International Journal of Biometeorology (2023) 67:285–297
https://doi.org/10.1007/s00484-022-02405-z

ORIGINAL PAPER

Weather integrated multiple machine learning models for prediction of dengue prevalence in India

Satya Ganesh Kakarla[1,2] · Phani Krishna Kondeti[1] · Hari Prasad Vavilala[1] · Gopi Sumanth Bhaskar Boddeda[1] · Rajasekhar Mopuri[1] · Sriram Kumaraswamy[1,2] · Madhusudhan Rao Kadiri[1,2] · Srinivasa Rao Mutheneni[1,2]

Received: 29 September 2021 / Revised: 21 July 2022 / Accepted: 4 November 2022 / Published online: 16 November 2022

Abstract

Dengue is a rapidly spreading viral disease transmitted to humans by *Aedes* mosquitoes. Due to global urbanization and climate change, the number of dengue cases are gradually increasing in recent decades. Hence, an early prediction of dengue continues to be a major concern for public health in countries with high prevalence of dengue. Creating a robust forecast model for the accurate prediction of dengue is a complex task and can be done through various data modelling approaches. In the present study, we have applied vector auto regression, generalized boosted models, support vector regression, and long short-term memory (LSTM) to predict the dengue prevalence in Kerala state of the Indian subcontinent. We consider the number of dengue cases as the target variable and weather variables viz., relative humidity, soil moisture, mean temperature, precipitation, and NINO3.4 as independent variables. Various analytical models have been applied on both datasets and predicted the dengue cases. Among all the models, the LSTM model was outperformed with superior prediction capability (RMSE: 0.345 and R^2:0.86) than the other models. However, other models are able to capture the trend of dengue cases but failed in predicting the outbreak periods when compared to LSTM. The findings of this study will be helpful for public health agencies and policymakers to draw appropriate control measures before the onset of dengue. The proposed LSTM model for dengue prediction can be followed by other states of India as well.

Keywords Dengue · Kerala · India · Modelling · Machine learning · Deep learning

Introduction

Dengue is the most common mosquito-borne viral disease prevalent in tropical and sub-tropical regions of the world. According to the World Health Organization (WHO), about half of the World's population is at risk of dengue infection. The risk of dengue infection is spreading rapidly over the last two decades and it has increased more than eight fold. Dengue virus (DENV) is a *Flavivirus* belongs to family *Flaviviridae*, and there are four related but antigenically distinct, serotypes (DENV1, DENV2, DENV3, DENV4) of the virus that cause dengue fever. Prior to 1970, dengue was prevalent in nine countries, has now spread in more than 125 countries, and the number of dengue infections increased from 505,430 cases in 2000 to 2.4 million in 2010 and 5.2 million in 2019 (WHO 2020, Stanaway et al. 2016).

Dengue was prevalent in the America, South-East Asia (SEA), and Western Pacific regions of the World. The Asia represents ~ 70% of the global burden of disease (WHO 2020). In recent years, the dengue infections have increased over 400% in Asia and WHO estimated that there would be 100 million symptomatic cases and 300 million asymptomatic cases, annually (WHO 2012). Approximately 87% of the total population in the SEA region are at risk of dengue (WHO 2012).

In 2010, nearly 34% of globally occurred dengue infections contributed by India alone (approximately 96 million cases) (Bhatt et al. 2013). Dengue is endemic in India and reported frequent or continuous risk of dengue transmission across all states with substantial annual and geographic variation (Jentes et a. 2016). Recent findings suggest that

✉ Srinivasa Rao Mutheneni
msrinivas@iict.res.in

[1] ENVIS Resource Partner On Climate Change and Public Health, Applied Biology Division, CSIR-Indian Institute of Chemical Technology (CSIR-IICT), Tarnaka, Hyderabad 500007, Telangana, India

[2] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

the dengue was varied greatly between regions such as the highest seroprevalence was in southern regions (76·9%), followed by western (62·3%), and northern (60·3%) regions. Similarly, the urban areas (70·9%) contributing high dengue seroprevalence than rural areas (42·3%) (Wilder-Smith & Rupali 2019). The expansion of dengue cases in India may be due to unplanned rapid urbanisation, population immunological factors, population movement, climate change, poor surveillance, and inadequate vector control operations (Mutheneni et al. 2017).

Dengue virus is transmitted to humans by the bite of infected female mosquitoes *Aedes aegypti* and *Aedes albopictus*. The *Aedes aegypti* mosquito is considered as primary vector and *Aedes albopictus* is secondary vector for dengue transmission in Asia. These mosquitoes are also acting as vectors for chikungunya, yellow fever, and Zika viruses (WHO 2020, Stanaway et al. 2016). Both mosquito species are diurnal, biting mostly in the morning and evening rather than at night periods (Trpis et al. 1973).

Several studies have demonstrated the association between climate variability and dengue infections. Climate variables such as rainfall, relative humidity, and temperature were found to be associated with dengue transmission (Mutheneni et al. 2017; Kakarla et al. 2020; Bal & Sodoudi 2020; Mala & Jat 2019). However, the association between climate factors and dengue varied across geographical locations and socio-environmental strata (Arcari et al. 2007; Thammapalo et al. 2008). Rainfall provides the ample breeding sites for *Aedes* vectors whereas temperature helps to fasten the life cycle. Furthermore, the temperature also influences the virus development (extrinsic incubation period), mosquito biting rate, human to vector and vector to human transmission of parasite, and mosquito mortality (Kakarla et al. 2020). Similarly, the relative humidity influences the longevity of vectors which can lead the infected female mosquito to complete more than one cycle of virus replication, thus becoming infective (Donalisio & Glasser 2002).

Currently, there is no effective vaccine is available for to prevent the dengue infections and the available vaccine does not provide equal protection against all four serotypes (Thomas and Yoon 2019). Thus, controlling mosquito vectors and reducing human-vector contact to reduce dengue transmission are the primary prevention strategies (WHO 2020). To address this critical public health issue, many health authorities have focused their attention towards early warning system to reduce disease burden and to allocate health resource effectively. Hence, the present study has applied statistical method vector auto regression (VAR), machine learning methods support vector regression (SVR) and generalized boosted regression model (GBM), and deep learning method long short-term memory (LSTM) to forecast dengue well in advance. Machine learning models were successfully applied to predict dengue in the Philippines, China, Singapore, and Europe (Guo et al. 2017; Carvajal et al. 2018; Ong et al. 2018; Salami et al. 2020). Early and accurate forecast of dengue using various modelling approaches might minimize the threat and also helps the policy makers for effective disease control. In the same way, the present study explored various machine learning and deep learning algorithms on integrated datasets (epidemiology and climate data) to forecast dengue cases in the Kerala state of India.

Materials and methods

Study area

The state Kerala is situated in the south-western end of the Indian subcontinent and geographically located between 8° 17′ 30" and 12° 47′ 40" north latitude and 74° 27′ 47" and 77° 37′ 12" east longitudes. It lies between the Arabian Sea in the west and the Western Ghats in the east with an area of 38,863 km^2. The state shares boarder with Tamil Nadu and Karnataka states. Geographically Kerala is divided into three parts highland, mid plains and coastal areas. The area in and around the Western Ghats is mostly hilly and thick evergreen rainforests. The major rivers of Kerala originate from these highlands. The lakes and backwaters make Kerala a water rich land mass state. The state consist of rich bio-diversity and Western Ghats region is one of the biodiversity hotspots in the world.

Epidemiological data

Reported month wise positive dengue case data were obtained from the annual reports of the state surveillance unit, Integrated Disease Surveillance Programme (IDSP), Directorate of Health Services, Kerala for the study period from January 2003 to December 2017. The incidence rate of dengue was calculated by dividing the total number of cases with the total population of Kerala state for a particular year and district was multiplied by a factor of per million population.

Meteorological data

The monthly meteorological data of mean temperature (mean. temp in °C), total rainfall (mm), relative humidity (RH in %), soil moisture (SM), and NINO 3.4 were obtained from the Indian Meteorological Department (IMD), Pune and National Oceanic and Atmospheric Administration (NOAA), USA for the period of 2003 to 2017.

Data processing

The dataset consists of dengue cases and weather variables from January 2003 to December 2017 were used for the study. Prior to the analysis, feature engineering was carried out to generate new features of variables or transform original data in to more informative format. The study considers two datasets, the first dataset is raw data consist of weather parameters such as monthly mean temperature, rainfall, relative humidity (RH), soil moisture (SM), NINO 3.4, and dengue cases (ln: natural logarithm), and this dataset is denoted as DS1. The second dataset consists of delayed time lags to the first dataset to take temporal characteristics into account and this dataset is denoted as DS2. To capture the delayed effect of climatic factors on dengue transmission, cross correlation analysis was performed and determined the appropriate lag time for each meteorological factor between 0 to 12 months lag. The most significant lagged terms were observed those are Nino3.4 with two-month lag (Nino3.4_2), mean temperature with three months lag (Mean.Temp_3), soil moisture with one-month lag (SM_1), rainfall and relative humidity with zero-month lag (Rainfall_0 & RH_0) were considered as explanatory variables in lagged climate dataset (DS2). Similarly, the autocorrelation analysis (ACF) was performed to assess the impact of previous dengue cases on the current dengue cases. A strong autocorrelation was observed with 1-month lag; hence, ACF1 feature is considered as one of the explanatory variable along with significant lagged terms of climate factors.

Prior to the analysis, the datasets were split into training and testing data. Train data consists of observations from January 2003 to December 2015 for DS1 and April 2003 to December 2015 for DS2. Similarly, test data is considered from January 2016 to December 2017 for both DS1 and DS2. The training dataset was used to build the models whereas, the test data used to validate the models.

Data modelling

In the present study, statistical, machine learning and deep learning models such as vector auto regression (VAR), support vector regression (SVR), generalized boosted regression (GBM), and long short-term memory (LSTM) models were implemented to predict the dengue prevalence in Kerala. The overall framework of the study and stepwise modelling pathway is illustrated in Fig. 1. The natural logarithm (ln) values of dengue cases taken as dependent or response variable whereas, climate variables and their corresponding lags were used as independent variables. The comparison of the model performance proceeded using the inverse logarithm of the predicted value.

Vector autoregressive (VAR) model

The vector autoregressive (VAR) model is an approach to describe the interaction of variables through time in a complex multivariate system. The VAR model helps to assess the degree of dependence between components, the main reasons are as follows: 1) VAR models are dynamical models that can capture the temporal structure in the variations of individual components and in the interdependence between them. 2) The parameters of autoregressive model are relatively easy to estimate by solving a linear regression problem. 3) Random processes can be very well approximated by a sufficiently high order autoregressive (AR) model. 4) The VAR models form a natural context in which measures of directed influence based on the concept of Granger causality analysis (Goebel et al. 2003). In the present study, the VAR model was used to capture pattern and fluctuations of dengue cases through different time periods of climate parameters.

The mathematical notation for VAR model with p-Lags can be formed as

$$Y_t = k + \varnothing_1 Y_{t-1} + \varnothing_2 Y_{t-2} + \cdots + \varnothing_p Y_{t-p} + \epsilon_t, t = 1, ..., T \quad (1)$$

Equation 1 is a resemblance of multiple linear regression equation where $Y_t = (Y_{1t}, Y_{2t}, \ldots, Y_{nt})'$ is a $(n \times 1)$ vector of time series variables, k denotes the intercept of the time series curve, $\varnothing_i (i = 1, 2, \ldots .p)$ is the coefficient matrix of lagged values with the maximum entry of order p, and ϵ_t is white noise process with multivariate normal distribution of zero mean and a constant variance.

The VAR model works with inter correlated variables. That is, we can predict the variable with the past values of itself (auto correlations) along with past values of other variables (cross correlations) in the system. Granger causality test was performed to analyse the causal relationships between variables. The test is a statistical hypothesis test for determining whether one time series is useful for forecasting another. The test was computed between weather factors and dengue cases (ln), and the null hypothesis shows that the coefficients of past values in the regression equation is zero, more precisely it shows that the past values of time series (X) do not cause the other series (Y). Therefore, the p-value is less than the significance level ($P < 0.05$), and we can safely reject the null hypothesis.

Augmented Dickey Fuller (ADF) test was generally used to assess data stationarity. The ADF test was applied to the DS1 and DS2 datasets to assess stationarity in time series. If a series is said to be stationary, the p-value should be less than the significant level (5%) and it rejects the null hypothesis. The non-stationary series were made stationary by differencing the series and tested for stationarity until every non–stationary series achieve stationarity.

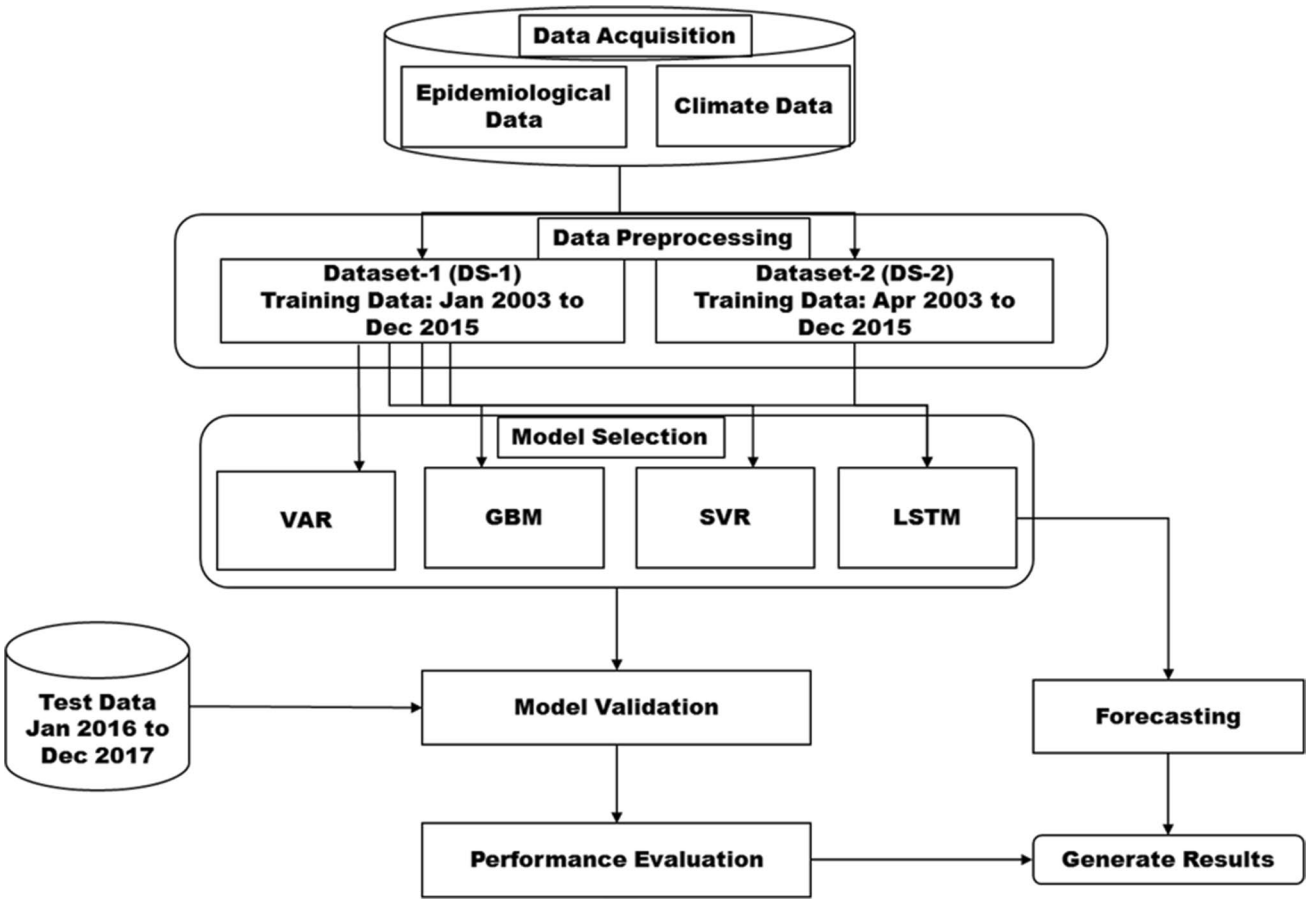

Fig. 1 The overall workflow of the study for dengue prediction

An appropriate lag-length was selected and the maximum lags constrained to 6 months for DS1 using select_order method of VAR class. These lag lengths are frequently selected using an explicit statistical criterion such as the Akaike information criteria (AIC), Bayesian information criteria (BIC), final prediction error (FPE) and Hannan-Quinn information criterion (HQIC) (Lütkepohl 2005). Followed by, Durbin-Watson test (Chris 2019) was performed to check the serial correlation of residuals. The Durbin-Watson statistic will always have a value between 0 and 4. A value of 2.0 indicates that there is no autocorrelation detected in the sample. Values from 0 to < 2 indicate positive autocorrelation and values from 2 to 4 indicate negative autocorrelation.

Generalized boosted regression models (GBM)

Generalized boosted regression models are non-parametric machine learning models which are extended models of Freund and Schapire's AdaBoost algorithm and Friedman's gradient boosting machine learning approach (Greg 2007). In GBM process primarily a regression tree was fitted for response variable and it was iteratively improved in a forward stepwise manner (boosting) by minimizing the variation of response variable. In each iteration, the variation was reduced by fitting error which was not explained in the previous iteration.

The GBM contains two categories of hyper parameters such as boosting parameters and tree specific parameters. Boosting parameters include number of trees and shrinkage or learning rate. The number of trees denotes total number of trees in the sequence. Each tree grown in sequence to reduce past tree's error but high number of trees can lead to over fitting of the model. Hence, optimal number of trees can be found with cross validation. Shrinkage or learning rate determines contribution of each tree on out coming response. Smaller learning rate increases the number of trees and vice versa. Smaller learning rate and larger number of trees is preferable for better model selection. The tree specific hyper parameters which tree depth it captures the interaction between predictor variables with each other (Carvajal et al. 2018). To find the optimal parameters, hyper parameter tuning was performed on train set of DS1, DS2 by creating a grid search function which uses different combinations of hyper parameters to develop each individual

model. The initial hyper parameters such as learning rates: 0.3, 0.1, 0.05, 0.01, and 0.005, interaction depth: 3, 5, 7; n.minobsinnode: 5, 10, 15;cv.folds:3,5,10 were used for the hyper parameter tuning whereas bag.fraction (0.5) was set as default parameter.

Support vector regression (SVR)

Support vector regression (SVR) is most popular and widely used machine learning algorithm. The function approximation can be addressed using $\varepsilon-$ insensitive Support Vector Regression (SVR) whose framework has been derived from Vapnik–Chervonenkis (VC) theory. VC theory helps the learning models to perform better on test data. SVR incorporates regularization term and slack variables to avoid infeasible solutions. SVR outputs a function that has the most training points that lie inside $\varepsilon-$ band (Harris et al. 1997; Smola and Schölkopf 2004; Jayadeva Khemchandani & Chandra 2007).

$$\text{Minimize}\frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$
$$\text{Subjectto}\begin{cases} y_i - w^T x_i - b \le \varepsilon + \xi_i \\ w^T x_i + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \quad (2)$$

Here ξ_i, ξ_i^* are slack variables and $C > 0$ is the regularization term. Let our training data be $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ and w be the projection of training points on the hyperplane. Equation 2 is the objective/cost function that can be solved by converting it into a quadratic programming problem (QPP) using Lagrangian multipliers (Harris et al. 1997). The QPP formulation can be seen in Eq. 3.

$$\text{Maximize}\begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)x_i^T . x_j \\ -\varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \end{cases} \quad (3)$$
$$\text{Subjectto}\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0; 0 \le \alpha_i, \alpha_i^* \le C$$

Here α_j, α_j^* are lagrange multipliers.

The hyperparameter C helps as tradeoff factor between overfitting and underfitting of SVR model. The higher C implies model has high variance, low bias and low C implies low variance, high bias. If C is large, the width of $\varepsilon-$ band is smaller and vice versa. In the present study the data was split into train (12 years), validation (1 year), and test (2 years) datasets. Followed by, the 3 datasets (train, test, validation datasets) were normalized and converted into supervised data using sliding window technique.

Here the window size (i), kernel (k), and the regularization parameter C are considered as hyperparameters. The best kernel was chosen based on hyperparameter tuning using optuna package. The optimal hyperparameters (i, k and C) were used for training the model by merging of both train and validation datasets and testing was done on test dataset.

The hyperparameters were optimized using linear and rbf kernel in hyperparameter space. During hyperparameter tuning, the epsilon (ϵ) is assigned as 0.1 and regularization parameter C consider between 10^{-3} to 1 using log uniform distribution. Similarly, the window size chosen from the integer range between 1 and 6 from uniform distribution. In SVR analysis the ACF 1 is excluded from DS2-SVR data as it inherently deals with auto correlation of dengue cases. The hyperparameter search analysis also ran for 100 trails to find best hyperparameters for SVR analyses. After obtaining the optimal parameters, both training and validation datasets were combined and were used to train the model. Followed by, test data was used to evaluate the model.

Long short-term memory (LSTM)

Long short-term memory (LSTM) is the most effective approach for time series forecasting (Choi & Lee 2018), where classical linear methods are difficult to use multivariate forecasting issues. Among various machine learning methods, the recurrent neural network (RNN) uses feedback loops; subsequently, the network can learn the sequence of information. However, standard RNN fails to remember the long-term information; hence, the present study proposes LSTM network for dengue prediction.

LSTM is the recurrent neural network (RNN) architecture and it was designed by the Hochreiter and Schmidhuber to address the vanishing and exploding gradient problems of traditional RNNs (Hochreiter & Schmidhuber 1997). These vanishing and gradient issues hinder the model accuracy. The LSTM consists of different blocks called memory cells (Figure-S1). Memory block contains memory cells and gates. Memory cells are able to remember and manipulating the temporal state of the memory or network by self-connections and the process of memory or information controlled through three gates. The first gate called as "Forget" gate (f_j), which is responsible for removing information. The second gate is "Input" gate (i_j), which is responsible for addition of information to the cell state. The final gate is "Output" gate (o_j), which selects the useful information from current cell and shows it as an output (Wang et al. 2017). These forget gate, input gate, and output gate of LSTM are described by Hochreiter and Schmidhuber (1997) using the following notations mentioned below:

$$f_j = \sigma(w_f.[h_{j-1}, x_j] + b_f)$$

$$i_j = \sigma(w_i.[h_{j-1}, x_j] + b_i)$$

$$o_j = \sigma(w_o.[h_{j-1}, x_j] + b_o)$$

$$\hat{C}=\tanh(w_c.[h_{j-1}, x_j] + b_c)$$

$$C_j = f_j * C_{j-1} + i_j * \hat{C}$$

$$h_j = o_j * \tanh(C_j)$$

where f, i, o are forget gate, input gate and output gate respectively. The input gate is used to decide the information to be stored in the cell, forget gate is used to decide what kind of information is to be dropped from cell and output gate is used to decide output from the cell. 'σ' represents the logistic sigmoid function.w, b, and h are weights, biases, and value of hidden layers respectively. $\hat{C}$ and C are vector of new candidate values and cell state respectively.

Predictive performance and model validation

Predictive performance of models (VAR, GBM, SVR, and LSTM) measured specifically by coefficient of determination (R^2) and root mean square error (RMSE) as model evaluation metrics. The R^2 measures variability between observed and predicted cases. Similarly, RMSE is a widely used performance measure of a developed model which takes the deviation between observed values and the predicted values. RMSE can be calculated by using the following formula

$$\text{RMSE} = \sqrt{\frac{\sum (y_t - \hat{y}_t)^2}{n}}$$

Here, yt is the observed dengue cases (ln) at time t,$\hat{y}_t$ is the predicted Dengue cases (ln).

Statistical software

R software version 3.6.2 and its packages such as gbm, H2O, dplyr, metrics, and ggplot2 were used for GBM analysis. Python 3.7.4 and its packages such as stats models 0.10.1, pandas, numpy, sci-kit learn, matplotlib, optuna (version 0.19.0), and keras were used for VAR, SVR, and LSTM model development.

Results

Dengue infection was first reported in Kottayam district of Kerala in 1997 with 14 cases, since then the cases are steadily increasing every year. In the year 2001 and 2002, the number of dengue infections are 74 and 163 only. The outbreak of dengue was occurred first time in Kerala in the year 2003 with 3546 cases and 68 deaths. Following the outbreak, the dengue cases were gradually decreased till 2009. From 2010 onwards, the number of dengue cases are gradually increasing and highest number of infections are reported during the year 2017 with 19,994 cases (incidence rate (IR): 555.13 per million population) and 37 deaths. Along with the year 2017, the highest number of cases are also observed in 2013 (7938 cases; IR:225.87) and 2016 (7218 cases; IR:232.08), while analyzing the data from 2003 to 2017 shows that the prevalence of dengue is fluctuated strongly from year to year, and between the months/seasons within a year. A total of 59,801 dengue cases were reported in Kerala during 2003 to 2017, with a mean of 3986 cases annually. During the study period, Kerala has faced six major dengue epidemics in the years 2003, 2010, 2012, 2013, 2016, and 2017 respectively. The year-wise dengue cases and incidence rate of Kerala are shown in Fig. 2.

The spatiotemporal distribution of dengue cases by district wise in Kerala is shown in figure-S2. Out of 14 districts, the highest incidence rate (IR) of dengue reported from Thiruvanathapuram district (IR:510.98), followed by Pathanamthitta (IR:156.95), Kollam (IR:137.92), Kasaragod (IR:108.14), Alappuzha (IR:101.13), Idukki (IR: 96.54), Wayanad (IR:96.16), and other districts contributes IR < 90 respectively (Figure-S3). The month and year wise, district and year wise incidence rates of Kerala state is shown in figure-S4A&B.

In Kerala, the transmission of dengue generally starts in May/June, reaches peak during the July to September and declined gradually in October, and reached the lowest level by November (Figure-S5). The highest monthly dengue cases (5251 cases) were reported in July 2017, whereas total monthly cases from 2003 to 2017 show that July has reported highest number of dengue cases (Figure-S5). The dengue infection is highly seasonal and the highest number of cases are observed in each year during the monsoon period, i.e., June to September (38,087 dengue cases) followed by summer/pre-monsoon season, i.e., February to May (11,032 cases) and winter/post-monsoon period, i.e., October to January (10,682 cases) (Figure-S4B & S5).

The time series plots of monthly weather variables such as mean temperature, rainfall, relative humidity, soil moisture, and Nino3.4 from 2003 to 2017 of Kerala state are shown in Fig. 3. The descriptive statistics of weather variables are shown in Table 1. The minimum and maximum monthly mean temperature were 26 °C and 31.8 °C and the monthly average rainfall 200 mm respectively (Table 1). It is observed that the mean temperatures are slightly increased for the period 2010, 2013, 2016, and 2017 (Fig. 3). The increasing temperature fastens the development of vector and virus replication, and increases the dengue transmission intensity leads to dengue outbreak.

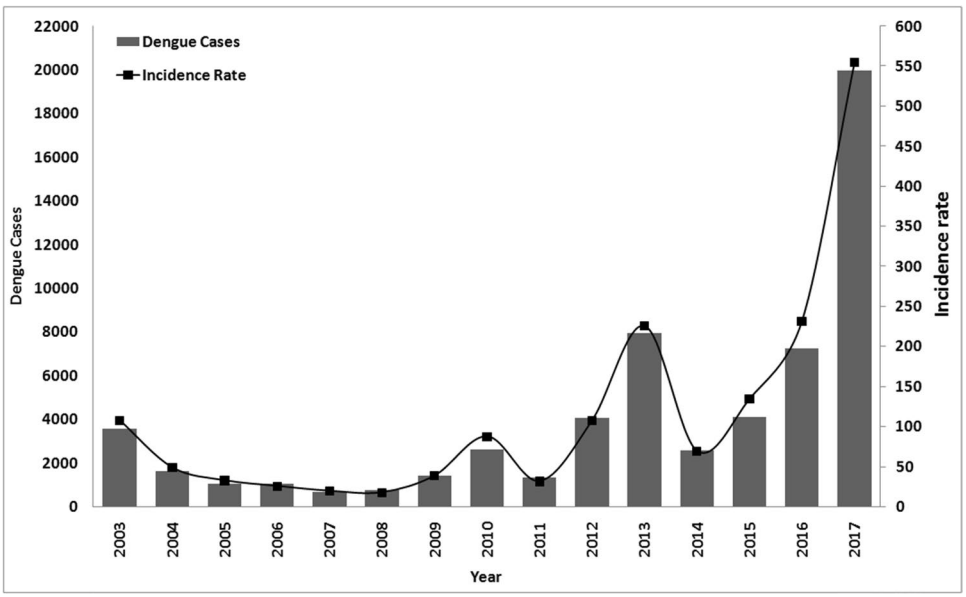

Fig. 2 Year-wise distribution of dengue cases and incidence rate (per million population) in Kerala, India, from 2003 to 2017

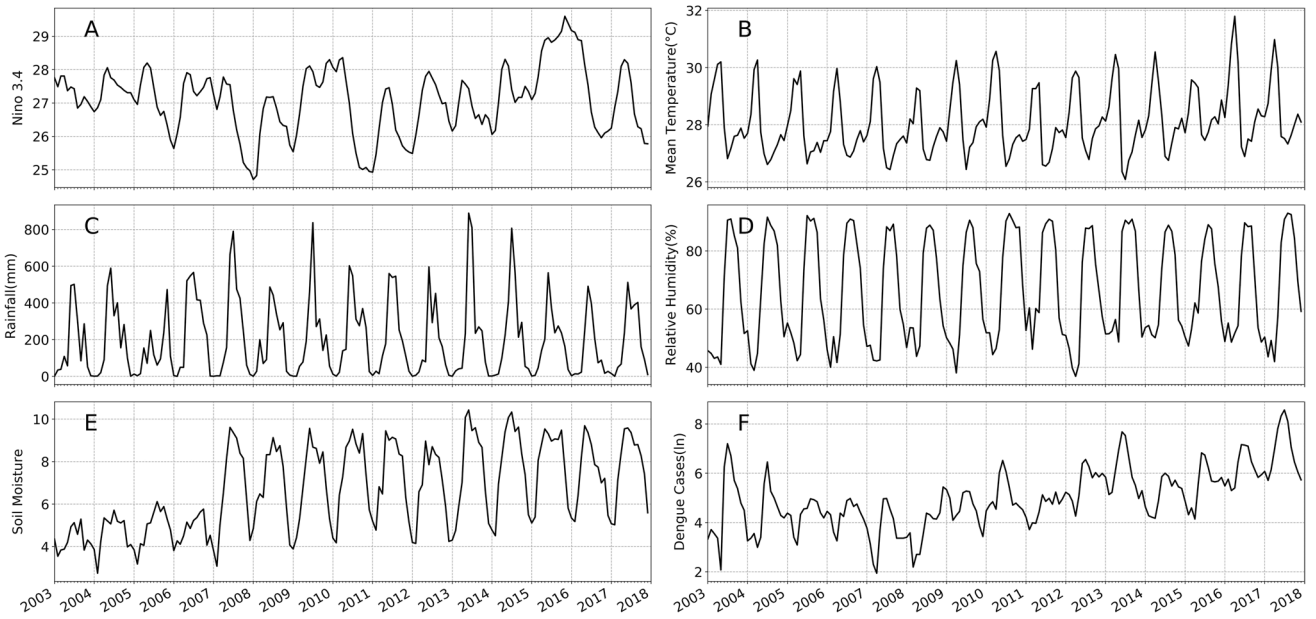

Fig. 3 Time series plots of monthly climate variables and dengue cases from January 2003 to December 2017. (A) Nino3.4, (B) mean temperature, (C) rainfall, (D) relative humidity, (E) soil moisture, (F) dengue cases (ln)

Table 1 Descriptive statistics of monthly weather variables and dengue cases from 2003 to 2017 in Kerala state, India

Variables	Number of months	Minimum	Maximum	Mean	Standard deviation
Dengue cases	180	7	5251	336	636
Mean temperature (°C)	180	26	31.8	28.1	1.136
Rainfall (mm)	180	0.07	889.1	200	208.597
Relative humidity (%)	180	37	93	66	18.002
Soil moisture	180	2.8	10.42	6.63	2.0148
Nino3.4	180	24.71	29.6	27	1.008

The Pearson correlation analysis showed that the weather variables are significantly associated with the dengue cases (Table 2). Similarly, cross correlations analysis was also performed to check any delayed correlations exist between weather factors and dengue cases. The analysis reveals that rainfall and relative humidity has direct association with dengue cases at zero lag. The other climatic factors such as soil moisture have 1-month time lag (t_{-1}) association with dengue cases, Nino3.4 with 2 months lag association and mean temperature with three months lag (t_{-3}) association was observed (Table 2, Figure-S6). Similarly, the auto correlation analysis also performed to find the association of previous months cases on present month (Figure-S6). Moreover, it is found that there is a strong autocorrelation exhibited by dengue cases with 1-month lag (Table 2).

VAR model

The Granger Causality test explains that there is a significant ($p < 0.05$) correlation between climatic variables and dengue cases. Similarly, the ADF test results were shown in table-S1. Before differencing the rainfall and relative humidity have shown stationarity whereas, other variables have shown stationarity after differencing (Table-S1).

The hyper parameters namely trend, lag order, and Information criteria (IC) were used to get the optimal predicted values for dengue prevalence. Followed by, to determine the appropriate number of lags, the Akaike information criteria (AIC) was used as common selection criteria and optimal model was obtained at a lag order of six months with AIC of 6.86234. Table-S2 represent the results obtained from Durbin-Watson test and most of the test values were closure to two or greater than two. Hence, the test explains that there is no serial correlation existing among the residuals of the model. The final results consisting of RMSE and R^2 values calculated based on VAR model presented in Table 3.

Generalized boosted regression model (GBM)

Grid search was performed for hyper parameter optimization and the optimal hyper parameters for both models are shown in table-S3. Based on cross-validation techniques, the optimal number of trees was set as 29 and 40 for DS1-GBM and DS2-GBM respectively (Figure-S7). The relative influence (RI) of each climate variables (DS1) and lagged climate variables (DS2) with generalized boosted regression model is shown in Fig. 4A&B. The DS1 dataset shows that relative humidity (RH) is the highly influencing variable whereas mean temperature is the least influencing variable on dengue transmission. Similarly, none of the climatic factors has shown zero influence; hence, each climate variable is influencing dengue prevalence at different levels. The descending order of climate variables based on RI values are relative humidity (RI = 32.285), rainfall (RI = 21.425), soil moisture (RI = 19.367), Nino3.4 (RI = 17.934), and mean temperature (RI = 8.98) (Fig. 4A). The model developed with metrological factors (DS1-GBM) has shown 0.5653 RMSE and explained 77% of variance observed with dengue

Table 2 Results of correlation, cross correlation (up to three lag periods), and auto correlation

	Correlation		Cross correlation					
Predictor	**Lag-0**	***p*-value**	**Lag-1**	***p*-value**	**Lag-2**	***p*-value**	**Lag-3**	***p*-value**
Nino3.4	0.148	0.048	0.305	< 0.0001	0.396	< 0.0001	0.388	< 0.0001
Mean Temp	− 0.246	0.0009	− 0.001	0.9893	0.306	< 0.0001	0.486	< 0.0001
Rainfall	0.388	< 0.0001	0.351	< 0.0001	0.217	0.0033	0.025	0.7383
Relative humidity	0.477	< 0.0001	0.306	< 0.0001	0.096	0.1986	− 0.1	0.1804
Soil Moisture	0.477	< 0.0001	0.51	< 0.0001	0.473	< 0.0001	0.343	< 0.0001
Lag Cases (Auto correlation)	1	< 0.0001	0.845	< 0.0001	0.636	< 0.0001	0.465	< 0.0001

Table 3 Predictive performance of all developed models

Model	Train data			Test data		
	RMSE	R^2	**R**	**RMSE**	R^2	**R**
VAR	2.48	0.73	0.85	0.572	0.67	0.81
DS1-GBM	0.565	0.77	0.87	1.656	0.36	0.60
DS2-GBM	0.317	0.91	0.95	0.503	0.70	0.84
DS1-SVR	0.401	0.85	0.92	0.441	0.77	0.88
DS2-SVR	0.447	0.81	0.9	0.41	0.8	0.9
DS1-LSTM	0.501	0.72	0.88	0.421	0.79	0.89
DS2-LSTM	0.453	0.81	0.9	0.345	0.86	0.93

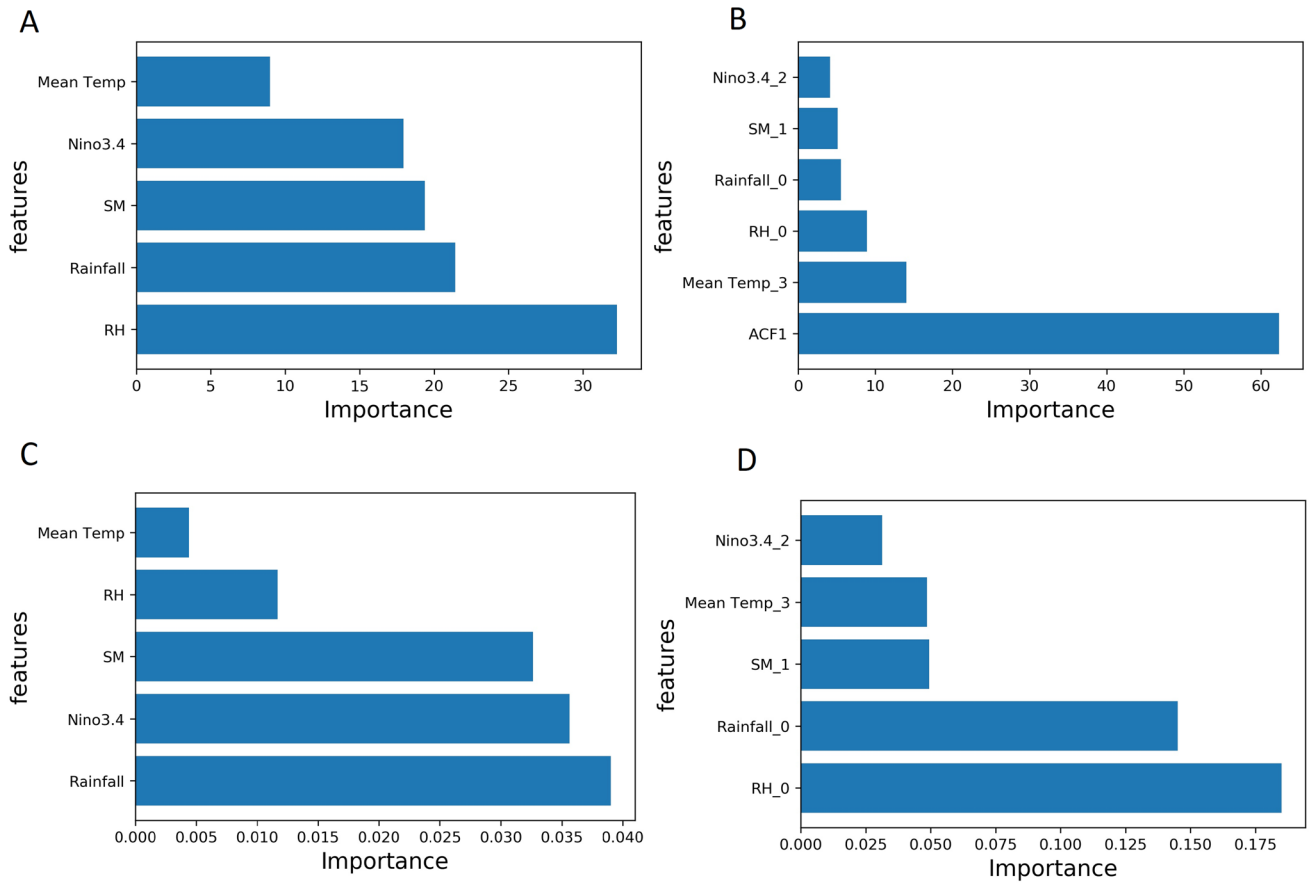

Fig. 4 The plots show the relative importance of features (sum of all features equal to 100) from generalized boosted regression (A, B) and support vector regression (C, D) models with meteorological factors (A, C) and lagged or delayed meteorological factors (B, D)

cases ($r=0.875$, $p=<0.0001$) on train data. Similarly, while validating the model with test dataset has shown RMSE of 1.65 and explained 36% of variance with observed dengue cases ($r=0.604$, $p=0.0014$). However, GBM with lagged dataset (DS2) shows ACF1 is the most influencing variable and Nino3.4 with 2-month lag is the least influencing variable. The descending order of climate variables based on RI values: ACF1 (RI = 62.3282), mean temperature with three months lag (RI = 14.01206), relative humidity (RI = 8.8909), and rainfall (RI = 5.55436) with zero-month lag, soil moisture with 1-month lag (RI = 5.1101) and Nino3.4 with 2 months lag (RI = 4.104) (Fig. 4B). The DS2-GBM analysis shows that 0.31737 RMSE and explained 91% variance with observed dengue cases ($r=0.95$, $p=<0.0001$) on train data. Whereas while validating the trained model with test set it shows 0.5034 RMSE and explained 70% of variance with observed dengue cases ($r=0.83$, $p<0.0001$).

Support vector regression (SVR)

The study used a similar hyper parameter search space for both DS1 and DS2 datasets to predict dengue prevalence. The optimal hypermeters for DS1 & DS2 datasets are listed in table-S4. Figure 4C&D shows that the relative feature importance for both the datasets. Among all the variables, SVR has given a high priority for rainfall and least priority for mean temperature in DS1. Similarly, in DS2, relative humidity is an important feature and Nino 3.4 as a non-influential feature in dengue prediction. Among two datasets, the DS2 is a featured engineered dataset and performed slightly better than DS1 on test data. The RMSE was used to calculate the prediction errors in two datasets and assessed the performance of the model. The RMSE values for train and the test datasets are almost similar and less than one observed. The values for train and test for DS1 and DS2 are shown in Table 3.

LSTM model

Primarily, the time series data was converted into supervised multivariate data. This dataset was split in to train data and test data to build the LSTM model followed by Min–Max normalization technique was used to normalize the data. Optuna, an automatic hyperparameter optimization software

framework, was used for hyperparameter tuning. The table-S5 represents the optimal hyperparameters obtained from the optuna for both DS1 and DS2 datasets. The test dataset was used to evaluate the optimal model for dengue prediction. Figure 5 shows the observed and predicted dengue cases in Kerala using LSTM model. By analyzing the graphs and metrics, it is concluded that LSTM model outperformed other three models. The predicted results of LSTM are quite similar with the real case scenarios of Kerala (Fig. 5).

Assessment of predictive performance

Models' performance and prediction accuracy were measured by using RMSE. The observed and predicted values of all developed models are shown in Fig. 5. The models were well captured the trend of dengue cases but some of the models performed poor in predicting the outbreak periods. Among all the models, LSTM and SVR showed the lowest RMSE and highest R^2 values (Table 3). However, the LSTM model was outperformed among all models on both the datasets but the lowest RMSE values (0.345) were observed with lagged dataset. Moreover, it was also shown that lagged meteorological factors dataset (DS2) has the lowest RMSE and high R^2 values compared to DS1 dataset across different models. Next to the LSTM, SVR model has shown better performance than the other models. Both GBM and VAR models captured the trend of dengue cases but failed in predicting the outbreak periods when compared to LSTM and SVR models.

Discussion

The novel analytical methods such as statistical, machine learning, and deep learning approaches and increasing of data availability on infectious diseases have led to growth in developing disease forecasting systems. Accurate disease-forecasting models would improve epidemic prevention and control capabilities. These forecast models help the policy decisions, public health authorities, and management of resources effectively during the disease outbreak. Dengue is a global public health problem that can affect the millions of people every year. Hence, prediction of dengue epidemics is a valuable objective that can help in preparing health-care system and implementing appropriate vector control interventions much in advance, and that could reduce the impact (Johansson et al. 2016; Alkhamis et al. 2018). In the present study, we have applied various statistical, machine and deep learning approaches to predict dengue prevalence in Kerala, India.

The prevalence of dengue in Kerala, India, was analyzed retrospectively from 2003 to 2017 and their spatial

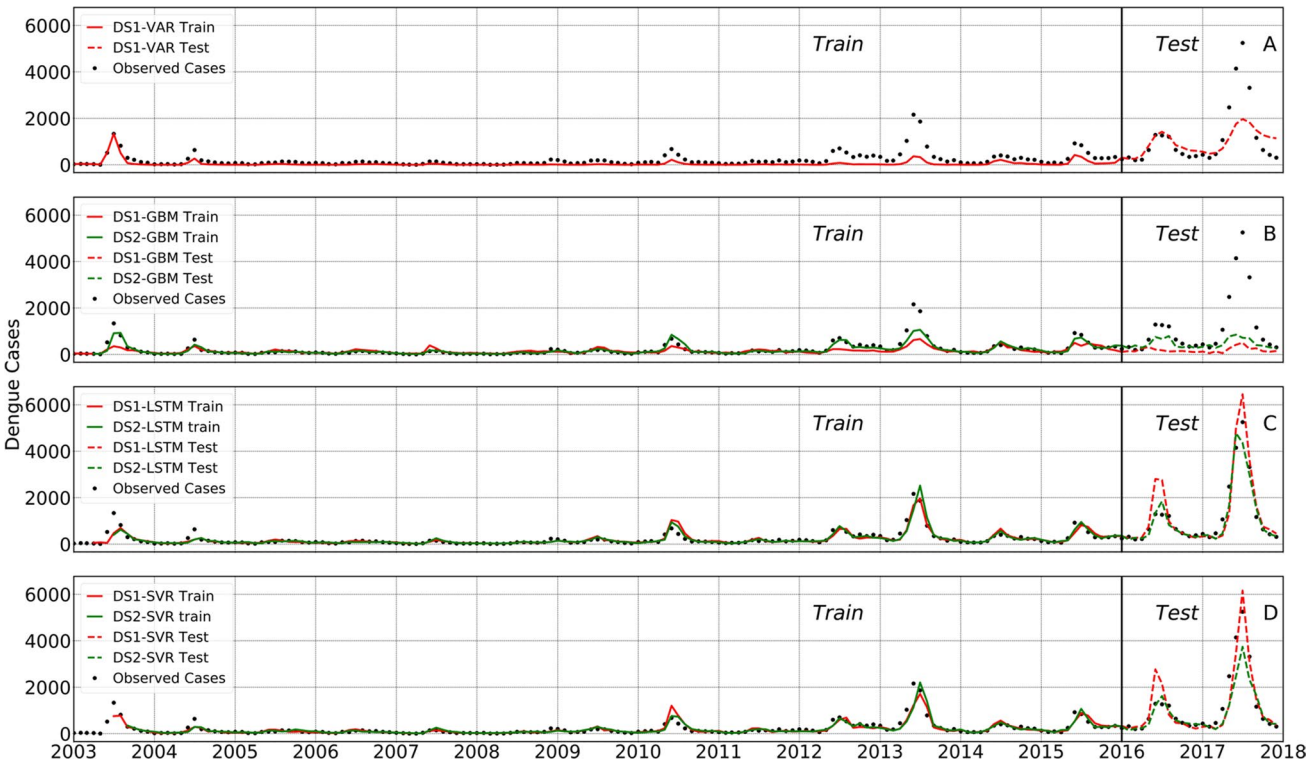

Fig. 5 Observed and predicted dengue cases in training (2003 January to 2015 December) and test data (2016 January to 2017 December) of different models. (A) VAR, (B) GBM, (C) LSTM, (D) SVR

distribution pattern reveals that the heterogeneity of cases was observed among the districts. The high dengue incidence rates were reported from the Thiruvananthapuram, followed by Kollam, Malappuram, Kozhikode, Ernakulam, and Thrissur district. The state Kerala is hyperendemic for dengue and it is also one of the leading states in deaths due to dengue infection (Karunakaran et al. 2014). Dengue was first time reported in Kerala in 1997 followed by the state has reported dengue epidemics during the years 2003, 2010, 2012, 2013, 2016, and 2017. During the year 2017, the state has reported the highest number of dengue cases. The district Thiruvananthapuram contributes nearly 40 to 50% total cases in the state and most of these cases are from the urban areas of the district (Anoop et al. 2010).

Kerala has experienced ample rainfall, seasonal variations in temperature, and high humidity throughout the year as it is located in the tropical wet climate region of the country. The weather conditions and urbanized environment favor the presence of *Aedes* mosquitoes and the transmission of dengue virus, thus making the region highly vulnerable to dengue (Reddy et al. 2017). During the correlation analysis between dengue cases and weather variables, rainfall, soil moisture, relative humidity, and Nino3.4 showed positive correlations, while mean temperature showed negative correlation with dengue. Similarly, various lag times between weather variables and dengue cases were analyzed to identify optimum lag periods including interactions of multiple lags. The mean temperature was shown high correlation at 3-month lag time in the selected model. These longer lag periods suggest that the *Aedes* mosquito eggs can survive for several months in dry containers and long egg hatching periods (Withanage et al. 2018). However, the rainfall and relative humidity have shown the zero lag association with dengue cases. The rainfall provides ample breeding grounds for *Aedes* mosquito, whereas, heavy rainfall can flush out breeding sites (Jeelani & Sabesan 2013). Similarly, relative humidity favors the survival rate and biting activity of adult mosquito. This allows the infected female *Aedes* mosquitoes to complete more than one replication cycle of the virus (Donalisio & Glasser 2002).

Statistical regression, ARIMA and SARIMA are most commonly used models for prediction of dengue, malaria in various geographic regions (Hii et al. 2012; Bhatnagar et al. 2012; Gharbi et al. 2011; Mopuri et al. 2020). In recent years, data-driven techniques based on machine learning and deep learning techniques have gained importance in predictive analytics (Kondeti et al. 2019; Bhimala et al. 2021). The machine learning techniques such as K-nearest neighbour, artificial neural network, support vector machine, naive Bayes, random forest, and gradient boosting are frequently used machine learning approaches in predicting dengue prevalence (Guo et al. 2017; Althouse et al. 2011; Gambhir et al. 2018; Laureano-Rosario et al. 2018; Scavuzzo et al. 2018). In the present study along with statistical (VAR), machine learning techniques (SVR, GBM) most advanced deep learning technique (LSTM) were applied for dengue prediction. LSTM was successfully utilised for prediction of Influenza, epidemics of hand, foot, and mouth disease, and COVID-19 (Bhimala et al. 2021; Liu et al. 2018; Wang et al. 2019).

The predictive models developed in this study were based on retrospective monthly rainfall, mean temperature, relative humidity, soil moisture, NINO3.4, and dengue cases. The study does not include any mosquito density data in the model as this data is unavailable with authors. Among all the developed models, the prediction accuracy of LSTM model is superior than the other models. Because the LSTM model learns temporal dependencies from the data and understand the situation based on past observations over time by incorporating memory state. The results show that the deep learning method LSTM model has the potential for predicting the cases and captured the epidemic trends with high accuracy than the VAR, GBM and SVR models. Among the two datasets (DS1 & DS2), the LSTM and SVR models were outperformed for predicting the dengue with lagged dataset (DS2). The RMSE of statistical, machine learning and deep learning models is also computed and it is evident that the LSTM (Train data: RMSE = 0.453, $R = 0.9$; Test data: RMSE = 0.345, $R = 0.9$) exhibits better performance (Table 3). Followed by, machine learning model, SVR (Train data: RMSE = 0.447, $R = 0.9$; Test data: RMSE = 0.41, $R = 0.9$) has shown better performance in predicting the dengue cases during 2016–2017 period.

The climate-based disease prediction models developed in the present study can be feasibly implemented as a dengue early warning system. The incorporation of real time climate forecasts and the current dengue case data will drive the dengue early warning system which is formulated using the machine learning framework to produce a dengue forecast for short term and long-term periods well in advance. Similarly, establishing a web version of these developed models and integrating with the local meteorological department on the real-time climate and disease data can enable the dengue early warning system to estimate the climate sensitive disease-risk timely and accurately, thus assisting the public health authorities to promptly target the risk areas.

Conclusion

Infectious diseases cause a significant burden on public health and economic stability of societies globally. The aim of this study was to design an infectious disease prediction model that is more suitable than existing models by using various statistical, machine learning and deep learning techniques. A very little or scared information is available on

dengue predictions in India using machine or deep learning applications. Hence, the present study utilized machine learning (SVR, GBM) and deep learning (LSTM) models along with statistical model (VAR), for dengue prediction. Among all the models, the LSTM model was outperformed (RMSE = 0.345, $R = 0.9$) with lagged dataset and predicted monthly dengue cases using weather variables and dengue cases in Kerala state. These predictions will be helpful for public health authorities, researchers, and planners for managing services and arranging medical infrastructure accordingly. The proposed LSTM model for dengue prediction can be followed by other states of India as well.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00484-022-02405-z.

Acknowledgements The authors are grateful to the Directors of the Council of Scientific and Industrial Research-Indian Institute of Chemical Technology, Hyderabad, for his encouragement and support. Srinivasa Rao Mutheneni acknowledges the Department of Science and Technology (DST/SSTP/Telangana/11/2017-18 (G)) and the Ministry of Environment, Forest & Climate Change (MoEF& CC), Government of India, for funding the project environmental information system (ENVIS: Resource Partner on Climate Change and Public Health). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript. CSIR-IICT communication number of the article is IICT/Pubs./2020/358.

Data availability The data used in this study are available from the corresponding author upon request.

Declarations

Ethical statement The authors declare that an ethical statement is not applicable because the case information has been gathered.

Conflict of interest The authors declare no competing interests.

References

Alkhamis MA, Brookes VJ, VanderWaal K (2018) Editorial: applications of novel analytical methods in epidemiology. Front in Vet Sci 5:243. https://doi.org/10.3389/fvets.2018.00243

Althouse BM, Ng YY, Cummings DAT (2011) Prediction of dengue incidence using serach query surveillance. PLoS Negl Trop Dis 5:e1258

Anoop M, Issac A, Mathew T, Philip S, Kareem NA, Unnikrishnan R, Sreekumar E (2010) Genetic characterization of dengue virus serotypes causing concurrent infection in an outbreak in Ernakulam, Kerala, South India. Indian J Exp Biol 48:849–857

Arcari P, Tapper N, Pfueller S (2007) Regional variability in relationships between climate and dengue/DHF in Indonesia. Singap J Trop Geogr 28:251–272

Bal S, Sodoudi S (2020) Modeling and prediction of dengue occurrences in Kolkata, India, based on climate factors. Int J Biometeorol 64:1379–1391. https://doi.org/10.1007/s00484-020-01918-9

Bhatnagar S, Lal V, Gupta SD, Gupta OP (2012) Forecasting incidence of dengue in Rajasthan, using time series analyses. Indian J Public Health 56:281–285. https://doi.org/10.4103/0019-557X.106415

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF, George DB, Jaenisch T, Wint GR, Simmons CP, Scott TW, Farrar JJ, Hay SI (2013) The global distribution and burden of dengue. Nature 496:504–507. https://doi.org/10.1038/nature12060

Bhimala KR, Patra GK, Mopuri R, Mutheneni SR (2021) Prediction of COVID-19 cases using the weather integrated deep learning approach for India. Transbound Emerg Dis. https://doi.org/10.1111/tbed.14102.Advanceonlinepublication.10.1111/tbed.14102

Carvajal TM, Viacrusis KM, Hernandez L, Ho HT, Amalin DM, Watanabe K (2018) Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila Philippines. BMC Infect Dis 18:183. https://doi.org/10.1186/s12879-018-3066-0

Choi JY, Lee B (2018) Combining LSTM network ensemble via adaptive weighting for improved time series forecasting. Math Probl Eng 2470171. https://doi.org/10.1155/2018/2470171

Chris B (2019). Introductory Econometrics for Finance. Cambridge, Cambridge University Press. https://doi.org/10.1017/9781108524872

Donalisio MR, Glasser CM (2002) (2002) Vigilancia entomologica e controle de vetores do dengue. Rev Bras Epidemiol 5:259–272

Gambhir S, Malik SK, Kumar Y (2018) The diagnosis of dengue disease: An evaluation of three machine learning approaches. Int J Healthc Inf Syst Inform 13:1–19

Gharbi M, Quenel P, Gustave J, Cassadou S, La Ruche G, Girdary L, Marrama L (2011) Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. BMC Infect Dis 11:166. https://doi.org/10.1186/1471-2334-11-166

Goebel R, Roebroeck A, Kim DS, Formisano E (2003) Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magn Reson Imaging 21:1251–1261. https://doi.org/10.1016/j.mri.2003.08.026

Greg R (2007) Generalized boosted models: a guide to the gbm package. https://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf. Acessed 18 Jan 2019

Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, Luo G, Li Z, He J, Zhang Y, Ma W (2017) Developing a dengue forecast model using machine learning: A case study in China. PLoS Negl Trop Dis 11:e0005973. https://doi.org/10.1371/journal.pntd.0005973

Harris D, Chris JCB, Linda K, Alex S, Vladimir V (1997) Support vector regression machines: Advances in neural information processing systems 9: 155–161. MIT Press

Hii YL, Zhu H, Ng N, Ng LC, Rocklöv J (2012) Forecast of dengue incidence using temperature and rainfall. PLoS Negl Trop Dis 6:e1908. https://doi.org/10.1371/journal.pntd.0001908

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural. Computing 9:1735–1780

Jayadeva Khemchandani R, Chandra S (2007) Twin Support Vector Machines for pattern classification. IEEE Trans Pattern Anal Mach Intell 29:905–910. https://doi.org/10.1109/tpami.2007.1068

Jeelani S, Sabesan S (2013) *Aedes* vector population dynamics and occurrence of dengue fever in relation to climate variables in Puducherry, South India. Int J Curr Microbiol App Sci 2:313–322

Jentes ES, Lash RR, Johansson MA, Sharp TM, Henry R, Brady OJ, Sotir MJ, Hay SI, Margolis HS, Brunette GW (2016) Evidence-based risk assessment and communication: a new global dengue-risk map for travellers and clinicians. Journal of travel medicine 23:taw062. https://doi.org/10.1093/jtm/taw062

Johansson MA, Reich NG, Hota A, Brownstein JS, Santillana M (2016) Evaluating the performance of infectious disease forecasts: A comparison of climate-driven and seasonal dengue forecasts for Mexico. Sci Rep 6:33707. https://doi.org/10.1038/srep33707

Kakarla SG, Bhimala KR, Kadiri MR, Kumaraswamy S, Mutheneni SR (2020) Dengue situation in India: Suitability and transmission potential model for present and projected climate change scenarios. Sci Total Environ 739:140336. https://doi.org/10.1016/j.scitotenv.2020.140336

Karunakaran A, Ilyas WM, Sheen SF, Jose NK, Nujum ZT (2014) Risk factors of mortality among dengue patients admitted to a tertiary care setting in Kerala, India. J Infect Public Health 7:114–120. https://doi.org/10.1016/j.jiph.2013.09.006

Kondeti PK, Ravi K, Mutheneni SR, Kadiri MR, Kumaraswamy S, Vadlamani R, Upadhyayula SM (2019) Applications of machine learning techniques to predict filariasis using socio-economic factors. Epidemiol Infect 147:e260. https://doi.org/10.1017/S0950268819001481

Laureano-Rosario AE, Duncan AP, Mendez-Lazaro PA, Garcia-Rejon JE, Gomez-Carro S, Farfan-Ale J, Savic DA, Muller-Karger FE (2018) Application of Artificial Neural Networks for Dengue Fever Outbreak Predictions in the Northwest Coast of Yucatan, Mexico and San Juan, Puerto Rico. Trop Med Infect Dis 3(1):5. https://doi.org/10.3390/tropicalmed3010005

Liu L, Han M, Zhou Y, Wang Y (2018) LSTM Recurrent Neural Networks for Influenza Trends Prediction. Bioinform Res 10847:259–264

Lütkepohl H (2005) New introduction to multiple time series analysis. Springer-Verlag Berlin Heidelberg (1 ed.). https://doi.org/10.1007/078-3-540-27752-1

Mala S, Jat MK (2019) Implications of meteorological and physiographical parameters on dengue fever occurrences in Delhi. Sci Total Environ 650:2267–2283. https://doi.org/10.1016/j.scitotenv.2018.09.357

Mopuri R, Kakarla SG, Mutheneni SR, Kadiri MR, Kumaraswamy S (2020) Climate based malaria forecasting system for Andhra Pradesh. India J Parasit Dis 44(3):497–510. https://doi.org/10.1007/s12639-020-01216-6

Mutheneni SR, Morse AP, Caminade C, Upadhyayula SM (2017) Dengue burden in India: recent trends and importance of climatic parameters. Emerg Microbes Infect 6:e70. https://doi.org/10.1038/emi.2017.57

Ong J, Liu X, Rajarethinam J, Kok SY, Liang S, Tang CS, Cook AR, Ng LC, Yap G (2018) Mapping dengue risk in Singapore using Random Forest. PLoS Negl Trop Dis 12:e0006587. https://doi.org/10.1371/journal.pntd.0006587

Reddy MN, Dungdung R, Valliyott L, Pilankatta R (2017) Occurrence of concurrent infections with multiple serotypes of dengue viruses during 2013–2015 in northern Kerala. India Peerj 5:e2970. https://doi.org/10.7717/peerj.2970

Salami D, Sousa CA, Martins M, Capinha C (2020) Predicting dengue importation into Europe, using machine learning and model-agnostic methods. Sci Rep 10:9689. https://doi.org/10.1038/s41598-020-66650-1

Scavuzzo JM, Trucco F, Espinosa M, Tauro CB, Abril M, Scavuzzo CM, Frery AC (2018) Modeling Dengue vector population using remotely sensed data and machine learning. Acta Trop 185:167–175. https://doi.org/10.1016/j.actatropica.2018.05.003

Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222

Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, Hay SI, Bedi N, Bensenor IM, Castañeda-Orjuela CA, Chuang TW, Gibney KB, Memish ZA, Rafay A, Ukwaja KN, Yonemoto N, Murray C (2016) The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. Lancet Infect Dis 16:712–723. https://doi.org/10.1016/S1473-3099(16)00026-8

Thammapalo S, Chongsuvivatwong V, Geater A, Dueravee M (2008) Environmental factors and incidence of dengue fever and dengue haemorrhagic fever in an urban area, Southern Thailand. Epidemiol Infect 136:135–143. https://doi.org/10.1017/S0950268807008126

Thomas SJ, Yoon IK (2019) A review of Dengvaxia®: development to deployment Hum. Vaccin Immunother 15:2295–2314. https://doi.org/10.1080/21645515.2019.1658503

Trpis M, McClelland GA, Gillett JD, Teesdale C, Rao TR (1973) Diel periodicity in the landing of *Aedes aegypti* on man. Bull World Health Organ 48:623–629

Wang YB, Xu CJ, Zhang SK, Yang L, Wang ZD, Zhu Y, Yuan JX (2019) Development and evaluation of a deep learning approach for modeling seasonality and trends in hand-foot-mouth disease incidence in mainland China. Sci Rep 9:8046

Wang Y, Zhou J, Chen K, Wang Y, Liu L (2017) Water quality prediction method based on LSTM neural network. 12th International Conference on Intelligent Systems and Knowledge Engineering, Nanjing 1–5. https://doi.org/10.1109/ISKE.2017.8258814

Wilder-Smith A, Rupali P (2019) Estimating the dengue burden in India. Lancet Glob Health 7:e988–e989. https://doi.org/10.1016/S2214-109X(19)30249-9

Withanage GP, Viswakula SD, Nilmini Silva Gunawardena YI, Hapugoda MD (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 (WHO) Situation of dengue/dengue haemorrhagic fever in South-East Asia region: World Health Organization, 2007. Available at http://209.61.208.233/en/Section10/Section332_1098.htm. Accessed 20 March 2012

World Health Organisation (WHO) https://www.who.int/news-room/fact-sheets/detail/dengueandseveredengue#:~:text=The%20number%20of%20dengue%20cases,and%204.2%20million%20in%202019. Accessed on 17.06.2020