International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 11 | Nov 2024 www.irjet.net p-ISSN: 2395-0072

Predicting Dengue in India through Machine Learning: A Literature Review

Hemashree H C[1], Shashikala S R[2], Siddesh T S[3], Ruchita R Naik[4]

[1]Ass. Prof. Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, affiliated to VTU Belagavi, Karnataka, India.
[2]Bachelor of Engineering, Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, affiliated to VTU Belagavi, Karnataka, India.
[3]Bachelor of Engineering, Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, affiliated to VTU Belagavi, Karnataka, India.
[4]Bachelor of Engineering, Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, affiliated to VTU Belagavi, Karnataka, India.

---***---

Abstract - *The rising incidence of dengue fever in India and other countries is a severe public health issue that is being made worse by human migration and climate change. This study examines several methods for predicting the epidemiology of dengue fever from a One Health standpoint.*

In order to integrate the larger environmental elements into a thorough comprehension of the small-scale processes as they impact disease incidence, this study examines how machine learning techniques have been applied to it and focuses on the risk factors for dengue in India. Given that a variety of indicators can be used to forecast dengue outbreaks, the best predictors have not yet been identified through a large-scale comparison of several predictors over broader geographic areas than those that have been researched.

Key Words: Dengue; Climate Change; One Health; Machine Learning; Epidemiology; Prediction; India.

1.INTRODUCTION

Dengue fever is a rapidly spreading, mosquito-borne viral illness that is mostly carried by female Aedes mosquitoes [1,2], particularly A. aegypti, A. albopictus, and A. vitattus. These illustrations are common ectoparasites that eat blood in tropical regions. Over the course of the last 60 years, this zoonotic illness has moved from just 9 nations experiencing severe epidemics to becoming endemic in over 100 countries globally, including affecting non-tropical or subtropical locations [4,5]. It originated in African or Asian non-human primates 500–1000 years ago [3].

Moreover, the symptomatic condition [4] produced by its four serotypes affects over 100 million people annually [6]. Considering how much environmental changes affect the spread of disease, the One Health concept is desperately needed to integrate ecological, animal, and human health. This work aims to shed light on methods for developing One Health-based prediction models for the future, primarily with reference to India but also with applications world wide. Furthermore, especially for India and other regions, the research offers insights into methods that can be applied to prediction models in the future. With their potential to transform public health and epidemiology and provide fresh insights into dengue and other infectious illnesses, techniques like machine learning (ML) have seen an exponential surge in popularity.

1.1 The Indian dengue scenario

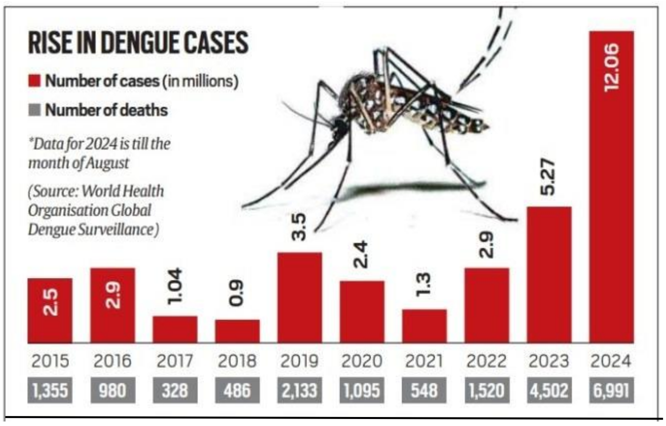

Chart-1: Rise in Dengue Cases

From 2015 to 2024, dengue cases and related deaths increased, as seen in the chart (data for 2024 is up to August). This is an in-depth explanation:

1. **Cases of dengue (in millions):**

Between 2023 and 2024, the number of dengue cases increased from 5.27 million to 12.06 million, a notable increase over the previous years.

Dengue case trends are shown in the chart, with significant peaks in 2019 (3.5 million cases), 2022 (2.9 million cases), and 2024.

2. Deaths Due to Dengue:

Additionally, dengue-related mortality have increased, with a discernible uptick in recent years.6,991 deaths were reported by August of 2024, which was a considerable increase over previous years and a dramatic increase from 4,502 deaths in 2023.

3. Analysis of Trends:

The data indicates deteriorating dengue epidemics, with both cases and deaths showing a worrisome upward trend. Urbanization, changing temperature conditions, and insufficient vector control efforts are some of the possible causes of this increase.

4. Cases and Deaths Comparison:

The mortality rate appears to be rising as both measurements climb at the same time, even though the number of cases in millions is far higher than the number of fatalities. The data, which comes from WHO worldwide Dengue Surveillance, emphasizes how urgently greater prevention, early detection, and treatment strategies are needed to counteract dengue's growing worldwide effect.

2. Dengue Control Through a One Health Approach

A general trend over the past few decades has been the vast geographical expansion of dengue fever worldwide, which has given rise to a renewed interest in finding crucial variables in disease transmission [7]. Comprehending the intricate interplay among environment, human, and biological factors is crucial in comprehending the trajectory of emerging infectious diseases and its consequences for the public health domain [9]. In this regard, the 2004 One World-One Health program recommends a more transdisciplinary and holistic approach as a worldwide tactic in the battle against those infectious diseases [8], especially dengue fever in the India. As a result, this new knowledge prompts us to adopt a more comprehensive global health perspective on dengue transmission scenarios. From this angle, the risk factors for dengue fever can be divided into four basic categories. This viewpoint allows us to categorize the dengue fever risk variables into four basic groups according to the One Health approach: (i) dengue serotypes, (ii) vector ecology, (iii) human conditions, and (iv) environment (Figure.1)

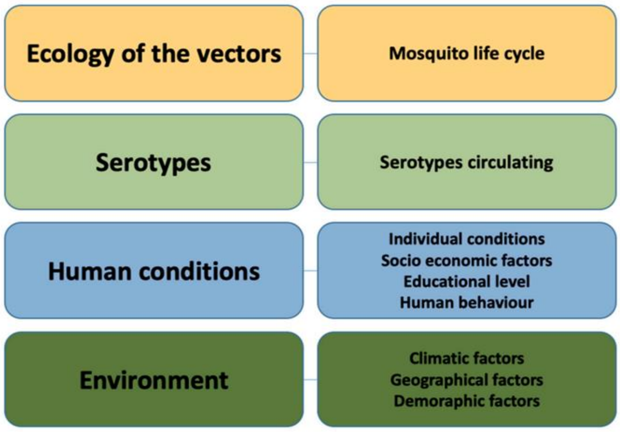

Fig-1: Shows the four One Health groups into which dengue fever risk factors are categorized: (i) vector ecology, (ii) serotypes, (iii) human conditions, and (iv) environment.

The One Health framework is seen in Figure 1 along with the following subcategories: (i) dengue serotypes; (ii) the ecology of the vector (mosquito life cycle); (iii) human conditions (individual conditions, socioeconomic factors, educational attainment, human behavior); and (iv) the environment (climatic circumstances, geographic factors, demographic factors). To better understand each cofactor's role in the existence and severity of the disease, these will be utilized as guidelines.

3. Human Conditions

3.1 Socioeconomic Factors:

In terms of dengue fever transmission, novel techniques based on various methodologies have offered crucial insights into how the urban environment or social network affects residents' well-being [8] and how those findings could direct tactics related to public health in the battle against illness. Numerous studies carried out globally reveal a robust correlation between the severity of the disease and the lack of a public water supply [1, 16–21]. Insufficient coverage of the water supply encourages the use of man-made containers to retain water, which in turn provides mosquito breeding grounds [22]. Ineffective urban planning and development may be the cause of this deficiency in sanitary services [16,21,23].

Moreover, some research supports the notion that greater population density and a lower GDP per capita could be predictive factors for the severity of dengue fever. Poverty and inadequate public health care can be caused by low per capita GDP. Other research discovered that a high percentage of abandoned and vacant homes in the neighbourhood, together with unsatisfactory housing conditions, are risk factors for the illness [10]. Unplanned urbanization, poor health infrastructure, ineffective disease control programs, and inadequate or non-existent piped water and sewage services are the major pillars of a marginalized society, and these conditions are frequently the result of bad state administration [11,16].

3.2 Human Behavior

Consistent human actions, such as the neglect of washing and bathing water containers, also raise the possibility of dengue outbreaks globally. Because most people keep their water storage containers uncovered, these authors—who were working in India—found that this unintentionally creates an ideal environment for mosquito reproduction. The neighbourhood's surrounding area's uncovered trash is another element influencing human behaviour.

It's possible that precipitation will fill up discarded garbage bottles, cans, plastic containers, and tires, creating mosquito breeding grounds. In several other nations, like Venezuela [20], Brazil [17,18], and Thailand, similar dengue outcomes based on measurements of trash surrounding the area have been confirmated. Seasonal trends are significant in this context due to their impact on human behaviours such as school schedules, vacations, travel patterns, weekday or weekend activities, and outdoor exposures, in addition to their influence on climatic and environmental factors [13].

4.Environmental Factors

Climatic Factors:

Dengue fever is a climate-sensitive disease, with temperature, humidity, and rainfall being important drivers, as evidenced by numerous research that link outbreaks of the disease to climatological parameters [1,14,15,19]. Further research is necessary to fully comprehend the immediate and long-term impacts of those environmental variables, even if the seasonal pattern of dengue fever has been extensively explained in terms of climatic parameters [13]. Furthermore, several specialists advise creating models centered on regional or local contexts as opposed to broader global models to improve the forecast sensitivity for dengue outbreaks .

Increasing ambient temperature decreases mosquito larval size, which results in smaller developed adults that eat blood more quickly. This is just one way that warmer temperatures influence the disease's spread. As such, these adults need more blood in order to provide nutrition for their eggs [1]. Second, changes in climate have made places that were previously thought to be unsuitable for the growth of mosquito populations better, or even ideal.

5. Machine Learning in India for Dengue Predictive Applications

Machine Learning Overview:

The review of ML-based epidemiological predictive modelling carried out in the India region is the main focus of this work. A brief summary of machine learning ideas is presented in this section, followed by a review of applied machine learning studies on dengue disease, with a focus on predictive model applications to epidemics. Recent advances in Big Data and Artificial Intelligence (AI) have led to these disciplines becoming increasingly recognized as scientific disciplines across various study domains. Computerized systems capable of intelligent behavior are referred to as artificial intelligence (AI) .

Supervised learning is the first of three learning approaches used in machine learning. It trains the model using inputs and outputs that are known. The second type of learning is called unsupervised learning, which occurs when a system attempts to learn without any prior knowledge and discovers innate structures or hidden patterns in the data. In conclusion, intelligent agents or Reinforcement Learning systems have the ability to monitor their surroundings, operate in the background, and receive reinforcement in return. The learning types commonly used in machine learning (ML) are diagrammed in Figure 2 along with the related applications.

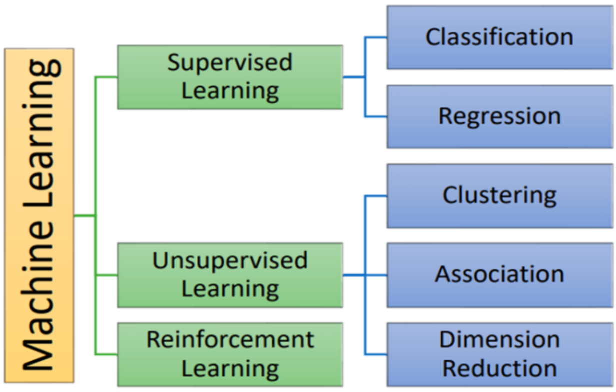

Fig-2.: Machine learning often uses a variety of learning methods.

The many frameworks available for ML use include interfaces, libraries, or tools that make it simple to build ML models. Among the most popular machine learning frameworks are: TensorFlow is an open-source machine learning library created by Google; PyTorch is an open-source cloud platform utilized by IBM and Facebook; (iii) MatLab (MathWorks) is a for-profit platform for programming and numerical computation that offers a range of Toolboxes for creating and executing ML and DL models; (iv) Scikit-learn is a free machine learning library for the Python programming language.

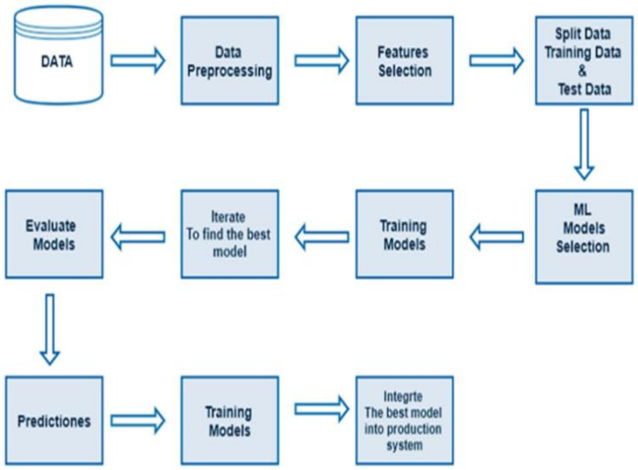

Fig-3: General workflow for ML project

The process depicted in figure 3 is typically followed in an ML project. The picture shows a general, step-by-step process for a machine learning (ML) project, including data collection, preprocessing to clean and organize the data, feature selection to identify relevant attributes for modelling, data splitting into training and testing sets, iterative training of various ML models to identify the best model, evaluation of the models based on performance metrics, and integration of the best model into a production system for predictions. This process guarantees that the model selected can produce dependable results in real-world applications and that it generalizes well.

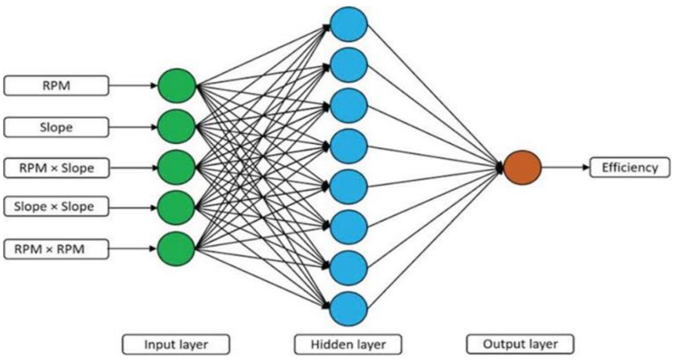

Fig- 4: General architecture of an ANN.

A network of interconnected nodes called an artificial neural network is comparable to the enormous network of neurons found in a human brain. They are a model that was influenced by how the human brain works. An artificial neural network (ANN) is designed as follows: (i) the input layer accepts input data and forwards it to the first hidden layer; (ii) the hidden layers use the inputs to execute mathematical computations; and (iii) the output layer delivers the predicted value. This is illustrated in Figure 4.

5.1 Machine Learning Utilized for Dengue in India

More recently, machine learning (ML) has been applied to research dengue fever at various stages of the disease or for prognostic purposes. Three key study areas were found in a systematic review of studies using machine learning (ML) to model dengue fever [12]. Utilizing signs and symptoms is crucial for constructing models that address the diagnosis and severity of the condition, which can be addressed by these modelling approaches. Therefore, they are known as models of prescription; secondly, models of epidemics are produced by analyzing the severity of dengue in a particular population; these models are utilized for forecasting purposes; and lastly, models of intervention are created, which comprise the optimization.

But according to a study, no model can be fully adopted for predicting outbreaks of mosquito-borne diseases (MBD), primarily because none of the models can fulfil all the real-world requirements. The primary limitation, as stated by the authors, arises from an inadequate and unstructured set of information, as well as from the absence of open-source material available online and insufficient data that is currently available. With the Entomological Index component, an improved framework is therefore suggested in the literature assessment of the MBD epidemic prediction framework in America. Additionally, ML is used to improve the forecast of future MBD outbreaks.

Table -1: Most widely used machine learning approaches for forecasting [23].

Abbreviation	Name of ML Technique
LoR	Logistic Regression
RF	Random Forest
LiR	Linear Regression
GAM	Generalized Additive Model
GLM	Generalized Linear
DT	Decision Trees
SVM	Support Vector Machines
ANN	Artificial Neural Networks
GBM	Gradient Boosting Machine
KNN	K-Nearest Neighbors
BRT	Boosted Regression Trees

Certain studies concentrate only on machine learning methods for dengue prediction in the Americas; multiple studies assess the relative merits of various predictive machine learning approaches. Using a neural network model based on Long Short-Term Memory (LSTM) to forecast future dengue cases, is one of the more intriguing examples. Between 2016 and 2019, a weekly set of data were gathered from 397 ovitraps located throughout the municipality of Natal, Brazil, to determine the dengue incidence and the Egg Density Index. When dengue cases from prior weeks were used to predict dengue incidence, the LSTM models and ovitrap data forecasts demonstrated a goodness-of-fit indicated by correlation coefficients of 0.92 and 0.87, respectively. The findings demonstrated that an earlier dengue prediction was made possible by ovitrap data.

Dengue fever epidemics in Yucatan, Mexico, and San Juan, Puerto Rico, were predicted using artificial neural networks. The information gathered correlated with six years of dengue cases in Mexico and 19 years of infections in Puerto Rico. Environmental and demographic information was incorporated into the analysis, including population sizes, dengue case epidemiology data, air temperature, humidity, precipitation, and sea surface temperature. In each area, the authors used two models: one to assess the size of the vulnerable population and the other to forecast occurrence. They were able to anticipate with above 70% accuracy.

In a different study, the authors used data from 2010 to 2014 to predict the rates of dengue fever (DF) and dengue hemorrhagic fever (DHF) in urban areas of Rio De Janeiro, Brazil, using street-level images, such as those from Google Street View, processed by convolutional neural networks (CNN). Furthermore, the authors assessed both deep and simple Siamese CNN architectures and compared the outcomes with both models. The results of the study showed that identifying DF/DHF hot spots in urban areas may be advantageous when doing so with deep CNNs and street-level imagery.

Table -2: Lists potential interventions for dengue epidemics that address the behavior, physiology, habitats, and epidemiological significance of yellow fever.

Mosquito Diversity	
Species richness	Total number of species sampled at each sampling point.
Shannon–Wiener index	Measure of species diversity weighted by the relative abundance.
Functional richness (FRic)	Represents the quantity of functional space filled by the community, where low FRic implies that some resources are unused or unavailable in the ecosystem.
Functional evenness (FEve)	Describes the distribution of abundance in a functional space of traits, where low FEve indicates that some parts of the functional niche are underutilized.
Functional divergence (FDiv)	A measure of the functional similarity among the dominant mosquito species of a community. FDiv is high when the most abundant species have extreme functional trait values.
Functional dispersion (FDis)	A multivariate measure of the dispersion of mosquito species in the trait space represents the mean distance of species to the centroid of the community, weighted by mosquito species abundance.
Haemagogus relative abundance	The number of Haemagogus mosquitoes is divided by the number of mosquitoes collected at each sampling point.
Haemagogus minimum infection rate (MIR)	Represents the minimum number of infected mosquitoes, assuming that only one was infected in each positive mosquito pool. It was calculated for each sampling point using the formula MIR = number of YFV-positive.
Ecological Indexes	
Environment of Mosquito sampling, inside the forest	Within dense forests connected to other forests.
Rural fragment	Within forests smaller than 100 hectares and surrounded by pastures.
Rural peri-domicile	Around homesteads and country houses.
Urban fragment	Within forests inside cities.
Urban intra-domicile	Within human houses inside cities.
Vertical distribution in the forest	

Geo-Environmental Indexes	
Altitude, landcover/land use, forest fragment size.Normalized Difference Vegetation Index (NDVI).	
Functional Diversity	
Physiology	Egg resistance to larval development speed.
Habitats	Seasonal distribution; primary habitat.
Epidemiological	Epidemiological importance Importance concerning disease.
Behavior	Main hourly biting activity;

Lastly, even though the goal of the current study was to apply machine learning to dengue fever outbreaks, the comprehensive approach offers a chance to draw lessons from other infectious diseases spread by vectors, like yellow fever [25]. Thus, in order to forecast regions that are susceptible to yellow fever outbreaks, our study used a comprehensive set of entomological data as well as landscape composition. Table 2 lists the measures pertaining to behavior, physiology, habitats, and epidemiological significance at each sampling point.

6. CONCLUSIONS

Applying machine learning (ML) techniques to dengue outbreak prediction has proven to be a useful strategy for addressing public health issues in India. By utilizing models like artificial neural networks (ANNs) and data-driven frameworks, ML can analyse a variety of factors that influence dengue transmission, such as environmental, behavioural, and epidemiological data. These techniques improve the ability to forecast outbreaks, enabling timely interventions and resource allocation. However, the successful application of ML for dengue prediction depends on the availability of high-quality, comprehensive datasets and the appropriate adaptation of models to local contexts. Although promising, these systems still have some drawbacks, such as the complexity of real-world data and the requirement for domain-specific customizations.

Future developments in machine learning technology, when paired with stakeholder collaboration, have the potential to further increase the accuracy and efficacy of dengue epidemic prediction and control.

REFERENCES

[1] 1. Cabrera, M.; Taylor, G. Modelling Spatio-Temporal Data of Dengue Fever Using Generalized Additive Mixed Models. Spat. Spatio-Temporal Epidemiol. 2019, 28, 1–13. [CrossRef] [PubMed]

[2] 2. Castrillón, J.C.; Castaño, J.C.; Urcuqui, S. Dengue En Colombia: Diez Años de Evolución. Rev. Chil. Infectol. 2015, 32, 142–149. [CrossRef] [PubMed]

[3] 3. Holmes, E.C.; Twiddy, S.S. The Origin, Emergence and Evolutionary Genetics of Dengue Virus. Infect. Genet. Evol. 2003, 3, 19–28. [CrossRef]

[4] 4. Messina, J.P.; Brady, O.J.; Pigott, D.M.; Brownstein, J.S.; Hoen, A.G.; Hay, S.I. A Global Compendium of Human Dengue Virus Occurrence. Sci. Data 2014, 1, 140004. [CrossRef]

[5] 5. Saydam, F.N.; Erdem, H.; Ankarali, H.; Ramadan, M.E.E.-A.; El-Sayed, N.M.; Civljak, R.; Pshenichnaya, N.; Moroti, R.V.; Mahmuodabad, F.M.; Maduka, A.V.; et al. Vector-Borne and Zoonotic Infections and Their Relationships with Regional and Socioeconomic Statuses: An ID-IRI Survey in 24 Countries of Europe, Africa and Asia. Travel Med. Infect. Dis. 2021, 44, 102174. [CrossRef]

[6] 6. Iqbal, N.; Islam, M. Machine Learning for Dengue Outbreak Prediction: An Outlook. Int. J. Adv. Res. Comput. Sci. 2017, 8, 93–102

[7] 7. Tuladhar, R.; Singh, A.; Varma, A.; Choudhary, D.K. Climatic Factors Influencing Dengue Incidence in an Epidemic Area of Nepal. BMC Res. Notes 2019, 12, 131. [CrossRef]

[8] 8. Destoumieux-Garzón, D.; Mavingui, P.; Boetsch, G.; Boissier, J.; Darriet, F.; Duboz, P.; Fritsch, C.; Giraudoux, P.; Roux, F.L.; Morand, S.; et al. The One Health Concept: 10 Years Old and a Long Road Ahead. Front. Vet. Sci. 2018, 5, 14. [CrossRef]

[9] 9. Arora, M.; Curtin, P.; Curtin, A.; Austin, C.; Giuliani, A.; Birnbaum, L.S. Environmental Biodynamics: A New Science of How the Environment Interacts with Human Health; Oxford University Press: Oxford, UK, 2021; ISBN 978-0-19-758294-7.

[10] 10. Jácome, G.; Vilela, P.; Yoo, C. Social-Ecological Modelling of the Spatial Distribution of Dengue Fever and Its Temporal Dynamics in Guayaquil, Ecuador for Climate Change Adaption. Ecol. Inform. 2019, 49, 1–12. [CrossRef]

[11] 11. Ibarra, A.M.S.; Ryan, S.J.; Beltrán, E.; Mejía, R.; Silva, M.; Muñoz, Á. Dengue Vector Dynamics (Aedes Aegypti) Influenced by Climate and Social Factors in Ecuador: Implications for Targeted Control. PLoS ONE 2013, 8, e78263. [CrossRef]

[12] 12.Hoyos, W.; Aguilar, J.; Toro, M. Dengue Models Based on Machine Learning Techniques: A Systematic Literature Review. Artif. Intell. Med. 2021, 119, 102157. [CrossRef] [PubMed]

[13] 13.Sippy, R.; Herrera, D.; Gaus, D.; Gangnon, R.E.; Patz, J.A.; Osorio, J.E. Seasonal Patterns of Dengue Fever in Rural Ecuador: 2009–2016. PLOS Negl. Trop. Dis. 2019, 13, e0007360. [CrossRef] [PubMed]

[14] 14. Hii, Y.L.; Rocklöv, J.; Ng, N.; Tang, C.S.; Pang, F.Y.; Sauerborn, R. Climate Variability and Increase in Intensity and Magnitude of Dengue Incidence in Singapore. Glob. Health Action 2009, 2, 2036. [CrossRef]

[15] 15. Yu, H.-L.; Yang, S.-J.; Yen, H.-J.; Christakos, G. A Spatio-Temporal Climate-Based Model of Early Dengue Fever Warning in Southern Taiwan. Stoch. Environ. Res. Risk Assess. 2010, 25, 485–494. [CrossRef]

[16] 16. Guha-Sapir, D.; Schimmer, B. Dengue Fever: New Paradigms for a Changing Epidemiology. Emerg. Themes Epidemiol. 2005, 2, 1. [CrossRef]

[17] 17. Ferreira, G.S.; Schmidt, A.M. Spatial Modelling of the Relative Risk of Dengue Fever in Rio de Janeiro for the Epidemic Period between 2001 and 2002. Braz. J. Probab. Stat. 2006, 20, 29–47.

[18] 18. Mondini, A.; Chiaravalloti-Neto, F. Spatial Correlation of Incidence of Dengue with Socioeconomic, Demographic and Environmental Variables in a Brazilian City. Sci. Total Environ. 2008, 393, 241–248. [CrossRef]

[19] 19. Lowe, R.; Bailey, T.C.; Stephenson, D.B.; Graham, R.J.; Coelho, C.A.S.; Carvalho, M.S.; Barcellos, C. Spatio-Temporal Modelling of Climate-Sensitive Disease Risk: Towards an Early Warning System for Dengue in Brazil. Comput. Geosci. 2011, 37, 371–381. [CrossRef]

[20] 20. Barrera, R.; Delgado, N.; Jiménez, M.; Valero, S. Eco-Epidemiological Factors Associated with Hyperendemic Dengue Haemorrhagic Fever in Maracay City, Venezuela. Dengue Bolletin 2002, 26, 84–94.

[21] 21. Patz, J.A.; Martens, W.J.; Focks, D.A.; Jetten, T.H. Dengue Fever Epidemic Potential as Projected by General Circulation Models of Global Climate Change. Environ. Health Perspect. 1998, 106, 147–153. [CrossRef] [PubMed]

[22] 22. Caprara, A.; de Lima, J.W.O.; Marinho, A.C.P.; Calvasina, P.G.; Landim, L.P.; Sommerfeld, J. Irregular Water Supply, Household Usage and Dengue: A Bio-Social Study in the Brazilian Northeast. Cad. De Saúde Pública 2009, 25, S125–S136. [CrossRef]

[23] 23. Brightmer, M.I.; Fantato, M.G. Human and Environmental Factors in the Increasing Incidence of Dengue Fever: A Case Study from Venezuela. GeoJournal 1998, 44, 103–109. [CrossRef]

[24] 24. Andersson, V.O.; Birck, M.A.F.; Araujo, R.M. Towards Predicting Dengue Fever Rates Using Convolutional Neural Networks and Street-Level Images. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018.

[25] 25. De Abreu, F.V.S.; de Andreazzi, C.S.; Neves, M.S.A.S.; Meneguete, P.S.; Ribeiro, M.S.; Dias, C.M.G.; de Albuquerque Motta, M.; Barcellos, C.; Romão, A.R.; de Magalhães, M.A.F.M.; et al. Ecological and Environmental Factors Affecting Transmission of Sylvatic Yellow Fever in the 2017–2019 Outbreak in the Atlantic Forest, Brazil. Parasites Vectors 2022, 15, 23. [CrossRef]550. [CrossRef]