 2013 were used to train the model and 2014 to 2016 were used to test the model.[14]	LSTM-ATT shows the highest performance, scoring 1.60 for RMSE-based ranking and 1.95 for MAE-based ranking.	Favouring	LSTM-AT Tis method can accurately predict the dengue cases and outbreak months up to 3 months ahead.
6	Forecast weekly dengue incidence in 790 cities in Brazil LSTM model was used and compared with LASSO and Random Forest Regression.[15]	Mean prediction errors in quantile scale for three different cities were calculated for all three model.	Favouring	LSTM model gives noticeable lowest error value among three models.

3.1 Presented results validation with previous results of relevant literature

As mentioned above, LSTM gives better results compared to other models. The LSTM model is trained with data from Gujarat State in India. Metrological data and dengue incidence in 10 different cities of Gujarat State were collected from 2010 to 2019. The model was trained with data from 2010 to 2016. It gives an accuracy of up to 89% on the in-sample value and 84% on the out-sample value. The RMSE value is noticeably small, which is 0.04.

4 Conclusion

The LSTM method has yielded a highly effective forecasting model, leveraging key factors such as the month of year, population density, average temperature of the month, rainfall two months' prior, average humidity, the number of dengue incidences one month prior, and the number of dengue incidences during the same month in the previous year. This model was developed with a data partition ratio of 70:30, utilizing the Adam optimizer, 150 epochs, a learning rate of 1e-3, and a batch size of 4. Impressively, the training dataset achieved an exceptionally low root mean square error (RMSE) value of 0.02, with the testing dataset exhibiting a similarly favorable RMSE of 0.04. Furthermore, the model's Coefficient of Determination (R2) scores were notably high, with recording values of 0.89 for training and 0.84 for testing. The conducted experiments underscore the model's sensitivity to variations in data partition proportions, chosen variables, dataset sources, and parameter values. When compared against other variants like stacked LSTM and bidirectional LSTM, the LSTM-based forecasting model consistently outperformed them. Notably, its predictive prowess exhibited superior accuracy and precision. Looking ahead, potential avenues for future research include incorporating district-level data on a daily basis from Gujarat State. This augmentation promises a richer dataset, enhancing the model's predictive capabilities and accuracy. Additionally, expanding the set of independent variables and thoughtfully selecting those highly correlated with the occurrence of dengue fever cases holds promise for further improving the model's forecasting accuracy and insights.

References

1) Mussumeci E, Coelho FC. Large-scale multivariate forecasting models for Dengue - LSTM versus random forest regression. *Spatial and Spatio-temporal Epidemiology*. 2020;35:100372. Available from: https://doi.org/10.1016/j.sste.2020.100372.
2) Santosh T, Ramesh D, Reddy D. LSTM based prediction of malaria abundances using big data. *Computers in Biology and Medicine*. 2020;124:103859. Available from: https://doi.org/10.1016/j.compbiomed.2020.103859.
3) Ding Y, Zhu Y, Feng J, Zhang P, Cheng Z. Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing*. 2020;403:348–359. Available from: https://doi.org/10.1016/j.neucom.2020.04.110.
4) Chang YS, Chiao HT, Abimannan S, Huang YP, Tsai YT, Lin KM. An LSTM-based aggregated model for air pollution forecasting. *Atmospheric Pollution Research*. 2020;11(8):1451–1463. Available from: https://doi.org/10.1016/j.apr.2020.05.015.
5) Abdulsalam FI, Yimthiang S, La-Up A, Ditthakit P, Cheewinsiriwat P, Jawjit W. Association between climate variables and dengue incidence in Nakhon Si Thammarat Province, Thailand. *Geospatial Health*. 2021;16(2):1–14. Available from: https://doi.org/10.4081/gh.2021.1012.
6) Rahman KM, Sharker Y, Rumi RA, Khan MUIUI, Shomik MS, Rahman MW, et al. An Association between Rainy Days with Clinical Dengue Fever in Dhaka, Bangladesh: Findings from a Hospital Based Study. *International Journal of Environmental Research and Public Health*. 2020;17(24):1–9. Available from: https://doi.org/10.3390/ijerph17249506.
7) Chovatiya M, Dhameliya A, Deokar J, Gonsalves J, Mathur A. Prediction of Dengue using Recurrent Neural Network. In: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE. 2019;p. 926–929. Available from: https://doi.org/10.1109/ICOEI.2019.8862581.
8) Kakarla SG, Caminade C, Mutheneni SR, Morse AP, Upadhyayula SM, Kadiri MR, et al. Lag effect of climatic variables on dengue burden in India. *Epidemiology and Infection*. 2019;147:1–10. Available from: https://doi.org/10.1017/s0950268819000608.
9) Nayak SDP, Narayan KA. Prediction of dengue outbreaks in Kerala state using disease surveillance and meteorological data. *International Journal Of Community Medicine And Public Health*. 2019;6(10):4392–4400. Available from: https://doi.org/10.18203/2394-6040.ijcmph20194500.
10) Aleixo R, Kon F, Rocha R, Camargo MS, de Camargo RY. Predicting Dengue Outbreaks with Explainable Machine Learning. In: 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), 16-19 May 2022, Taormina, Italy. IEEE. 2022;p. 940–947. Available from: https://doi.org/10.1109/CCGrid54584.2022.00114.
11) Jain R, Sontisirikit S, Iamsirithaworn S, Prendinger H. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. *BMC Infectious Diseases*. 2019;19(1):1–16. Available from: https://doi.org/10.1186/s12879-019-3874-x.
12) Doni AR, Sasipraba T. LSTM-RNN Based Approach for Prediction of Dengue Cases in India. *Ingénierie des systèmes d information*. 2020;25(3):327–335. Available from: https://doi.org/10.18280/isi.250306.
13) Majeed MA, Shafri HZM, Zulkafli Z, Wayayok A. A Deep Learning Approach for Dengue Fever Prediction in Malaysia Using LSTM with Spatial Attention. *International Journal of Environmental Research and Public Health*. 2023;20(5):1–22. Available from: https://doi.org/10.3390/ijerph20054130.
14) Van-Hau Nguyen, Tuyet-Hanh TT, Mulhall J, Van Minh H, Duong TQ, Van Chien N, et al. Deep learning models for forecasting dengue fever based on climate data in Vietnam. *PLOS Neglected Tropical Diseases*. 2022;16(6):1–22. Available from: https://doi.org/10.1371/journal.pntd.0010509.
15) Mussumeci E, Coelho FC. Machine-learning forecasting for Dengue epidemics - Comparing LSTM, Random Forest and Lasso regression. 2020. Available from: https://doi.org/10.1101/2020.01.23.20018556.