



Original Article

A Machine Learning Tool with An Integrated Dataset Towards the Construction of An Early Warning System for Dengue in Zulia State, Venezuela

Máquina de aprendizaje con conjunto de datos integrado para el desarrollo de un sistema de alerta temprana de dengue en el estado de Zulia, Venezuela

DOI

<https://doi.org/10.35434/rcmhnaaa.2025.182.2815>Cabrera Maritza^{1,a}, Naranjo-Torres José^{2,b}, Cabrera Ángel^{3,c},
Zambrano Lysien^{4,d}, Rodríguez-Morales Alfonso J.^{5,6,e*}

ABSTRACT

Background: The present study arises in response to the sustained increase in dengue outbreaks in Latin America, with special emphasis on the state of Zulia, Venezuela. This region, composed of 21 municipalities, is highly vulnerable to dengue transmission. Given this scenario, it is essential to have tools that allow early detection of outbreaks and, thus, optimize prevention and public health intervention strategies. The main objective is to develop an early warning system for dengue outbreaks using machine learning (ML) techniques. **Materials and methods:** Several data sources are integrated: epidemiological information, meteorological parameters, El Niño and La Niña (Niño 3.4 Index), socioeconomic and demographic variables. Two ML models were used: Support Vector Machine for regression (SVM-R) and Gaussian Process Regression (GPR). **Results:** The predictions obtained showed remarkable agreement with the actual dates on which the outbreaks were recorded, warning of the onset of dengue 2 to 3 weeks in advance, depending on the locality. However, in certain municipalities the predictions were less accurate, a finding that agrees with previous studies. **Conclusions:** In conclusion, the integration of epidemiological, climatological, and socioeconomic variables using ML techniques is presented as a promising tool for establishing.

Keywords: Dengue fever; Machine Learning; Epidemiology; Disease Outbreaks (Source: DeCS-BIREME).

RESUMEN

Introducción: El presente estudio surge como respuesta al sostenido aumento de los brotes de dengue en América Latina, con especial énfasis en el estado Zulia, Venezuela. Esta región, compuesta por 21 municipios, enfrenta una alta vulnerabilidad frente a la transmisión del dengue. Ante este escenario, resulta fundamental contar con herramientas que permitan la detección temprana de los brotes y, de esta forma, optimizar las estrategias de prevención e intervención en salud pública. El objetivo principal es desarrollar un sistema de alerta temprana para brotes de dengue utilizando técnicas de machine learning (ML). **Material y métodos:** Se integran diversas fuentes de datos: información epidemiológica, parámetros meteorológicos, El Niño and La Niña (Niño 3.4 Índice), variables socioeconómicas y demográficas. Se emplearon dos modelos de ML: Support Vector Machine para regresión (SVM-R) y el Gaussian Process Regression (GPR). **Resultados:** Las predicciones obtenidas mostraron una notable concordancia con las fechas reales en que se registraron los brotes, advirtiendo la aparición del dengue con una anticipación de 2 a 3 semanas, dependiendo de la localidad. Sin embargo, en ciertos municipios las predicciones fueron menos precisas, hallazgo que concuerda con estudios previos. **Conclusiones:** En

FILIIATION

1. Centro de Investigación de Estudios Avanzados del Maule (CIEAM), Universidad Católica del Maule, Talca 3480094, Chile.
2. Academic and ML Consulting Department, Global Consulting H&G, 8682 Sorrento Street, Orlando, FL 32819, USA.
3. Facultad de Ingeniería, Universidad Fermín Toro, Calle Principal C.C. Chucho Briceño Cabudare, Lara, Venezuela.
4. Department of Morphological Sciences, School of Medical, Sciences, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras.
5. Master of Clinical Epidemiology and Biostatistics, Universidad Científica del Sur, Lima, 15067, Peru.
6. Grupo de Investigación Biomedicina, Faculty of Medicine, Fundación Universitaria Autónoma de las Américas-Institución Universitaria Visión de las Américas, Pereira, Colombia.
 - a. PhD in Spatial Epidemiology.
 - b. Doctor in Science and Technology.
 - c. Engineer.
 - d. MD, MSc, PhD.
 - e. MD, MSc, HonDSC.

ORCID

1. Cabrera Maritza / [0000-0001-7279-9921](https://orcid.org/0000-0001-7279-9921)
Naranjo-Torres José / [0000-0002-3653-9601](https://orcid.org/0000-0002-3653-9601)
Cabrera Ángel / [0009-0003-6216-895X](https://orcid.org/0009-0003-6216-895X)
Zambrano Lysien / [0000-0001-9002-5807](https://orcid.org/0000-0001-9002-5807)
Rodríguez-Morales Alfonso J. / [0000-0001-9773-2192](https://orcid.org/0000-0001-9773-2192)

CORRESPONDENCE

Rodríguez-Morales Alfonso J.
Tel: +573008847448.

EMAIL

arodriguezmo@cientifica.edu.pe

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FINANCING

The authors declare that they have not received any funding.

AUTORS CONTRIBUTIONS

MC: Conceptualization, investigation, methodology, Writing - review & editing, supervision, Data curation, Formal analysis. JNT: investigation, methodology, Writing - review & editing, supervision, Software, Formal analysis. AC: Formal analysis. LZ: Writing - review & editing, supervision, Software, Formal analysis. AJRM: Writing - review & editing, supervision, Software, Formal analysis.

ACKNOWLEDGMENTS

Dr Rodríguez-Morales would like to dedicate this publication to the memory of Prof. Olinda Delgado, PhD (1930-2024), an expert in tropical and parasitic diseases such as leishmaniasis and toxocaríasis, from Venezuela, his major mentor and excellent human being, who passed away in Caracas, December 2024, R.I.P. This article has been registered in the Research Proposal Registration of the Coordination of Scientific Integrity and Surveillance of Universidad Científica del Sur, Lima, Peru, under the number PI-50-2025-0237.



This work is under a Licencia Creative Commons Atribución 4.0 Internacional.
Printer Version: ISSN: 2225-5109
Electronic Version: ISSN: 2227-4731
Cross Ref. DOI: 10.35434/rcmhnaaa
OJS: <https://cmhnaaa.org.pe/ojs>

conclusión, la integración de variables epidemiológicas, climatológicas y socioeconómicas mediante técnicas de ML se presenta como una herramienta prometedora para el establecimiento de sistemas de alerta temprana, aunque se recomienda continuar afinando los modelos de manera específica según las particularidades de cada municipio

Palabras Clave: Fiebre de Dengue; Aprendizaje automático; Epidemiología; Brotes de enfermedades. (Fuente: DeCS-BIREME).

PEER REVIEW

Received: 16/03/2025
Accepted: 10/07/2025

HOW TO CITE

Cabrera M, Naranjo-Torres J, Cabrera Ángel, Zambrano L, Rodríguez-Morales AJ. A Machine Learning Tool with An Integrated Dataset Towards the Construction of An Early Warning System for Dengue in Zulia State, Venezuela. Rev. Cuerpo Med. HNAAA [Internet]. 1 de agosto de 2025 [citado 28 de septiembre de 2025];18(2):2815e. DOI: 10.35434/rcmhnaaa.2025.182.2815

INTRODUCTION

Dengue fever is a vector-borne disease that has become endemic in tropical and sub-tropical regions worldwide over the last century and now poses a threat to half of the world’s population⁽¹⁾. *Aedes* mosquitoes, the species responsible for the dengue transmission, are now found in more than 130 countries, making dengue the single most significant human disease burden of any arbovirus with around 10.000 deaths and more than 100 million symptomatic infections per year^(1,2). This grim global scenario is exacerbated by other *Aedes*-like species reported in many parts of the world, coupled with factors such as migratory displacement, political conflict, climate change, unplanned urbanization, and inefficient public services⁽²⁾.

In this context, the Neglected Tropical Diseases (NTDs) initiative, established as part of the United Nations Sustainable Development Goals (SDGs) in 2015 to address the global challenges of dengue fever and other infectious diseases, emphasizes the necessity of a multidisciplinary approach to dengue eradication, requiring an unprecedented level of cooperation and collaboration between different sectors⁽³⁾.

In the case of dengue in Venezuela, it has shown sustained increases, both in the interannual magnitude of cases and in the frequency of outbreaks⁽⁴⁾. In this regard, six major dengue epidemics were documented in Venezuela between 2007 and 2016⁽⁵⁾. The largest was in 2010, the third largest ever recorded in the Americas, with a peak of about 1,000,000 cases, of which more than 10,000 were severe⁽⁵⁾. Zulia state

is one of the most affected regions, as shown in Fig. 1. In this regard, recent spatio-temporal models of dengue in Zulia state using Generalized Additive Mixed models⁽⁶⁾ suggest the need for urgent intervention to improve the inhabitants’ living conditions and reduce the growing escalation of dengue in that part of the country.

Likewise, the World Health Organization (WHO) has stated that the prediction of dengue outbreaks is one of the main objectives for effective public global health^(8,9); and Machine Learning (ML) techniques are a promising approach for epidemiological studies⁽¹⁰⁾, that can provide for knowledge of dengue outbreaks, with the aim of the development of Early Warning Systems (EWS) to control the disease^(8,10-14).

Based on the above, the main objective of this study is to evaluate the possibility of creating an Early Warning System for Dengue Prevention in Venezuela appropriate for the local environment of Zulia State using climatic data in conjunction with socioeconomic and demographic data. Using the algorithms Support Vector Regression (SVR), which is an extension of Support Vector Machines (SVM) applied to regression problems^(15,16); and the Gaussian Process Regression (GPR) is a non-parametric supervised learning method used to solve regression and probabilistic classification problems^(17,18).

Even though there are others advanced algorithms currently available for predictive purposes using ML approach, this paper is focused on the implementation of Support Vector Machine (SVM) and Gaussian Process Regression (GPR) in predicting future dengue outbreak in Zulia State Venezuela.

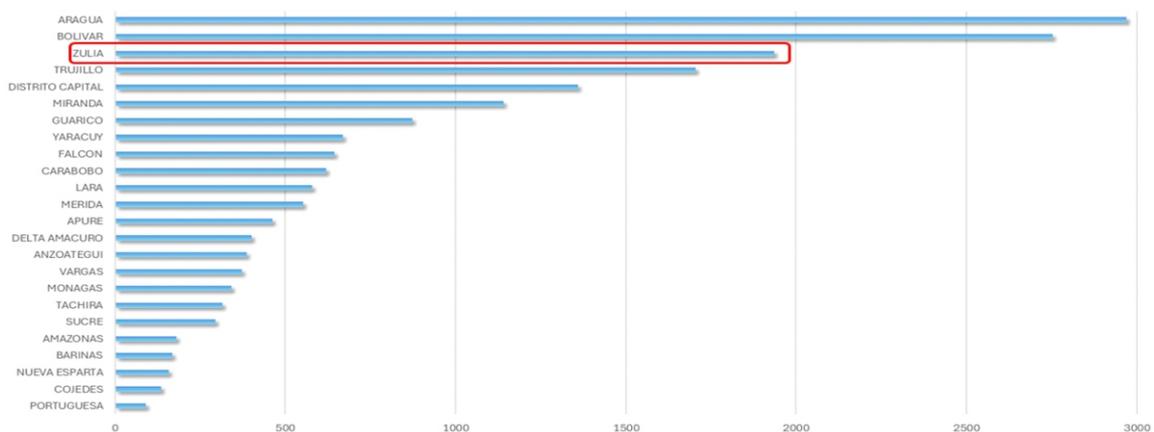


Figure 1.

Total dengue cases in Venezuela by state for the year 2018 Zulia State (7). It is observed that the Zulia state is the third most affected in the country.

A study conducted in Jakarta, Indonesia⁽¹⁹⁾, benefited from this algorithm by arguing a better interpretation of the results to the community including, governments, health officials and policymakers. This interpretability facilitated better communication for strategic planning and initiative identification in dengue fever prevention. The authors emphasized the simplicity on data scaling to apply on a range of features before being fed into the training model. In this regard, SVM models can be used based on different kernel functions including Gaussian radial basis function (RBF), polynomial function and linear function according to the aforementioned researchers.

Moreover, the Gaussian Process Regression (GPR) is implemented in a wide range of artificial intelligence problems due to its ability to achieve a fine and precise balance between data fitting and smoothing⁽²⁰⁾. Another relevant benefit of this method is its efficiency when dealing with a large dataset and assessing the impact of uncertainties⁽²⁰⁾.

Preliminaries

The typical of these studies is that the authors use climate variables as predictors of dengue incidence using the Generalized Additive Model for Location, Scale, and Shape (GAMLSS) and Random Forest (RF). In this⁽⁸⁾, the authors applied Vector Auto Regression (VAR), Generalized Boosted models (GBM), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) to predict dengue prevalence using meteorological data. In⁽¹⁴⁾, the authors used climate data to train a Support Vector Machine (SVM) classifier with Radial Basis Function (RBF) kernel; In⁽²¹⁾, the authors compared LSTM time series forecasting for dengue prediction with Support Vector Regression (SVR).

A study performed in China⁽²²⁾ For instance, weekly data from 2011 to 2014 were used in conjunction with meteorological factors with appropriated delayed effects plus Internet search engine query statistics to construct predictive models of dengue. Several models were tried to compare and identify the optimal performance through the Goodness of Fit validation of Root Mean-Square Error (RMSE) and R-squared measures.

In⁽²³⁾ the authors perform a systematic review of ML techniques applied to the study of dengue, creating a list of the most common ML regression models for predictive applications. Similarly, in⁽²¹⁾ a review of ML models was carried out to estimate possible dengue outbreaks. The diagram in Figure. 2 shows some of the most used ML models for dengue outbreak forecast studies.

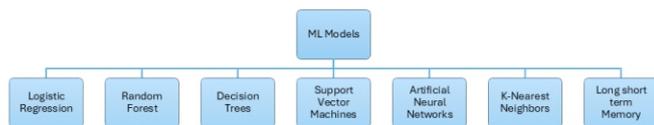


Figure 2.

ML models are most frequently used to conduct studies to predict dengue outbreaks

The study performed in China⁽²²⁾ considered a delayed climatic factor effect identified through a Cross-Correlation

Analysis, the obtained results were: The Nino 3.4 with 2 months lag; mean temperature at 3 months lag; soil moisture at 1-month lag and both rainfall and relative humidity with a 0-month lag.

Another study performed in India⁽⁸⁾ confirmed those values but also found the most significant lagged terms of the Nino 3.4: mean temperature with a 2-month lag, soil moisture with a 3-month lag, and rainfall and relative humidity with a zero-month lag, respectively.

MATERIALS AND METHODS

Study area

This study used data related to Zulia state, which is located in the northwest part of Venezuela, between 8.20° to 11.79° North latitude and between 70.73° and 73.37° West longitude⁽⁶⁾. The state is divided into 21 municipalities, covering 50,230 square kilometers around Lake Maracaibo. These municipalities (Figure. 3) are: Almirante Padilla, Baralt, Cabimas, Catatumbo, Colon, Francisco Javier Pulgar, Jesús Enrique Lossada, Jesús María Semprún, La Cañada de Urdaneta, Lagunillas, Machiques de Perijá, Mara, Maracaibo, Miranda, Indígena Bolivariano Guajira, Rosario de Perijá, San Francisco, Santa Rita, Simón Bolívar, Sucre and Valmore Rodríguez.

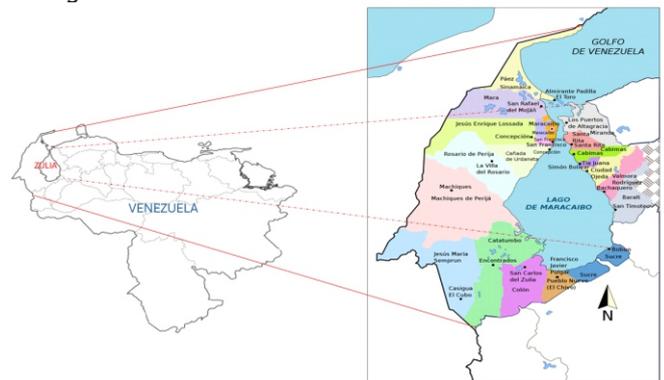


Figure 3.

Zulia State municipalities, Venezuela

Data Integration

This study integrated a series of data (Figure. 4) from various sources, such as epidemiological data, socioeconomic and demographic data, and local climatic data. The characteristics of these data are presented below.

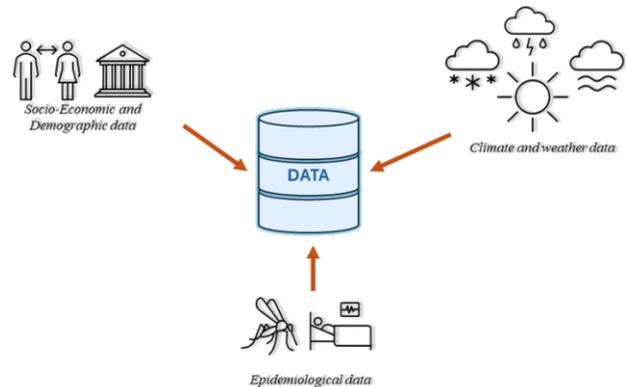


Figure 4.

Data integration

Epidemiological data

Hospital admissions of dengue fever were provided by the Venezuelan Ministry of Health as weekly reported dengue fever cases from January 2008 to December 2016. A case was operationally defined as hospital admission, regardless of disease severity. This study is similar to a previous one performed in Zulia State by Cabrera M & Taylor G⁽⁶⁾, framed on an overall dengue admission rate and range across the Zulia state municipalities.

Climate and weather data

Meteorological data from Zulia state was obtained from NASA Earth Science's Applied Program's POWER (Prediction Of Worldwide Energy Resources) project⁽²⁴⁾. This project provides NASA's solar and meteorological data sets to support renewable energy schemes, building energy efficiency, and agricultural needs. Meteorological parameters are achieved from NASA's GMAO MERRA-2 assimilation model plus GEOS-5.12.4 FP-IT. The Data Access Viewer was used to search the geographical coordinates across the zone: -74° West to -70° East and 12° North to 8° South (Lat 10.6417 Long -71.6295)⁽²⁴⁾ using the following meteorological parameters:

- Temperature at 2 Meters Maximum (°C)
- Temperature at 2 Meters Minimum (°C)
- Specific Humidity at 2 Meters (g/kg)
- Relative Humidity at 2 Meters (%)
- Precipitation Corrected (mm/day)
- Wind Speed at 2 Meters (m/s)
- Wind Speed at 2 Meters Maximum (m/s)

Since meteorological conditions in Venezuela are heavily influenced by El Niño, producing warmer temperatures and reduced precipitation, and La Niña with the opposite effect - cooler temperatures and more significant rainfall⁽⁴⁾ we used

the Niño3.4 Index, following earlier Venezuelan studies^(4,6). The Niño 3.4 index was obtained from the National Oceanic and Atmospheric Administration (NOAA)⁽²⁵⁾.

Socio-Economic and Demographic data

While multiple factors contribute to dengue incidence, the collection of socioeconomic data in the Zulia state is restricted by the limited information recorded in the 2011 census conducted by the National Institute of Statistics (INE)⁽²⁶⁾ due to the lack of available data for the country. In this regard, the socioeconomic variables within this category were aggregated at the municipality level in Zulia, as provided by the 2011 Census. These variables included the proportion of households living in poverty and the proportion of access to piped water supply.

Regarding demographic data, the annual population figures were obtained from INE⁽²⁶⁾, also based on the Venezuelan National Census of 2011. These figures were used to address gaps in demographic data from 2008 to 2016, aggregated at the municipality level. These gaps arose due to the absence of formal population surveys conducted by the Venezuelan national government since 2011. Therefore, the most recent population size data available for the municipalities in Zulia state is from 2011. While this decision may limit the results, it represents the most accurate approximation.

Machine Learning Algorithms

This study decided to use the Support Vector Regression (SVR) algorithm, fundamentally the Support Vector Machine (SVM) algorithm for regression, since it is one of the most widely used algorithms in Dengue studies, as previously mentioned. It was decided to use Gaussian process regression (GPR) because only a few previously published studies using this

Table 1. Common kernel functions for SVM-R implementation: (i) The simplest kernel function (linear kernel); (ii) Non-linear classical kernel function (polynomial kernel) and a powerful kernel (radial basis function RBF).

Kernel	Mathematical Function	Reference
Linear	$k(x, y) = x \cdot y + C$	(Kesorn, K. et al. 2015)
Polynomial	$k(x, y) = (x \cdot y)^d$	(Kesorn, K. et al. 2015; Mello-Roman, J., et al. 2019)
Radial basis function (RBF)	$k(x, y) = \left[\frac{-\ x - y\ ^2}{2\sigma^2} \right]$	(Kesorn, K. et al. 2015; Nordin, N. I. et al., 2020).

According to specialists, no efficient, structured method for selecting the optimal hyperparameters exists. Therefore, a simple approach optimizes one parameter at a time⁽¹⁵⁾. Hence, the evaluation of parameters of σ^2 for instance, with a fixed parameter C at value 1.0, through a wide range of (typically) exponential forms: $2^{(-8)}$; $2^{(-7)}$; $2^{(-6)}$... 2^1 is tried. Once the optimal value of σ^2 has been found, the optimal value of parameter C is determined⁽¹⁵⁾.

Gaussian Process Regression.

Gaussian process regression (GPR)⁽¹⁸⁾ is a regression approach gaining significant attention in Machine Learning⁽²⁹⁻³¹⁾, due to its nonparametric and Bayesian characteristics. This

methodology has shown remarkable performance with small datasets and its ability to provide uncertainty measurements on the predicted outcomes. In contrast to many widely used supervised Machine Learning algorithms that strive to learn precise values for each parameter in a given function, the Bayesian approach takes a different path by inferring a probability distribution across all potential values.

In essence, a Gaussian Process (GP) is defined by its mean function and covariance function (kernel). The covariance function, often parameterized by hyperparameters such as signal standard deviation (σ_f) and characteristic length scale (l), captures the relationships between input points and

determines the similarity of their corresponding output values.

The GPR model explains the responses by means of latent variables and explicit basis functions that project the inputs into a higher dimensional feature space. When treating a model with noisy observations, the joint distribution of the training results and the new predictions follows a Gaussian distribution, whose parameters, mean and variance, are derived from the kernel function and the training data. This makes GPR a powerful tool for tasks where understanding the reliability of predictions is as important as the predictions themselves. The detailed mathematical formulation of Gaussian Process Regression (GPR) is provided in Appendix A.

The methodology applied in the present study used dengue cases in Zulia state weekly aggregated in conjunction with a set of socioeconomic and local and global climatic variables and based on previously implemented machine learning algorithms used for other parts of the world^(8,15,22). The construction of Support Vector Regression (SVR) and Gaussian process regression (GPR) as traditional ML algorithms were performed due to their flexibility, practicality and typically excellent performance. A set of scenarios was proposed in the models for comparative purposes using both climatic and non-climatic factors and respective validations to obtain the optimum model. Figure 5. shows the process implemented in this study.

Data Integration: Various preliminary steps were used to prepare the data before being entered into the machine learning algorithms:

- i. Weekly epidemiological data of dengue cases in Zulia state were aggregated at the municipal level in conjunction with a set of climatic and non-climatic covariants. In this context it was necessary to integrate the existing data because of the different sources of information (as proposed Cabrera M & Taylor G.⁽⁶⁾). The present study also utilised remote satellite climatic data obtained from NASA as described previously.
- ii. Epidemiological data was missing from the Guajira municipality between 2013 to 2016, which resulted in this municipality being excluded from the study.
- iii. To settle discrepancies between weekly and daily

availability datasets, specifically the years 2008 and 2014, it was determined that these years should be treated as covering 53 weeks. This decision addresses a calendrical incongruity: a standard year of 52 weeks accounts for 364 days, leaving an accumulated deficit of seven days across the period spanning 2008 to 2014. By assigning an additional week to these years, this deficit is absorbed, ensuring temporal alignment across the datasets. For example, raw weather data, which was initially acquired on a daily temporal resolution, underwent a transformation process whereby it was aggregated and averaged over seven-day intervals to achieve a weekly representation, thus facilitating consistent analysis with the weekly availability data.

- iv. According to some authors⁽¹⁵⁾, the data can be sensitive to extreme values. Therefore, in some cases, it is convenient to normalize or standardize the data. In this study, raw, standardized, and normalized data were used for each model to be trained. In this way, choose the best model obtained. In Standardization: the software centers and scales each column of the predictor data according to the mean and standard deviation of the column. In Normalization: it scales each column of the predictor data between -1 and 1.

Model Construction: Two different machine learning algorithms were used for comparative purposes in this study, chosen for their reported accuracy in dengue forecasting in other parts of the world^(8,15,22,32).

To conduct the experiments, *the MatLab Statistics and Machine Learning Toolbox*⁽³³⁾ was used, which provides a framework for designing and implementing ML algorithms, and applications. The experiments were carried out with a PC Laptop of the following characteristics: CPU Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, 12,0 GB RAM, GPU NVIDIA GeForce RTX2060 with Max-Q design.

Figure. 6 illustrates the process of training, building, and applying machine learning models, which encompasses the following steps:

Dataset Partition: The initial step involves dividing the entire dataset into two distinct subsets: 80% for training data, from

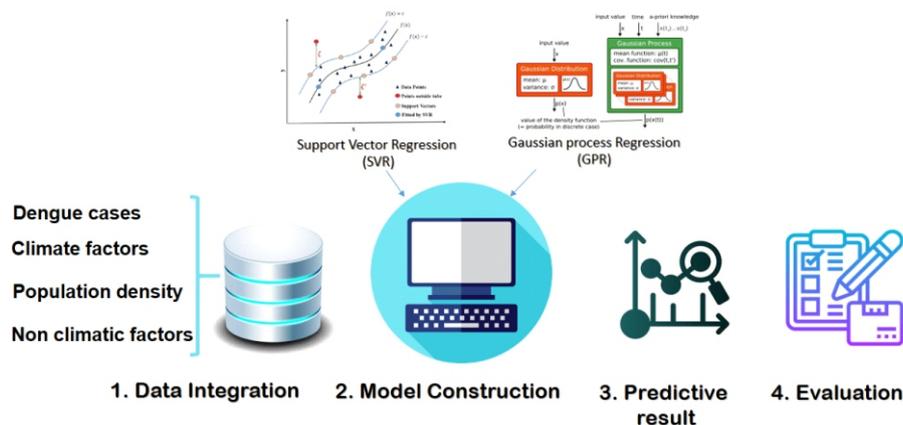


Figure 5.

Experimental processes for constructing a forecast tool for dengue cases in Zulia state, Venezuela based on the traditional kernel machines learning: SVR and Gaussian Processes due to their effectiveness, flexibility and ease of use.

which the model learns, and 20% for test data, utilized to assess the model's generalization capabilities on unseen data.

Model Training: The model training process employs 10-fold cross-validation on the training data. This internal procedure is conducted to ensure unbiased experimental results and to mitigate overfitting. The model undergoes "k" iterations of

training, each utilizing a unique fold, with the final performance representing the average of the "k" cross-evaluations.

Model Evaluation: Upon completion of model training, the model is evaluated using the test data set aside at the beginning of the process to determine its overall efficiency.

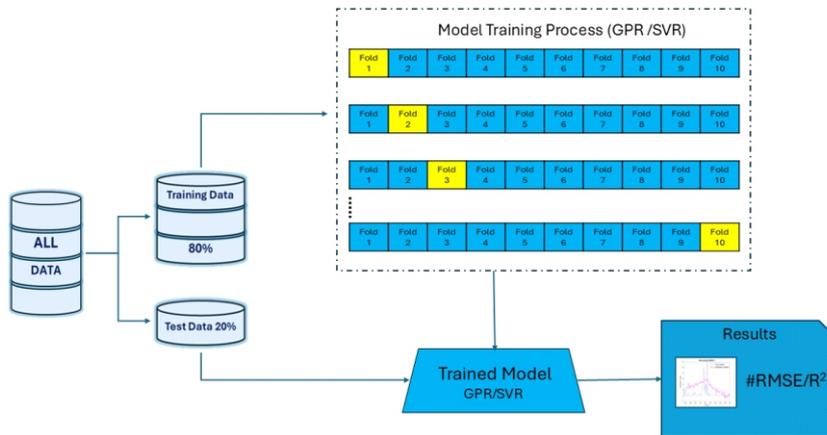


Figure 6. Stages of the training process and application of the ML models used in this study.

Fitting the GPR model requires estimating the following model parameters from the data:

- Covariance function $k(\theta)$ parameterized in terms of kernel parameters in vector θ
- Noise variance σ^2
- Coefficient vector of fixed-basis functions B

During this process of optimizing the hyperparameters of the GPR model, the system also performs a series of iterations, as shown in Figure. 7, testing different Base Functions, the kernel function and whether it standardizes the data, among other operations. This process is carried out until an optimal value is obtained.

Iter	Active workers	Eval result	Objective: log(1+loss)	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	Sigma	BasisFunction	KernelFunction	KernelScale	Standardize
1	5	Accept	1.5517	3.354	1.4867	1.4912	10.818	linear	matern32	0.090755	true
2	5	Best	1.4867	3.2914	1.4867	1.4912	0.56454	pureQuadrati	matern32	41.122	true
3	6	Accept	1.4867	2.4342	1.4867	1.4867	5.7709	pureQuadrati	matern32	50.429	true
4	6	Accept	1.5632	1.7391	1.4867	1.5221	0.51635	pureQuadrati	matern32	30.032	true
5	6	Accept	2.1174	5.7496	1.4867	1.4865	0.41938	pureQuadrati	matern52	16.751	true
6	6	Accept	1.4867	2.1992	1.4867	1.4855	8.0783	pureQuadrati	matern32	48.735	true
7	6	Accept	1.51	1.1062	1.4867	1.4855	0.060314	linear	matern32	0.28278	true
8	5	Best	1.4867	27.764	1.4867	1.5059	3.0377	pureQuadrati	ardsquaredex	-	true
9	5	Accept	1.5053	1.4225	1.4867	1.5059	0.0018455	linear	matern32	3.7534	true
10	5	Best	1.3767	0.93431	1.3767	1.3767	0.010457	linear	matern32	0.70933	true
11	5	Accept	1.4867	1.7586	1.3767	1.3767	16.427	pureQuadrati	exponential	51.415	true
12	5	Accept	1.4488	0.89435	1.3767	1.3767	1.7938	linear	matern32	1.303	true
13	5	Accept	1.3863	1.0968	1.3767	1.3767	0.00023305	pureQuadrati	exponential	12.23	true
14	6	Accept	1.6234	4.7015	1.3767	1.3976	0.00015077	linear	matern32	23.556	true
15	4	Best	1.2942	44.38	1.2942	1.296	1.8167	pureQuadrati	ardmatern32	-	false
16	4	Accept	1.4867	2.4143	1.2942	1.296	6.993	pureQuadrati	exponential	4.6754	true
17	4	Accept	1.3866	1.3925	1.2942	1.296	0.0098157	pureQuadrati	exponential	21.385	true
18	4	Accept	1.4685	0.69952	1.2942	1.296	0.00087169	pureQuadrati	exponential	0.20239	true
19	6	Accept	1.3138	73.803	1.2942	1.2958	0.0044969	pureQuadrati	ardsquaredex	-	false
20	6	Accept	1.4867	1.7363	1.2942	1.2958	0.0005247	pureQuadrati	exponential	0.081975	false

Figure 7. GPR model hyperparameter optimization process.

Fitting the SVR model requires solves an optimization problem that involves two parameters:

- The regularization parameter C and
- The error sensitivity parameter E.

As in the case of SVR model optimization, as shown in Figure.8, the system performs a series of iterations during the hyperparameter optimization process, the fundamental difference being that for SVR the Radial Basis Function (RBF) kernel is predefined.

Iter	Active workers	Eval result	Objective: log(1+loss)	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	BoxConstraint	KernelScale	Epsilon
1	2	Best	2.9052	1.7489	2.9052	2.9052	0.12683	0.54531	45.018
2	2	Accept	2.9405	1.6861	2.9052	2.9052	0.056826	0.5011	0.025879
3	2	Accept	2.9052	1.6187	2.9052	2.9052	154.67	815.07	235.49
4	2	Accept	2.9443	1.6295	2.9052	2.9052	0.0015276	5.9647	0.10548
5	2	Accept	2.9052	1.5804	2.9052	2.9052	4.1332	0.029216	30.733
6	6	Accept	2.9577	1.8492	2.9052	2.9052	276.17	9.1775	0.020594
7	2	Accept	2.9052	0.28569	2.8756	2.8756	3.508	0.0016038	117.38
8	2	Accept	2.9404	0.29796	2.8756	2.8756	0.015879	0.0090936	0.053999
9	2	Accept	2.9379	0.3198	2.8756	2.8756	0.0014853	8.6006	0.037165
10	2	Best	2.8756	0.31788	2.8756	2.8756	772.58	1.1465	1.3118
11	2	Accept	2.8907	0.30165	2.8756	2.8756	569.12	149.75	4.5172
12	6	Accept	2.9052	0.104	2.8756	2.8756	979.5	0.56738	74.602
13	2	Accept	2.9089	0.15961	2.7637	2.8758	1.9303	1.053	1.6613
14	2	Accept	2.9052	0.18501	2.7637	2.8758	0.0031709	0.97711	17.593
15	2	Best	2.7637	0.22891	2.7637	2.8758	13.777	5.2884	0.1274
16	2	Accept	2.9052	0.17736	2.7637	2.8758	0.0046746	943.68	80.136
17	2	Accept	2.8341	0.24153	2.7637	2.8758	319.96	32.373	0.033662
18	6	Accept	2.9206	0.23531	2.7637	2.8898	919.35	360.91	1.5803
19	3	Accept	2.9757	0.07906	2.7637	2.9079	513.97	377.02	0.13892
20	3	Accept	2.9052	0.10063	2.7637	2.9079	42.899	206.76	165.71
21	3	Accept	2.9052	0.1278	2.7637	2.9079	0.019289	220.94	80.697
22	3	Accept	2.9367	0.1371	2.7637	2.9079	0.0061477	0.082057	0.018793
23	6	Accept	2.9052	0.10299	2.7637	2.9078	30.277	0.031413	0.065036
24	3	Accept	2.9744	0.090259	2.7637	2.9288	0.36922	1.3471	0.25043
25	3	Accept	2.956	0.36662	2.7637	2.9288	151.82	694	0.0041846
26	3	Accept	2.9353	0.37787	2.7637	2.9288	0.0068183	3.5423	0.0088767
27	3	Accept	3.3325	0.37197	2.7637	2.9288	0.073441	5.7518	6.9474
28	6	Accept	2.8149	0.077888	2.7637	2.7637	12.754	11.544	0.039623
29	3	Best	2.7414	0.11007	2.7414	2.7415	13.637	6.8708	0.1578
30	3	Accept	3.005	0.091319	2.7414	2.7415	2.2519	0.016151	4.2383
31	3	Accept	2.9052	0.074558	2.7414	2.7415	95.796	13.858	44.52
32	3	Accept	2.9052	0.10059	2.7414	2.7415	36.813	0.6545	152

Figure 8. SVR model hyperparameter optimization process.

RESULTS

GPR outcomes

Once the data were integrated, we applied the GPR algorithm to the data. In this case we sought to optimize both the kernel functions and the hyperparameters. The results show that for each dataset the behavior is different, with the result that the algorithm works well for some municipalities whilst is unsatisfactory for others. This finding is consistent with results obtained from previous studies^(10,34), who similarly found that the same ML algorithm applied to dengue case data works well for some regions (states, municipalities, etc.) but not for others. The models were also evaluated, considering the effect of climate data on the incidence of dengue cases.

The results showed a lag of between two and three weeks in the effect of climate on the incidence of dengue cases in the municipalities of Zulia state. Table 2 shows the municipalities where the prediction trend reproduces the behavior of the real data of the considered municipalities.

Table 2. RMSE and lags.

MUNICIPALITY	Lags	RMSE
Baralt	2	2.94
Cabimas	3	7.13
Colon	3	6.26
Lossada	2	15.35
Mara	2	5.84
Maracaibo	2	35.36
Miranda	2	8.09
Rosario	2	7.8
San Francisco	3	16.2

Figure. 9 shows the prediction results along with the actual dengue cases in six of the municipalities; these cover weeks 377 to 470 in the dataset. In all plots, there is an outbreak of dengue cases approximately between weeks 410 and 430, as well as between the 400th and 420th weeks. It is also observed that the GPR model prediction shows a similar pattern compared to real values.

Those weeks correspond to the last 9 weeks of the year 2015 and the first 11 weeks of the year 2016, that is, from

November 2015 to the end of March 2016. It can be observed that, except for the Baralt municipality, from approximately week 460 to week 470, which corresponds to the last 10 weeks of 2016, dengue cases tend to increase.

This is consistent with the fact that it is in the same period where the previous outbreak was detected. In this way, the model predicted possible dengue outbreaks based on the meteorological conditions.



Figure 9.

Results of the GPR prediction along with the actual cases of dengue for certain municipalities with two-week lag in the effect of climate on the incidence of dengue cases.

SVR outcomes

The results for the SVR show almost exactly the same behavior as for the GPR. Similarly, for some municipalities, approximate values of the real behavior of the outbreaks are obtained, and for other municipalities, adequate results are not obtained. In both cases, they indicate the periods of increase of dengue cases and always with two or three weeks of lag with regards to the climate. Table 3 shows the municipalities where the prediction trend reproduces the behavior of the real data of the considered municipalities.

Table 3: RMSE and lags using SVR

MUNICIPALITY	Lags	RMSE
Baralt	2	2.57
Cabimas	2	8.14
Colon	3	9.43
Lagunillas	2	2.38
Mara	2	7.005
Lossada	3	16.12
Miranda	3	4.43
Padilla	2	2.16
Rosario	2	7.23
San Francisco	2	15.85

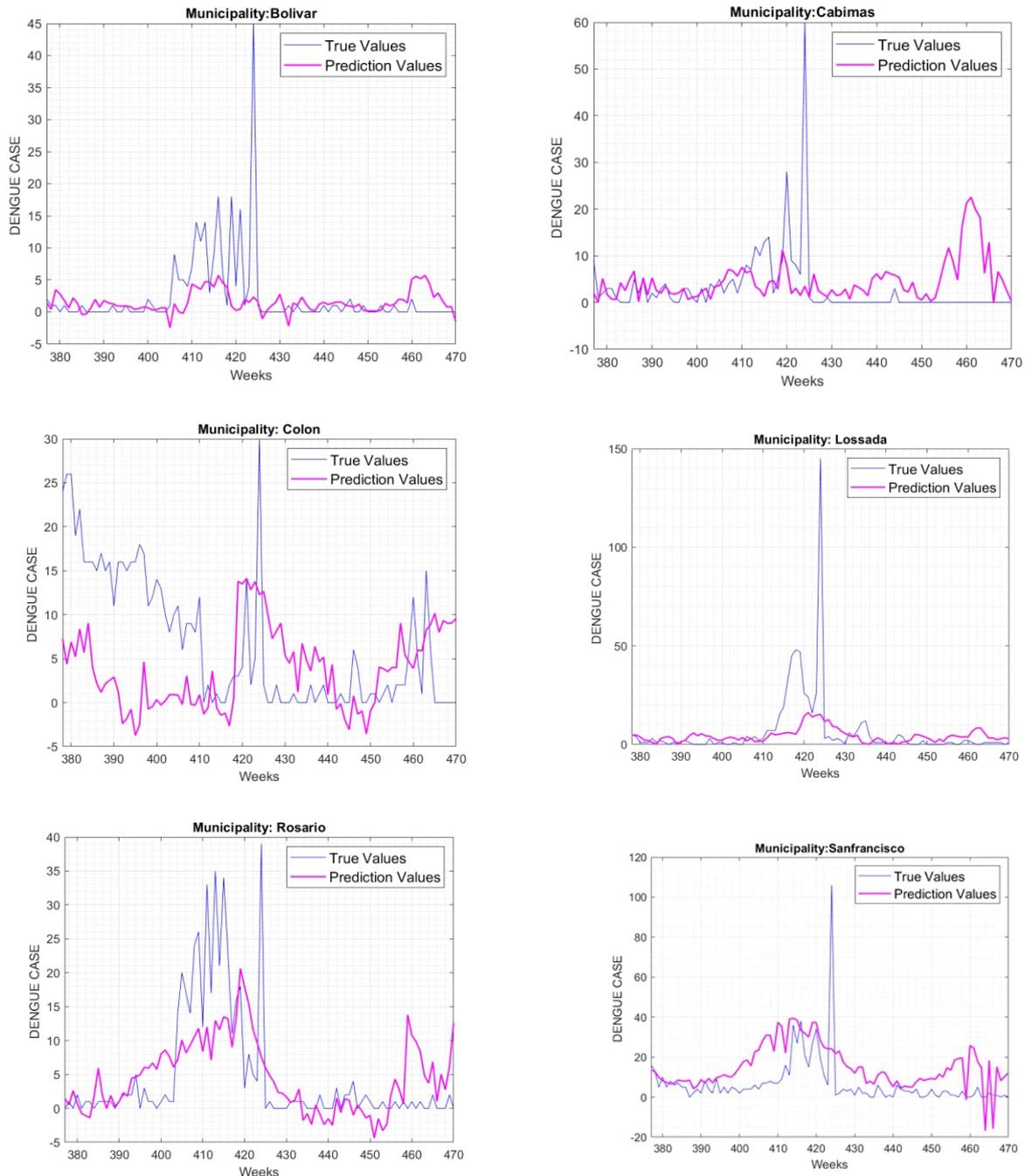


Figure 10. Results of the SVR prediction along with the actual cases of dengue for certain municipalities

DISCUSSION

Dengue continues to be a significant public health problem, especially in Latin America, where multiple countries are currently presenting an epidemic situation in 2024, with more than 13 million cases affecting Brazil and other territories⁽³⁵⁻³⁷⁾. Understanding this epidemiology and developing prediction models is critical for control and surveillance in endemic countries of the region, including Venezuela. This country is of deepening concern due to the country's current economic, political and social difficulties^(5,35). The economic crisis brought it with hyperinflation of 180.9% in 2015 and poverty levels that rose from 33.9% in the first half of 2015 to 82% in 2016⁽³⁵⁾. Consequently, a combination of poverty, socio-economic factors, precarious living conditions and grossly deficient public services such as water and electricity are the main drivers for its ongoing crisis in dengue infections⁽⁵⁾.

In addition, in terms of surveillance systems, the Epidemiological bulletins established in 1938 in Venezuela suffered an interruption from 2007-2014, and ultimately, the publication of healthcare indicators has wholly stopped since 2018, after 63 years of continuous activity⁽⁵⁾. The combination of all those factors has allowed the reemergence of infectious diseases that had been primarily eradicated^(5,35); as a result, the extreme damage to the health and welfare of its inhabitants is a matter that requires urgent actions⁽⁵⁾. Dengue has a significant impact in terms of morbidity, burden and costs⁽³⁶⁻³⁹⁾.

As for predicting mosquito-borne diseases⁽³⁸⁾, a study⁽¹⁰⁾ claims that no model could be completely effective, mainly because it cannot fit all the conditions of the real world. According to these authors, significant constraints are the insufficient available data and the lack of open-source information, in conjunction with an unstructured and inadequate set of information, amongst other drawbacks.

The evaluation of time series data in prediction probably needs further theoretical work. The prediction quality is usually given in terms of the difference between observed and predicted values at each instant. However, this is not really what is required for disease prediction. What is usually needed is the identification of outbreak peaks, and the accuracy of their predicted magnitude is probably less important than the accuracy of predicting them. Knowing that a peak is coming is more important than predicting its size. Similarly, the accuracy of the timing of the peak and the amount of notice that the predictive tool gives is more important than precisely predicting the size of the peak.

According to some authors⁽⁴⁰⁻⁴²⁾, an early warning system for detecting outbreaks of diseases such as dengue (Figure. 11), should consist of the following elements:

- Data Collection: Information is obtained from various sources: epidemiological data, meteorological information and variables.
- Data Preprocessing: Data are integrated to ensure consistency, facilitating the application of the ML model.
- Trained ML model: A model is trained using historical data. This model learns to identify patterns and

anomalies, allowing us to anticipate possible outbreaks.

- Outbreak Prediction and Detection: With real-time data, the model continuously evaluates the risk of outbreaks, generating predictions that facilitate early detection.
- Alert Generation: Automatic alerts are generated to inform the level of risk to health authorities.
- Authorities Intervention: Alerts allow health authorities to intervene appropriately, implementing preventive measures.

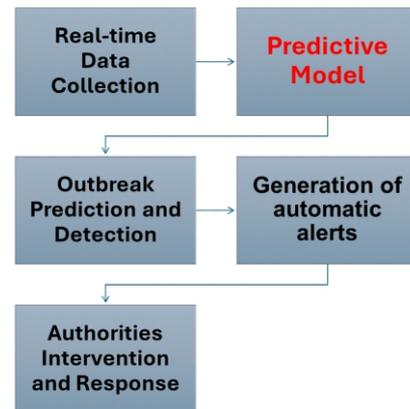


Figure 11.

Basic minimum structure of an early warning system for disease prediction.

Besides, as demonstrated by the findings of this study and supported by previous research^(10,43,44), there is no single machine learning (ML) or deep learning (DL) model that universally adapts to all possible conditions or serves as a definitive solution. In this context, we propose training multiple ML and DL models, selecting those with the best performance, and integrating them into a unified system. This system would incorporate a decision-making structure designed to enhance predictive accuracy, as illustrated in Figure. 12.

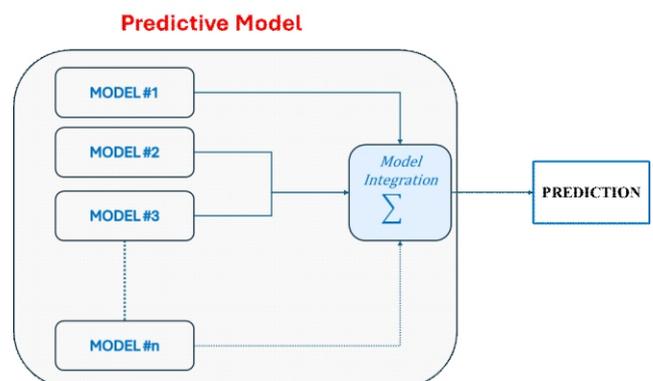


Figure 12.

Prediction model integrating several trained models.

CONCLUSIONS

This study produced findings that are consistent with previous research, showing that predictive models provide satisfactory accuracy in certain geographic areas, but present difficulties in others. In some locations, the models accurately capture relevant data trends and generate reliable forecasts. However, in other regions, accuracy does not meet expectations. These discrepancies may be attributed to the presence of critical factors (unique environmental, demographic, or temporal variables) that have not been clearly identified and adequately integrated into the model. As well, there may be complex interactions specific to each region that significantly alter model performance.

One possible proposal would be that future research should focus on isolating these undefined characteristics by incorporating additional data sets or adopting nonlinear methodologies that may better reflect the diverse nature of the different locations. Another proposal may be to integrate different models into a single predictor module and investigate how to integrate them and how they can be compared, in order to obtain better results.

BIBLIOGRAPHIC REFERENCES

- Messina JP, Brady OJ, Golding N, Kraemer MUG, Wint GRW, Ray SE, et al. The current and future global distribution and population at risk of dengue. *Nat Microbiol*. 2019 Sep;4(9):1508-15. doi:10.1038/s41564-019-0476-8
- Allan R, Budge S, Sauskojus H. What sounds like Aedes, acts like Aedes, but is not Aedes? Lessons from dengue virus control for the management of invasive Anopheles. *Lancet Glob Health*. 2023 Jan;11(1):e165-9. doi:10.1016/S2214-109X(22)00454-5
- Raviglione M, Maher D. Ending infectious diseases in the era of the Sustainable Development Goals. *Porto Biomed J*. 2017 Sep-Oct;2(5):140-2. doi:10.1016/j.pbj.2017.08.001
- Vincenti-Gonzalez MF, Tami A, Lizarazo EF, Grillet ME. ENSO-driven climate variability promotes periodic major outbreaks of dengue in Venezuela. *Sci Rep*. 2018 Apr 10;8(1):5727. doi:10.1038/s41598-018-24003-z
- Grillet ME, Hernández-Villena JV, Llewellyn MS, Paniz-Mondolfi AE, Tami A, Vincenti-Gonzalez MF, et al. Venezuela's humanitarian crisis, resurgence of vector-borne diseases, and implications for spillover in the region. *Lancet Infect Dis*. 2019 May;19(5):e149-61. doi:10.1016/S1473-3099(18)30757-6
- Cabrera M, Taylor G. Modelling spatio-temporal data of dengue fever using generalized additive mixed models. *Spat Spatio-Temporal Epidemiol*. 2019 Feb;28:1-13. doi:10.1016/j.sste.2018.11.006
- Gutiérrez LA. Pan American Health Organization / World Health Organization. 2024 [cited 2024 Jun 12]. PAHO/WHO Data - Venezuela - Dengue Cases | PAHO/WHO. Available from: <https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/dengue-subnacional-en/576-ven-dengue-casos-en.html>
- Kakarla SG, Kondeti PK, Vavilala HP, Boddada GSB, Mopuri R, Kumaraswamy S, et al. Weather integrated multiple machine learning models for prediction of dengue prevalence in India. *Int J Biometeorol*. 2023 Feb;67(2):285-97. doi:10.1007/s00484-022-02405-z
- World Health Organization. Global strategy for dengue prevention and control 2012-2020 [Internet]. Geneva: WHO; 2012 [cited 2023 Dec 6]. Available from: <https://iris.who.int/handle/10665/75303>
- Cabrera M, Leake J, Naranjo-Torres J, Valero N, Cabrera JC, Rodríguez-Morales AJ. Dengue Prediction in Latin America Using Machine Learning and the One Health Perspective: A Literature Review. *Tropical Med*. 2022 Oct 21;7(10):322. doi:10.3390/tropicalmed7100322
- Barboza LA, Chou-Chen SW, Vásquez P, García YE, Calvo JG, Hidalgo HG, et al. Assessing dengue fever risk in Costa Rica by using climate variables and machine learning techniques. *PLoS Negl Trop Dis*. 2023 Jan 13;17(1):e0011047. doi:10.1371/journal.pntd.0011047
- Nalini C, Shanthakumari R, VPrasanna R, Nikilesh A, S.M. NP. Prediction of dengue infection using Machine Learning. In: 2022 International Conference on Computer Communication and Informatics (ICCCI). Coimbatore, India: IEEE; 2022 [cited 2023 Dec 6]. doi:10.1109/ICCCI54379.2022.9740924
- Sanchez-Gendriz I, Souza GF de, Andrade IGM de, Neto ADD, Tavares A de M, Barros DMS, et al. Data-driven computational intelligence applied to dengue outbreak forecasting: a case study at the scale of the city of Natal, RN-Brazil. *Sci Rep*. 2022 Apr;12(1):6550. doi:10.1038/s41598-022-09311-9
- Souza C, Maia P, Stolerma LM, Rolla V, Velho L. Predicting dengue outbreaks in Brazil with manifold learning on climate data. *Expert Syst Appl*. 2022 Apr;192:116324. doi:10.1016/j.eswa.2021.116324
- Kesorn K, Ongruk P, Chompoonsri J, Phumee A, Thavara U, Tawatsin A, et al. Morbidity Rate Prediction of Dengue Hemorrhagic Fever (DHF) Using the Support Vector Machine and the Aedes aegypti Infection Rate in Similar Climates and Geographical Areas. *PLoS ONE*. 2015 May 11;10(5):e0125049. doi:10.1371/journal.pone.0125049
- Niu Y, Ye S. Data Prediction Based on Support Vector Machine (SVM)—Taking Soil Quality Improvement Test Soil Organic Matter as an Example. *IOP Conf Ser: Earth Environ Sci*. 2019 Jul 1;295(2):012021. doi:10.1088/1755-1315/295/2/012021
- Roberts S, Osborne M, Ebdon M, Reece S, Gibson N, Aigrain S. Gaussian processes for time-series modelling. *Phil Trans R Soc A*. 2013 Feb 13;371(1984):20110550. doi:10.1098/rsta.2011.0550
- Rasmussen CE. Gaussian Processes in Machine Learning. In: Bousquet O, Von Luxburg U, Rätsch G, editors. *Advanced Lectures on Machine Learning*. Berlin, Heidelberg: Springer; 2004. p. 63-71. (Lecture Notes in Computer Science; vol. 3176). doi:10.1007/978-3-540-28650-9_4
- Tanawi IN, Vito V, Sarwinda D, Tasman H, Hertono GF. Support Vector Regression for Predicting the Number of Dengue Incidents in DKI Jakarta. *Procedia Comput Sci*. 2021;179:747-53. <https://doi.org/10.1016/j.procs.2021.01.063>
- Lyu C, Liu X, Mihaylova L. Review of Recent Advances in Gaussian Process Regression Methods. In: Panoutsos G, Mahfouf M, Mihaylova LS, editors. *Advances in Computational Intelligence Systems*. Cham: Springer; 2024. p. 226-37. (Advances in Intelligent Systems and Computing; vol. 1454). doi:10.1007/978-3-031-55568-8_19
- Baker QB, Faraj D, Alguzo A. Forecasting Dengue Fever Using Machine Learning Regression Techniques. In: 2021 12th International Conference on Information and Communication Systems (ICICS). IEEE; 2021. doi:10.1109/ICICS52457.2021.9464619
- Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: A case study in China. *PLoS Negl Trop Dis*. 2017 Oct 16;11(10):e0005973. doi:10.1371/journal.pntd.0005973
- Hoyos W, Aguilar J, Toro M. Dengue models based on machine learning techniques: A systematic literature review. *Artif Intell Med*. 2021 Sep;119:102157. doi:10.1016/j.artmed.2021.102157
- NASA Earth Science. The POWER Project. [cited 2023 Dec 9]. Available from: <https://power.larc.nasa.gov/>
- NOAA. Climate Prediction Center - Monitoring & Data: Current Monthly Atmospheric and Sea Surface Temperatures Index Values. [cited 2023 Dec 9]. Available from: <https://www.cpc.ncep.noaa.gov/data/indices/>
- Instituto Nacional de Estadística. XIV CENSO NACIONAL DE POBLACIÓN Y VIVIENDA. Instituto Nacional de Estadística; 2014. Available from: <http://www.inec.gov.ve/anda4/index.php/catalog/18>
- Albinati J, Meira W, Pappa GL. An Accurate Gaussian Process-Based Early Warning System for Dengue Fever. In: 2016 5th Brazilian Conference on Intelligent Systems (BRACIS). IEEE; 2016. p. 43-8. doi:10.1109/BRACIS.2016.7839560
- Manogaran G, Lopez D. A Gaussian process based big data processing framework in cluster computing environment. *Cluster Comput*. 2018 Mar;21(1):189-204. doi:10.1007/s10586-017-1164-5
- Arias Velásquez RM, Mejía Lara JV. Forecast and evaluation of COVID-19 spreading in USA with reduced-space Gaussian process regression. *Chaos Solitons Fractals*. 2020 Jul;136:109924. doi:10.1016/j.chaos.2020.109924
- Cheng LF, Dumitrascu B, Darnell G, Chivers C, Draugelis M, Li K, et al. Sparse multi-output Gaussian processes for online medical time series prediction. *BMC Med Inform Decis Mak*. 2020 Dec;20(1):152.

- doi:10.1186/s12911-020-01239-5
31. Ketu S, Mishra PK. Enhanced Gaussian process regression-based forecasting model for COVID-19 outbreak and significance of IoT for its detection. *Appl Intell*. 2021 Mar;51(3):1492-512. doi:10.1007/s10489-020-01968-0
 32. Lubbe F, Maritz J, Harms T. Evaluating the Potential of Gaussian Process Regression for Solar Radiation Forecasting: A Case Study. *Energies*. 2020 Oct 21;13(20):5509. doi:10.3390/en13205509
 33. Statistics and Machine Learning Toolbox Documentation. [cited 2024 Jun 26]. Available from: https://www.mathworks.com/help/stats/index.html?s_tid=CRUX_lftnav
 34. Appice A, Gel YR, Iliev I, Lyubchich V, Malerba D. A Multi-Stage Machine Learning Approach to Predict Dengue Incidence: A Case Study in Mexico. *IEEE Access*. 2020;8:52713-25. doi:10.1109/ACCESS.2020.2980634
 35. Roa AC. Sistema de salud en Venezuela: ¿un paciente sin remedio? *Cad Saúde Pública*. 2018 Mar 5;34(3). Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2018000305003&lng=es&tlng=es
 36. Rodríguez-Morales AJ, Montenegro-Idrogo JJ, Celis-Salinas JC, Angerami R, Villamil-Gómez WE, Sarute N, et al. Unraveling the unparalleled 2024 epidemic of Dengue in the Americas. *Rev chil infectol*. 2024 Jun;41(3):421-8. <https://mail.revinf.cl/index.php/revinf/article/view/2145/987>.
 37. Rodríguez-Morales AJ, López-Medina E, Arboleda I, Cardona-Ospina JA, Castellanos JE, Faccini-Martínez AA, et al. The Epidemiological Impact of Dengue in Colombia: A Systematic Review. *Am J Trop Med Hyg*. 2025 Jan 8;112(1):182-8. doi:10.4269/ajtmh.25-0025
 38. Mattar S, López Y, Arrieta G, Mattar AS, Bandeira AC, Paniz-Mondolfi A, et al. The next arbovirus epidemic in Latin America and beyond: A question of when, not if - Mayaro, Oropouche, Usutu or Disease X? *Travel Med Infect Dis*. 2025 Mar;64:102818. doi:10.1016/j.tmaid.2025.102818
 39. Pareek A, Singhal R, Pareek A, Chuturgoon A, Rodríguez-Morales AJ. The unprecedented surge of dengue in the Americas: Strategies for effective response. *J Infect Public Health*. 2024 Dec;17(12):102585. doi:10.1016/j.jiph.2024.102585
 40. Dey SK, Rahman MdM, Howlader A, Siddiqi UR, Uddin KMM, Borhan R, et al. Prediction of dengue incidents using hospitalized patients, meteorological and socio-economic data in Bangladesh: A machine learning approach. *PLoS ONE*. 2022 Jul 20;17(7):e0270933. doi:10.1371/journal.pone.0270933
 41. Arrubla-Hoyos W, Solano-Barliza A. Contribuciones del aprendizaje automático en el descubrimiento del dengue: un análisis científico. *Rev cuba inf cienc salud*. 2024 Jul 3;35(0). Available from: <https://acimed.sld.cu/index.php/acimed/article/view/2630>
 42. Universidad Nacional Pedro Ruiz Gallo, Maquen-Niño GLE, Bravo J, Alarcón R, et al. Una revisión sistemática de Modelos de clasificación de dengue utilizando machine learning. *risti*. 2023 Jun 30;(50):5-27. <https://scielo.pt/pdf/rist/n50/1646-9895-rist-50-5.pdf>.
 43. Kamarudin ANA, Zainol Z, Kassim NFA. Forecasting the Dengue Outbreak using Machine Learning Algorithm: A Review. In: 2021 International Conference of Women in Data Science at Taif University (W i D S T a i f) . I E E E ; 2 0 2 1 . <https://ieeexplore.ieee.org/document/9430193/>.
 44. Anggraini Ningrum DN, Li YJ, Hsu CY, Solihuddin Muhtar M, Pandu Suhito H. Artificial Intelligence Approach for Severe Dengue Early Warning System. *Stud Health Technol Inform*. 2024 Jan 25;310:881-885. doi: 10.3233/SHT1231091.
 45. Popa CL, Dobrescu TG, Silvestru CI, Firulescu AC, Popescu CA, Cotet CE. Pollution and Weather Reports: Using Machine Learning for Combating Pollution in Big Cities. *Sensors*. 2021 Nov 3;21(21):7329. doi:10.3390/s21217329
 46. Bassman Otelie L, Rajak P, Kalia RK, Nakano A, Sha F, Sun J, et al. Active learning for accelerated design of layered materials. *npj Comput Mater*. 2018 Dec 10;4(1):74. doi:10.1038/s41524-018-0100-3
 47. Neal RM. *Bayesian Learning for Neural Networks*. New York, NY: Springer; 1996. (Lecture Notes in Statistics; vol. 118). doi:10.1007/978-1-4612-0745-0.