



Temporal and Spatiotemporal Arboviruses Forecasting by Machine Learning: A Systematic Review

Clarisse Lins de Lima¹, Ana Clara Gomes da Silva¹, Giselle Machado Magalhães Moreno², Cecília Cordeiro da Silva³, Anwar Musah⁴, Aisha Aldosery⁴, Livia Dutra², Tercio Ambrizzi², Iuri V. G. Borges², Merve Tunali⁵, Selma Basibuyuk⁵, Orhan Yenigün⁵, Tiago Lima Massoni⁶, Ella Browning⁷, Kate Jones⁷, Luiza Campos⁸, Patty Kostkova⁴, Abel Guilhermino da Silva Filho³ and Wellington Pinheiro dos Santos^{9*}

¹ Nucleus for Computer Engineering, Polytechnique School of the University of Pernambuco, Poli-UPE, Recife, Brazil, ² Department of Atmospheric Sciences, IAG-USP, University of São Paulo, São Paulo, Brazil, ³ Center for Informatics, Federal University of Pernambuco, CIn-UFPE, Recife, Brazil, ⁴ Centre for Digital Public Health and Emergencies, Institute for Risk and Disaster Reduction, University College London, London, United Kingdom, ⁵ Boğaziçi University, Institute of Environmental Sciences, Istanbul, Turkey, ⁶ Department of Systems and Computing, Federal University of Campina Grande, Campina Grande, Brazil, ⁷ Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, United Kingdom, ⁸ Department of Civil Environmental and Geomatic Engineering, University College London, London, United Kingdom, ⁹ Department of Biomedical Engineering, Federal University of Pernambuco, DEBM-UFPE, Recife, Brazil

OPEN ACCESS

Edited by:

João Valente Cordeiro,
New University of Lisbon, Portugal

Reviewed by:

Thaddeus Marzo Carvajal,
De La Salle University, Philippines
Alina Deshpande,
Los Alamos National Laboratory
(DOE), United States

*Correspondence:

Wellington Pinheiro dos Santos
wellington.santos@ufpe.br

Specialty section:

This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

Received: 19 March 2022

Accepted: 03 May 2022

Published: 03 June 2022

Citation:

Lima CLd, da Silva ACG, Moreno GMM, Cordeiro da Silva C, Musah A, Aldosery A, Dutra L, Ambrizzi T, Borges IVG, Tunali M, Basibuyuk S, Yenigün O, Massoni TL, Browning E, Jones K, Campos L, Kostkova P, Silva Filho AGd and dos Santos WP (2022) Temporal and Spatiotemporal Arboviruses Forecasting by Machine Learning: A Systematic Review. *Front. Public Health* 10:900077. doi: 10.3389/fpubh.2022.900077

Arboviruses are a group of diseases that are transmitted by an arthropod vector. Since they are part of the Neglected Tropical Diseases that pose several public health challenges for countries around the world. The arboviruses' dynamics are governed by a combination of climatic, environmental, and human mobility factors. Arboviruses prediction models can be a support tool for decision-making by public health agents. In this study, we propose a systematic literature review to identify arboviruses prediction models, as well as models for their transmitter vector dynamics. To carry out this review, we searched reputable scientific bases such as IEE Xplore, PubMed, Science Direct, Springer Link, and Scopus. We search for studies published between the years 2015 and 2020, using a search string. A total of 429 articles were returned, however, after filtering by exclusion and inclusion criteria, 139 were included. Through this systematic review, it was possible to identify the challenges present in the construction of arboviruses prediction models, as well as the existing gap in the construction of spatiotemporal models.

Keywords: digital epidemiology, computational intelligence, arboviruses forecast, machine learning, systematic review, dengue, chikungunya, Zika virus

1. INTRODUCTION

Vector-borne diseases present a major public health challenge for many countries around the world (1–3). Arboviral diseases are diseases caused by arthropod-borne viruses which are viruses that need a vertebrate host and a hematophagous arthropod (the transmitting vector) in order to maintain themselves in nature (4–6). Arboviruses transmitted by *Aedes aegypti*, e.g., manage to maintain themselves in nature through a human-mosquito cycle. In other words, for the transmission of one of these diseases, it is only necessary for the hematophagous arthropod to inject its infectious saliva into the blood of a non-viremic individual at the time of the bite.

However, non-vertical transmission is also possible, such as during sexual intercourse, from mother to child during pregnancy or childbirth, in addition to transmission of blood, bone marrow, and organ transplantation (6).

Since arboviruses are part of the Neglected Tropical Diseases (NTDs) group, they impact directly and indirectly the countries wherein they are endemic (7). The direct impact is related to the number of people infected and the number of deaths caused by arboviruses. On the other hand, the indirect impact is more associated with socioeconomic impacts (7). Dengue, Zika, and chikungunya fever, transmitted by *Aedes* mosquitoes, are examples of diseases that belong to the group of NTDs. According to the World Health Organization, dengue fever is present in more than 100 countries around the world. Furthermore, in the last decade, there has been an increase of around 300% in the number of cases of the disease (2). Chikungunya, in turn, has been identified in more than 60 countries since 2004, when it first spread to countries in Europe and the Americas (8), whereas the Zika virus is currently present in a total of 86 territories around the world (9). Thus, the arboviral diseases rapid global spread amplified the challenges faced by the scientific and governmental communities (10).

The arboviruses dynamics are associated with several heterogeneous factors that involve demographic, climatic, and environmental aspects of a region. Demographic changes arising from intense migratory flows from rural to urban areas have led cities to grow inordinately. The swelling of urban populations along with urban population mobility associated with other factors, such as poor sanitation, also plays an important role in transmission vector proliferation. In addition, the lack of water distribution, as well as the difficult access to health systems, also bring barriers to controlling the vector (3, 11, 12). Another aspect associated with arbovirus dynamics is the local climatic and environmental conditions. Luminosity, rainfall, relative humidity, and temperature, act directly on the mosquitoes' development and interfere with the eggs' hatch, as well as their lifetime and dispersion (3, 11, 13).

With climate change and the increase in the number and frequency of international flights, two new arboviruses transmitted by the *A. aegypti* mosquito have emerged in Brazil: Chikungunya and the Zika virus. Raising, in this way, new challenges regarding the control and monitoring of the vector (14–19).

Hence, considering the impact caused by the vector-borne diseases, several research groups have directed their efforts to understand the dynamics of arboviruses through mathematical and computational models for the creation of prediction models (3, 20, 21). We believe that prediction models can be a good tool for health authorities to implement public policies for rapid monitoring and control of the arboviruses spread. Therefore, this document proposes a systematic review of the literature to identify models for predicting arboviruses cases transmitted by the *A. aegypti*—dengue fever, Zika virus disease, and chikungunya—as well as the mosquito dynamics. In particular, this review seeks to answer the following research questions. In particular, this systematic review seeks to analyze what are the biggest challenges when it comes to implementing arboviruses

prediction models. In addition, we sought to identify the main techniques for predicting mosquito cases or foci and which are the main variables that interfere in the dynamics of disease transmission and the dynamics of the transmission vector.

2. METHOD

The strategy for conducting this systematic review is detailed in **Figure 1**. First, we performed an automatic search in scientific databases, such as IEE Xplore, PubMed, Science Direct, Springer Link, and Scopus. We searched for articles published between 2015 and 2020 wherein the metadata, titled or abstract contained the terms defined in the following search string: ["Arboviruses" OR "arthropod-borne virus" OR "dengue" OR "chikungunya" OR "mosquito-borne disease"] AND ["Machine Learning" OR "Deep Learning" OR "neural network" OR "artificial intelligence"] AND ["forecast" OR "prediction"].

In the following step, we identified the number of articles that were retrieved from each scientific database. We then checked if the articles met the exclusion criteria. In this review, we excluded works that were not in English, works that were not completed, and documents classified as posters, tutorials, editorials, or calls for articles. We also excluded works that did not include arbovirus or breeding site prediction and works that did not include computational techniques.

After filtering according to the exclusion criteria, we briefly read the article's abstract, introduction, and conclusion. This step was essential in order to select the articles according to the inclusion criteria. The works selected in that phase were those which met at least one of the following criteria:

1. Works with computational intelligence methods to predict arboviruses cases.
2. Works with computational intelligence methods to predict mosquito breeding sites.
3. Works with computational intelligence methods to predict the mosquitoes' dynamics.
4. Works with statistical learning (Bayesian and other probabilistic methods).
5. Works involving forecasting with differential equations.

The remaining articles after filtering by the inclusion criteria were fully read and evaluated according to the quality criteria described in **Table 1**. We used a 0–1 scale to assess study quality, where Yes (Y) = 1; Partially (P) = 0.5, and No (N) = 0. Three reviewers performed articles assessment, independently, and the disagreements were resolved by discussion among the reviewers.

From the articles selected by the inclusion criteria, we extract the following information: the title of the article, the name of the authors, the institution, the application of the study, the methodology applied to the study, the prediction model, results, the advantages, and the disadvantages of the method.

3. RESULTS AND DISCUSSION

The search process returned 51 articles from IEEE Xplore, 95 articles from PubMed, 238 from Scopus, 20 from Science Direct,

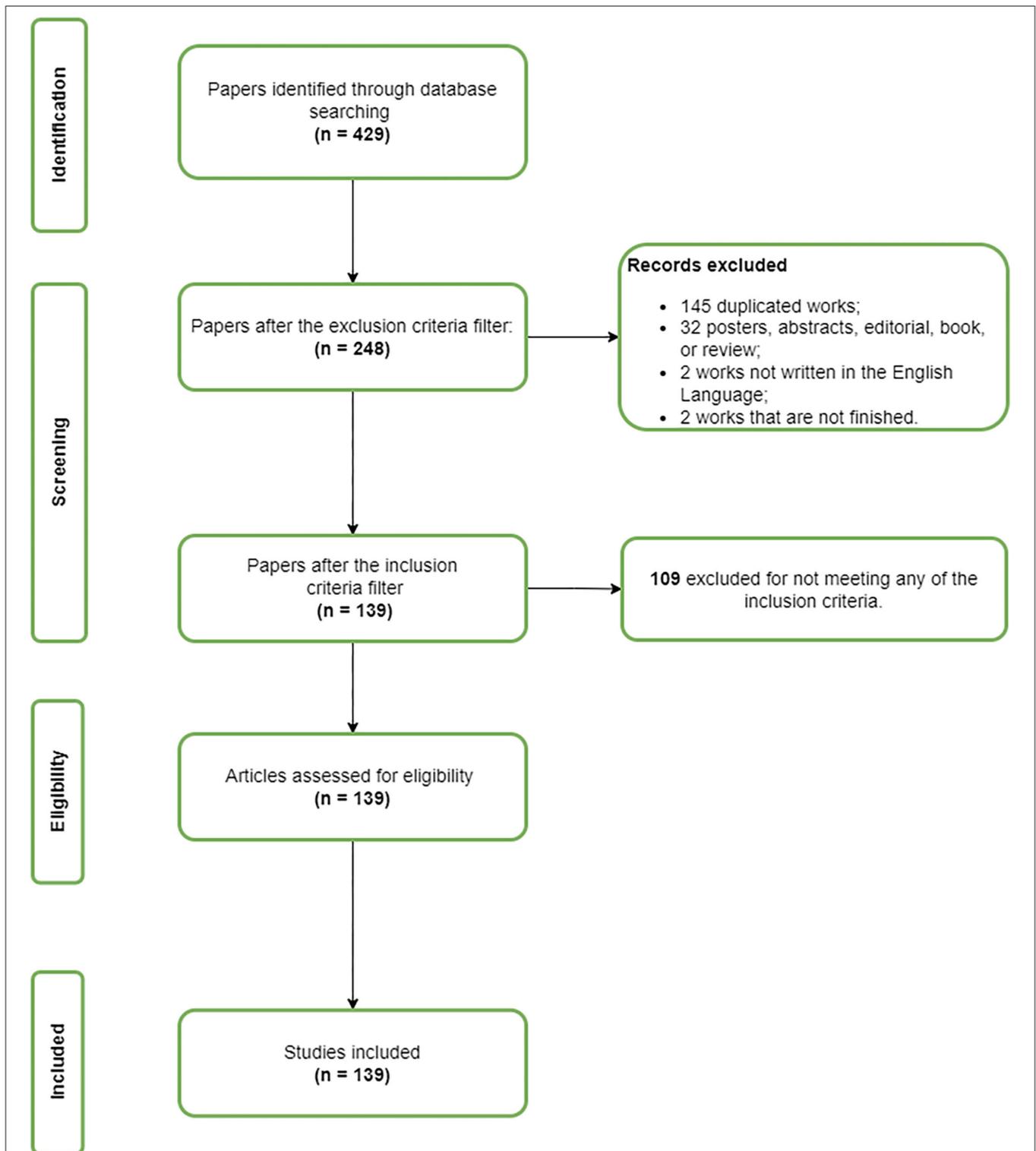


FIGURE 1 | This system consisted of the following steps: (1) First, we performed a search of scientific databases (IEEE Xplore, PubMed, Scopus, Science Direct, and Springer Link). (2) We then filtered the returned articles according to the exclusion criteria. (3) In the next step, we selected the article that remained from the previous stage according to the inclusion criteria. (4) After completing the previous step, we read, evaluated, and summarized the studies included in the review. (5) In the last step of this review, we grouped the studies considering their common characteristics.

TABLE 1 | Quality criteria used to evaluate the selected studies.

ID	Quality criteria	Answer
QC1	Are the objectives clearly stated?	Y/P/N
QC2	Are the data sources clearly described?	Y/P/N
QC3	Do the authors present the variables to build their models?	Y/P/N
QC4	Do the author explicitly defined which computational techniques or prediction model they used as well as their architectures and parameters?	Y/P/N
QC5	Do the authors report which metrics they used in order to evaluate their models?	Y/P/N
QC6	Are the conclusions coherent to the study findings and also with the set objectives?	Y/P/N
QC7	Do the authors detail the weakness of their work?	Y/P/N

and 25 from Springer Link. It is important to emphasize that, for the Science Direct database, the search string had to be reduced. For this database, the number of Boolean operators in the original string search was not supported. In this case, we used the terms: (“dengue” OR “zika” OR “chikungunya”) AND (“Machine Learning” OR “artificial intelligence” OR “regression”) AND (“forecast” OR “prediction”). From the 429 works collected, 181 were excluded in the filtering by the exclusion criteria stage. Among these 181 articles, 145 were duplicated studies, 32 were posters, abstracts, books, proceedings, or systematic literature review. In addition, two of them were excluded because they were not in English, and two articles were unfinished. We then screened the remaining 248 studies by reading the title, abstract, and conclusion. After the inclusion criteria stage, 109 were removed from this study for not meeting any of the inclusion criteria. Hence, 139 articles were included in this systematic review.

In the last step of the systematic review, we grouped the 139 selected articles according to their common characteristics (Table 2). The studies were divided into six groups: Arboviruses (counts) prediction (Group 1), Arboviruses detection (Group 2), Outbreaks and Risk prediction (Group 3), Models of mosquitoes dynamics, breeding sites models (Group 4), Clustering, modeling, and spatiotemporal prediction of arboviruses (Group 5), and Other Approaches (Group 6). In Group 1, we considered only the studies that presented models for counting arboviruses. In Group 2, we only included the studies that involved arboviruses detection. Group 3, in turn, is composed of studies that present models for predicting arbovirus outbreaks, as well as predicting the risk of an outbreak. The studies that presented vector monitoring and prediction models were included in Group 4. Those articles that investigated arboviruses prediction models with a spatiotemporal approach were included in Group 5. Finally, the studies that presented more than one of the approaches mentioned above—or that did not fit into any of the previous groups—were included in Group 6.

3.1. Arboviruses (Counts) Prediction

Among the 139 selected studies, about 80 studies are related to the prediction of the incidence of arboviruses cases (Table 2).

TABLE 2 | Number of studies per group, considering the following stratification: Group 1: prediction of arboviruses by counting; Group 2: detection of arboviruses; Group 3: prediction of risk and epidemiological outbreaks of arboviruses; Group 4: modeling the dynamics of mosquitoes and breeding sites; Group 5: spatio-temporal modeling; Group 6: other approaches.

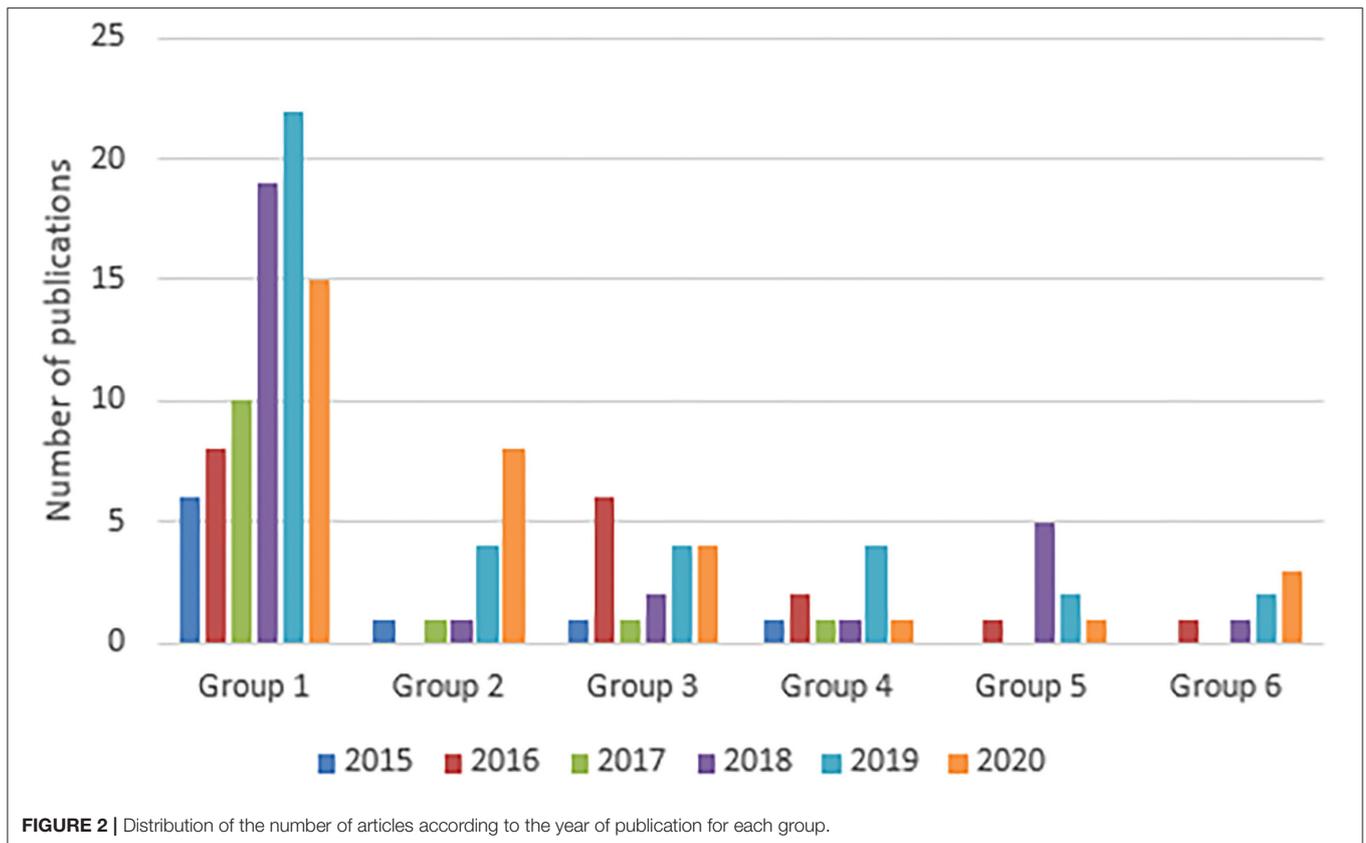
Group	Description	Number of studies
Group 1	Arboviruses (count) prediction	80
Group 2	Arboviruses detection	15
Group 3	Arboviruses Outbreaks and Risk prediction	18
Group 4	Models of mosquitoes dynamics, breeding sites models	10
Group 5	Clustering, spatiotemporal modeling	9
Group 6	Other approaches	7

Considering the year of publication, we observed that most of the studies in this group were published in 2018 and 2019. The number of articles published in 2018 and 2019 was 19 and 22, respectively, with a drop in the number of publications related to this topic in the year 2020 (Figure 2). Regarding the scores referring to quality criteria, we noticed that most scores were high, with the exception of QC7. For this criterion, the average score achieved by the studies was 0.33 (Figure 3).

The *A. aegypti* is the transmitter vector of three different type arboviral diseases. Taking into account the types of arboviruses transmitted by this mosquito, we found a significant amount of work focused on the construction of dengue fever transmission models. In these studies, the authors, in most cases, do not distinguish the serotype of the disease. In other words, dengue cases are generally considered as: dengue fever, dengue hemorrhagic fever and dengue shock syndrome, including local and imported cases. However, the studies of (22–27) are only focused on prediction models for dengue hemorrhagic fever. Regarding the other two diseases transmitted by the *Ae. aegypti*, we found a small number of articles addressing Zika virus disease and chikungunya’s numeric prediction models (28–33).

The returned studies also brought a great variety related to the attributes used to build arboviruses prediction models. It is noted that, in several studies, prediction models are built taking into account only past values of disease cases (17, 25, 29–31, 34–38). However, arboviruses are diseases that need a transmitting vector for the arbovirus cycle in nature to complete. Furthermore, climatic factors directly affect the life cycle of the transmitting mosquito. In this context, several studies have investigated prediction models considering the effect of climatic and environmental variables on arbovirus transmission. Therefore, we observed a wide variety of studies that used at least one of the following variables as model attributes: temperature, rain, and relative humidity. However, some studies included other parameters in their models such as the number of rainy days (39–41), number of stormy days, and wind speed (41).

Furthermore, environmental variables obtained through remote sensing were also explored in a relevant number of studies. The most common were normalized difference vegetation index (NDVI) (42–44), vegetation index (45),



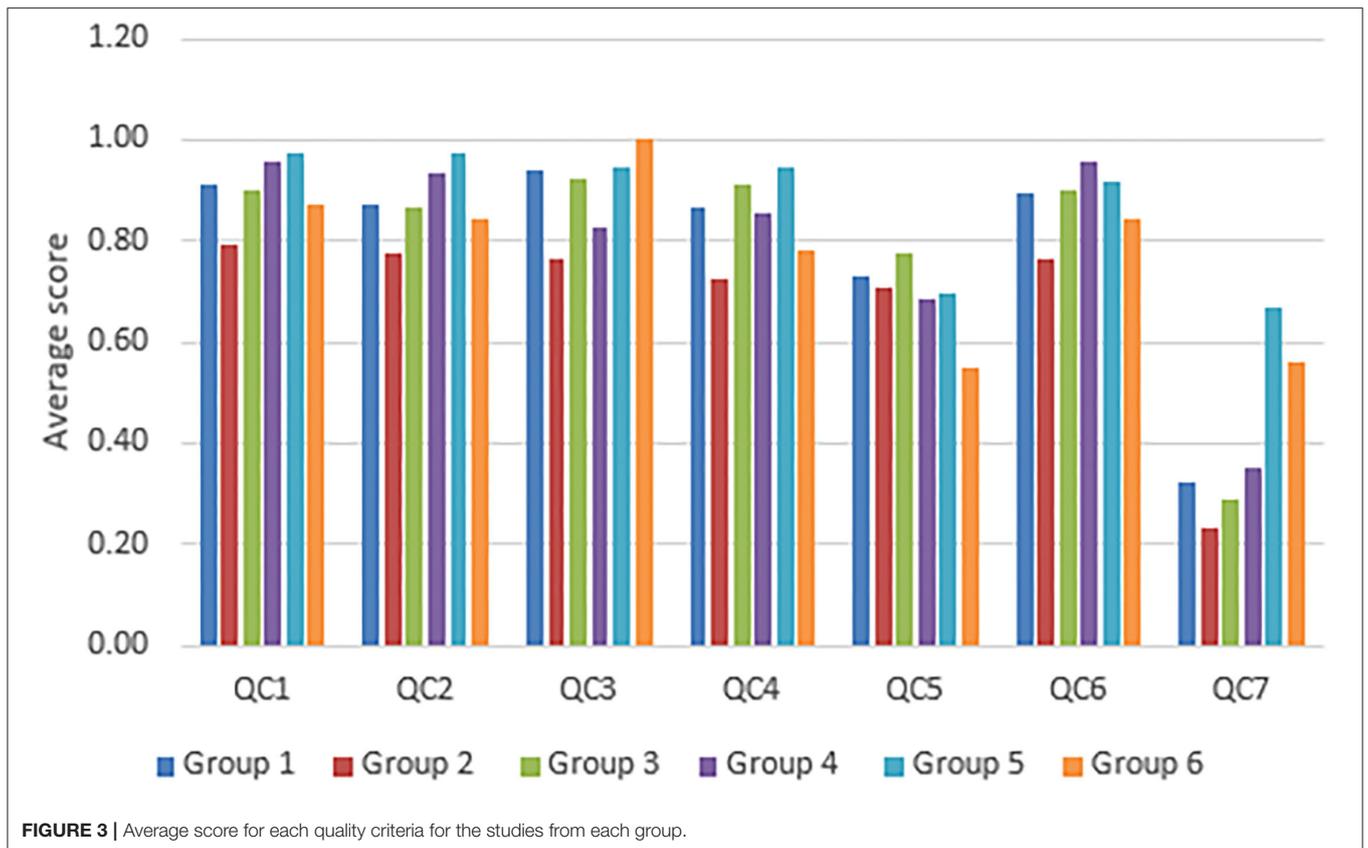
enhanced vegetation index (46), smoothed vegetation index, smoothed brightness temperature index, vegetation condition index, vegetation health index (44), land surface temperature (43, 46, 47), Southern Oscillation Index (SOI), and Sea Surface Temperature Anomaly (SSTA) (48). In the studies of (47) and (44), the authors included information on the EL Niño phenomenon as well as (47)—that included variables related to the El Niño Southern Oscillation Index—and (44)—that included the Oceanic Niño Index variable in their model.

The research groups also explored attributes other than climate variables, such as epidemiological surveillance variables and sociodemographic variables. Among the epidemiological surveillance variables, the most used were: the number of larva-free, house index (39, 49), weekly breeding percentage (50), container (49), breteau index (49, 51, 52), standard space index, adult mosquito density, *Ae. aegypti* larvae infection, and female mosquito infection rate (52). Mosquito dynamics interfere with arbovirus dynamics. Including this information in predictive models can be an outlet for the search for more robust models that can understand the arboviruses' dynamics in a given region. Sociodemographic aspects also influence the arboviruses dynamics. Considering this fact, Dharmawardana et al. (45) also implemented in their model a mobility model in order to predict the dengue cases's incidence curve. Still considering sociodemographic information, other researchers included in their models' population density (32, 40), poverty percentage (32), population (41, 46, 53), Gini Index—a

measure of income inequality—, education coverage (24), and unavailability of the garbage dump. In the model developed by (50), the population size attribute was considered for both the resident population and non-resident foreign population. Models considering sociodemographic factors can help us to understand how population dynamics are related to arboviruses cases. In this way, it can help to guide socio-educational actions and to direct the implementation of basic sanitation and infrastructure policies.

Continuing the analysis regarding the variables included in the prediction models, we observed that data from social media and search volume reported by search engines can be a powerful tool in monitoring arbovirus-borne diseases. In the study of (54), data from Baidu (a popular search tool in China) and social media are used to model the incidence of dengue in Guangzhou, China. Data referring to the number of comments, number of likes, and number of forwarding that are associated with dengue as a primary keyword are captured. In the studies of (30) and (27), the authors use Google Trends data to generate models for predicting Zika and dengue hemorrhagic fever, respectively. Espina and Estuar (36), in turn, use Twitter data to identify infodemiological content to be used in predicting dengue. In a world where information is gaining speed at every moment, the implementation of arbovirus models using social media can be an alternative for monitoring, surveillance, and disease prediction.

Taking into account the datasources used by the authors, we identified that in a vast majority of the returned studies,



the data were obtained through government institutions. These institutions were responsible for either epidemiological surveillance or meteorological monitoring of the study area. One of the limitations presented in studies that use government data is the underreporting cases (55). Usually, when the individuals do not have the most severe form of the disease, they do not seek health services. Hence, under these conditions, those individuals are not included in the statistics. Moreover, health data usually have other limitations such as missing values, e.g., (55). However, some works use alternative sources to obtain data. Data can also be obtained through social media and search engines (27, 29, 30, 36, 54) and through data from the WHO (29, 31). On the other hand, we observed that in some studies, the authors do not explicit the origins of the collected data (37, 39, 43, 44, 56–60). The lack of information regarding the datasources can affect the study's reproducibility since the databases' original conditions to generate the models are not clear.

When we evaluated the studies regarding the types of models used in the predictions, we observed that the vast majority of authors investigated moving average models (27), such as the Autoregressive Integrated Moving Average (ARIMA) (17, 23, 29, 35, 41, 43, 46, 56, 61–63), Seasonal Autoregressive Integrated Moving Average (SARIMA) (55, 63–66), Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) (67). Several works have also presented a wide variety of models using artificial neural networks, mainly the LSTM (59, 68–70). But models using backpropagation neural

networks (BPNN), GANN networks (60), Elman Recurrent Neural Network Levenberg Marquardt Algorithm (ERMN/LMA) (22), and Deep feed-forward neural networks (28) were also investigated. Although neural networks have been extensively explored, in many studies, the authors did not explain the type of network they were investigating (23, 45, 46, 61, 66, 71, 72).

When working with prediction, we also prioritize the computational cost associated with the implemented technique. In this sense, optimization algorithms can help us to reduce the computational cost by reducing model training time. Optimization algorithms do this by looking for attributes that represent the dataset being studied. Therefore, some studies in this group investigated some optimization techniques. In the study of (57), the authors investigated several optimization algorithms associated with the Least Square Support Vector Machine (LSSVM). The investigated algorithms were: Moth Flame Optimization (MFO), Gray Wolf Optimizer (GWO), Firefly Algorithm (FA), and Artificial Bee Colony (ABC) algorithm. Saptarini et al. (22), in turn, used Genetic Algorithm, as well as (68). Finally, we notice that, when it comes to predicting the count or incidence of arboviruses, there are a wide variety of model applications.

In the articles evaluated by this group, we observed that among the main diseases transmitted by *A. aegypti*, the models for predicting dengue cases are the most explored by research groups. As the diseases transmitted by this vector present similar symptoms in their milder forms, in regions where dengue,

zika, and chikungunya viruses circulate, the models may present errors related to the low distinction between the diseases. The models that included climatic variables and/or variables of sociodemographic aspects performed better than those that only took into account the historical series of confirmed cases of the disease.

As for the origin of the data sources, we observed that the data obtained by governmental institutions provide greater reliability to the models. However, these models have limitations, mainly in relation to the underreporting of cases. Underreporting can impair the performance of case prediction models. The data that are obtained through the analysis of user behavior in social networks can act as an important tool in the prediction of arboviruses. Since, in some regions, the public system may take a long time to update notifications, alternative data sources can be a good solution in the development of robust models, especially in critical situations such as during arbovirus outbreaks. On the other hand, models generated using data solely from social networks may not be applied in regions where access to the Internet and mobile devices are scarce. That is, in some peripheral regions, the model may not be able to identify disease cases in that region.

Regarding the prediction models selected, we observed that most of the studies that used Artificial Intelligence opted for deep learning models. Despite the promising results that were obtained, the use of deep networks, such as LSTM, is linked to large memory consumption. In other words, it takes a lot of training time and resources to create applications for the real world. Moving average models, in turn, are good tools for capturing trends, periodic changes, and random distortions in historical series. In addition, they are simple and quick to apply.

Thus, the models of the historical series are very relevant and can be a very useful tool in the planning of public policies to combat arboviruses. However, these models are not able to provide information regarding the spatial distribution of diseases. That is, they are not able to point out which areas are being more or less affected by diseases transmitted by *A. aegypti*.

3.2. Arboviruses Detection

For this group, 15 studies were selected from the 139 included in this review (Table 2). The publication years for this group varied between 2015 and 2020, wherein the majority of the studies were published in 2019 and 2020 (4 and 8 articles, respectively). For the years 2015, 2017, and 2018, there was only one publication on this topic (Figure 2). Considering the quality criteria evaluated, the vast majority reached a low score in QC7 as presented in Figure 3.

Among these 15 articles in this group, we noticed that they focus on two of the three arboviral diseases transmitted by the *A. aegypti*. That is, 12 articles focused on dengue fever prediction models, whereas three of them focused on Zika virus disease. It is also important to highlight that, in all articles, the authors used machine learning algorithms in order to build their prediction model.

For Zika virus disease prediction, we noticed that the authors investigated different algorithms to predict positive cases of the disease. Jarrin et al. (73) evaluated support vector

machines (SVM) and logistic regression to build their models, whereas Jarrin et al. (74) and Mahalakshmi and Suseendran (75) investigated Random Forest and Multilayer Perceptron (MLP) algorithms, respectively.

Jarrin et al. (73) investigated SVM and RL algorithms—implemented in Python 3.7—to classify individual samples into “infected” or “uninfected” with the Zika virus. According to the author’s results, the classifier showed a better accuracy for the “infected” class. The method presented by (73) can be used for the early diagnosis of ZIKV infection. Jarrin et al. (74), on the other hand, used mass spectrometry approaches to detect ZIKV by RT-PCR using RNA samples extracted from serum and urine to classify the diagnosis. The problem presented by (74) was modeled using Random Forest using MATLAB R2017a. The model presented by the authors is a robust platform that can be implemented in routine laboratories in order to help to support the diagnosis. Mahalakshmi and Suseendran (75) used the Multilayer Perceptron (MLP) artificial neural networks classifier. The data used is synthetic and was collected from the Internet. For prediction, the Weka software (version 3.8) was used. As the study was carried out with synthetic data, it is essential that tests be carried out with data from real databases. Having been trained only with artificial instances, when coming into contact with a real-world dataset, the accuracy of the generated model can drop significantly.

To predict dengue, the selected studies used different predictors. Mello-Román et al. (76) have developed a system in which data collection is based on the symptoms of the disease. The dataset is composed of cases registered by the Paraguayan health system, e.g., patients admitted due to fever and complete dengue diagnosis. Mello-Román et al. (76) carried out their tests using the IBM SPSS Modeler software in order to train their MLP and SVM algorithms. According to the author’s results, the MLP showed better accuracy over the SVM classifiers.

(77) provide a prediction of the types of dengue cases. In order to investigate the best classifier, the authors evaluated the Decision Tree (DT) and Random Forest (RF) algorithms. Ho et al. (78) group explored a different method to speed up dengue diagnoses in the laboratory. The authors analyzed the decision tree (DT), deep neural network (DNN), and logistic algorithms. In this way, through the clinical parameters identified in the study, it is possible to help with the burden of laboratories for the diagnosis of dengue. Alam et al. (79) in their approach bring a prototype of a new framework for analyzing biomedical data called biocloud. The data gathered on this framework is modeled with a support vector machine to classify the disease’s cases. This type of technology can provide services at a low cost and can be used in remote areas.

Ganthimathi et al. (80) developed an early dengue diagnosis system using Artificial Intelligence. In their research, Ganthimathi et al. (80) investigated two separate machine learning algorithms: support vector machine as well as k-nearest neighborhood. According to Ganthimathi et al. (80)’s findings, both algorithms presented good performances, however, the SVM showed superior performance compared to KNN’s performance. Kapoor et al. (81) also treated the dengue prediction problem as a classification problem. Thus, in the

study, Kapoor et al. (81) investigated four different classifiers, namely Random Tree, Random Forest, Support Vector Machine (SVM), and artificial neural networks. An interesting aspect of (81)'s model is that they used as their model's not only demographic information but also symptomatological data and clinical trial reports. Ariffin and Aris (82), in turn, created a system to help individuals in the self-diagnosis of dengue cases. As a classifier, the authors used artificial neural networks. As shown by the results obtained by (82), the model developed achieved high reliability for detecting the disease. Despite being a disease that helps in self-diagnosis, it is important to emphasize that medical guidelines are not dispensed with using the tool. Dharap and Raimbault (83) brought a different approach from the others studies commented so far. In their approach, Dharap and Raimbault (83) assessed the effectiveness of medical hematology analyzers that flags arboviruses' presence in blood samples. The machine learning algorithms used in their study were regression and Random Forest. With their results, they demonstrated that it is possible to screen arboviruses infection using a low cost, but also an effective predictor.

Srivastava et al. (84) bring a classification of dengue using online learning. Thus, learning takes place with just a few training examples. No retraining of the model or redeployment of the prediction engine is required. The following algorithms were used: Adaptive Regularization of Weights (AROW) and its Variants, Gradient Descent Online (OGD), Confidence Weighted Learning (CW) and Soft Variants (SCW 1 and SCW 2), Normalized HERD (NHERD), Passive Aggressive (PA) and its variants PA1 and PA2, Improved Ellipsoid Method (IELLIP), Approximate Large Margin Algorithm (ALMA), Second Order Perceptron and Perceptron (SOP), Relaxed Online Maximum Margin Algorithm (ROMMA), and Aggressive Romma (AROMMA). The evaluation of the classifiers was done offline in the Weka software with SVM and RF, and later, the classifiers were evaluated online. Additionally, this system is a health system helping to signal patients with a high probability of being diagnosed with dengue. Sasongko et al. (85) focused on finding the best backpropagation algorithms for early detection of dengue with the addition of multilayer perceptron (MLP) optimization through five algorithms. The backpropagation algorithms used were Gradient Descent (GD), BFGS Quasi-Newton (BQN), Conjugate Gradient Descent—Powel (CGD), Resilient Backpropagation (RB), and Levenberg Marquardt (LM). Additionally, the Levenberg Marquardt algorithm proved to be the best for detecting dengue. In other words, this algorithm solves the data outlier problem well.

Iqbal and Islam (86)'s group performed a performance evaluation of different dengue outbreak prediction classifiers. The methods were evaluated by eight different performance parameters. Iqbal and Islam (86) evaluated K-nearest neighbor (kNN), Support vector machine (SVM), Artificial neural network (ANN), Naive Bayes classifier, Decision tree, and Logistic regression classifier (LogitBoost) algorithms. The experiments were carried out with the Weka learning software. Among the trained algorithms, according to the authors, the one with the best performance was LogitBoost. This classifier had the best classification accuracy, sensitivity, and specificity metrics.

Balamurugan et al. (87) created a classifier for detecting dengue cases based on combinatorial characteristics based on weighted entropy scores based on ideal classification. The algorithms used to extract the most important attributes were Correlation based Feature Selection (CFS), Genetic Algorithms (AG) and Particle Swarm Optimization (PSO), in addition to the Optimized Classification Algorithm based on Weighted Entropy Score (EWSORA). Finally, the data were submitted to conventional classifiers such as Naïve Bayes, J48, Multilayer Perceptron (MLP), and Support Vector Machine (SVM). For evaluation, the Weka software was used. As metrics to evaluate the best models for predicting dengue, accuracy, true positive rate, precision, Recall, F Measure, and ROC were used. After applying the Genetic Algorithm (GA), Particle Swarm Algorithm (PSO), and Correlation-Based Resource Selector (CFS) algorithms for resource selection, the J48 and MLP classifiers proved to be better. EWSORA has greatly improved the accuracy performance for several classifiers, mainly for Genetic Algorithm (GA), Particle Swarm Algorithm (PSO), and Correlation Based Resource Selector (CFS).

3.3. Arboviruses Outbreaks and Risk Prediction

Among the studies evaluated in this systematic review, we observed that 18 articles were related to the prediction of the occurrence of arboviruses outbreaks, or the prediction of the risk of the occurrence of disease outbreaks (**Table 2**). Taking into account the years of publication of the articles, it is observed that most were published in 2016, as shown in **Figure 2**. Regarding the quality criteria, the studies achieved scores above 0.7, as shown in the graph in **Figure 3**. On the other hand, it is important to highlight that among the evaluated studies, the average QC7 score was quite low. In other words, most of the studies did not explicitly state the limitations of the investigated models.

Considering the types of arboviruses, we observe that the vast majority of the studies evaluated focused on the prediction of outbreaks or risk of dengue fever (18, 88–99). Two of the studies involving dengue risk prediction or dengue outbreak were focused on only one dengue serotype (dengue hemorrhagic fever) (89, 90). However, Brett and Rohani (95) show in their study an approach using each serotype for their prediction. In the study of (100) and (101), the authors investigated models for predicting the risk of Zika virus outbreaks, while (102) took into account all cases of arboviruses (dengue, Zika virus, and chikungunya) in their model.

As for the variables used to generate the prediction models, we observed that several studies included the climatic variables (90) where the most common are temperature (18, 91–94, 96, 98, 100, 102), rainfall (18, 91–94, 96–98, 100). Other climatic variables appear less frequently, such as wind speed (90, 91), vapor pressure (100), sunshine (91), atmospheric, and SST predictors (97). Regarding the predictors used for model generation, we observe a greater variety of predictors that are not associated with climate variables, when compared to the predictors used in Section 3.1. The population density, number of travelers, temperature, health expenditure per capita, gross domestic product per capita, water

coverage, ZIKV transmission in nearby countries were examples of predictors used in the study of (100). In contrast, Akhtar et al. (101) used gross domestic product per capita, physicians per 1,000 people, and beds per 1,000 people, population densities, in addition to Zika cases. Predictors based on transmitter vector monitoring data have been extensively explored, such as mosquito occurrence (100), breteau index, and ovitrap index (98). It is important to highlight that two of the studies evaluated did not clarify which variables were used in the investigated prediction models.

Regarding the data sources, in a considerable amount of works, the authors usually obtained their databases through local government data sources (90–94, 96, 98, 99, 102). However, some works obtained their databases through other international sources, such as the US National Oceanic and Atmospheric Administration (95), Pan American World Health Organization (PAHO), International Air Transport Associate, World Bank, US Bureau of Economic Analysis (101), and World Health Organization (WHO) (103). Two of the works included in this group did not present information related to the origins of the data sources obtained. In addition, no works were found that used alternative sources such as data originated by means of search engines, as well as data generated through social networks.

Finally, for risk predictions or prediction of arbovirus outbreaks, we found several approaches. Models were investigated using artificial neural networks (89, 101, 102, 104), decision trees (89, 99), gradient boosting regression tree (GBRT) (100), naïve Bayes (89), extreme learning machines (90), Least Absolute Shrinkage and Selection Operators (LASSO) and Ridge (92), support vector machines (SVM) (18, 91, 93). Moreover, early warning signals (EWs) derived from the theory of critical slowing down (95), the Shewhart model (98), population loss value at risk model (103) were also investigated.

In the articles evaluated for this group, we also observed that dengue was the arbovirus that received the most attention in terms of creating models for predicting outbreaks or disease risk. In the models of the studies, the climatic variables of temperature and precipitation are the predictors that appear most frequently in the prediction models. However, regarding predictors that are not related to climate variables, there is no assessment of which factors most impact the performance of the prediction model. That is, none of the studies presented the performance of models with different predictor configurations in a comparative way.

As for the types of models used, we observed that most studies used non-deep machine learning algorithms to generate the prediction models. Despite the promising results, it is difficult to indicate which algorithms had the best performances. The authors used different predictor configurations for the different models, which made it difficult to carry out a more in-depth analysis of the types of models used.

3.4. Models of Mosquitoes Dynamics, Breeding Sites Models

Of the 139 selected articles, 10 were predicted with vector control (Table 2). The years of publication range from 2015 to 2020. For the years 2015, 2016, 2017, 2018, 2019, and 2020, 1, 2, 1, 1, 4,

and 1 were selected, in this order (Figure 2). Analyzing Figure 3, we found that the articles scored relatively low on quality criteria 5 and 7. Among these, 7 developed models based on machine learning and 5 based on statistical methods.

Haddawy et al. (105) featured a pipeline design to detect mosquitoes' breeding sites using geotagged images with a machine learning approach. In Haddawy et al. (105)'s model, they use container count with resultant in order to create container density maps. The relationship between the densities of the eight types of recipients and the larval survey data was calculated using multivariate linear regression and obtained good precision. For object recognition, Haddawy et al. (105)'s group evaluated a convolutional neural network (R-CNN). Thus, creating geotagged container density maps is favorable for providing large-scale detailed hazard maps.

Raja et al. (106) developed early *Aedes* outbreaks prediction models using a machine learning approach. In order to build the prediction models, they used temperature, precipitation, start date notification, and notification date, as well as vector indices such as *Aedes albopictus*, *A. aegypti*, and larvae count.

Raja et al. (106) ran experiments using Bayesian Network Models method in order to create the prediction models. The system interface was implemented in C++ (backend) and the frontend implemented in JavaScript, CSS, and HTML5. The system is able to make predictions considering a 7 days horizon.

Asmai et al. (107)'s proposal was to create a Mobile Application for the Intelligent Detection of Mosquito Larvae (iMOLAP). The mobile app uses the convolutional neural networks (CNN) method, which is the Inception V3 model. The image that is captured is compared to a collection of predefined images to measure accuracy. Therefore, iMOLAP can classify *Aedes* and larvae species by imaging, and detecting the affected area of the site. This application can be a very important tool to assist in the surveillance and combat of mosquitoes. Lee et al. (108)'s, on the other hand, focused on developing a model to predict mosquito abundance. Thus, they considered climatic variables such as temperature, air humidity, wind speed, and precipitation as model predictors. The authors evaluated different approaches in order to build their model. They investigated using multiple linear regression (MLR), and artificial neural networks (ANN) algorithms. The correlation between climatic variables was assessed using the cross-correlation function. The metrics used were the correlation coefficients, the RMSE, and the agreement index. The results of models made with ANN were better than the MLR in all metrics. The approach brought by this study is interesting and can be very useful in mosquito monitoring. However, the authors did not describe well the apparatus necessary for collecting data on the number of mosquitoes. They did not describe the ANN configurations evaluated. In addition to not being described the number of tests to obtain a result with statistical significance.

In the study of (109), the authors developed a mosquito washing prediction system for *Aedes* in Recife, Brazil. The authors evaluated several types of regressors to build models to predict the number of properties with the presence of mosquitoes in Recife. Among the regressors, Extreme Learning Machines are Single Layer Feedforward Networks (SLFNs), Fuzzy Extreme

Learning Machine, Bayesian Extreme Learning Machine, Interval Type-2 Radial Basis Function, Neural Network (IT2-RBFNN), and Online Extreme Learning Machine (OLEM). First, the spatial distribution of the number of properties that contained water containers contaminated with *Aedes* mosquito larvae was performed. Then, the spatial distribution of properties with mosquito larvae was performed and stratified by the type of water reservoir. Finally, the models are implemented on the real-time surveillance data. As metrics, percentage RMSE and training time were used. In this way, the prediction system shows the mosquito's hotspots. This study takes a spatiotemporal approach, so research can help managers by giving direction to location-based mosquito population control policies, helping to limit transmission to humans.

Bennett et al. (110) brought a mosquito classification to detect *A. aegypti*. The database was created by the authors themselves. They collected samples of larvae present in garages that traded used tires in Panama. Additionally, with mass spectrometry, the types of larvae are identified. Finally, using the Supervised Neural Network (SNN) a classifier is built to identify the type of mosquito present. The model created had a very high capacity for recognizing and classifying training data. This study brings a look at the garages, which can be a strategic point for epidemiological surveillance policies.

Considering the statistical models, we highlight the study of (111). Their group has developed time prediction models for *A. aegypti* oviposition. Both model validation and application were applied in the dengue outbreak in 2016. For this purpose, time series of MODIS (moderate resolution image spectroradiometer sensor) products of normalized difference vegetation index and daytime surface temperature were created. The MODIS model consists of: (1) linear regression modeling and (2) the creation of two models, one with and one without lag times on the independent environmental variables. The environmental variables were standardized and the developed models were compared using the Akaike Information Criteria (AIC) to determine the ideal model in terms of goodness of fit and number of parameters. The model without latency was the best. Both models developed in this article showed that MODIS environmental variables (NDVI and LST) are good predictors because both environmental variables are present in both models, providing acceptable fit and validation results. We can understand that the NDVI increment may be due to precipitation in the near past followed by an increment in the vector activity which is verified by the increments of the oviposition activity. Furthermore, a model based on MODIS has the possibility to envision an operational forecast program at national level.

Estallo et al. (112) created a prediction model evaluating the weather variability associated with the seasonal fluctuation of the oviposition dynamics of *A. aegypti* in a City of Orán, Argentina. To create the model, precipitation data, photoperiod, water vapor pressure, temperature and relative humidity (maximum and minimum) and ovitrap sampling were used. A multiple linear regression analysis was performed with the set of meteorological variables considering the time lag that correlates with oviposition. And the model is validated. The prediction model created allows the prediction of the growth or decrease of the ovitraps activities

of *A. aegypti* based on meteorological data. The prediction of these activities can be predicted three or 4 weeks in advance. Because this model brings a more localized and comprehensive assessment with site-specific data that can be used in disease prevention policies.

Hettiarachchige et al. (113) built a data transmission risk prediction model based on high resolution meteorological data. Additionally, this risk is predicted through vector prediction. Routine entomological surveillance data for dengue and meteorological data from a prediction system with high spatial and temporal resolution were used. The risk prediction system was divided into two stages to assess dengue transmission *via A. aegypti*. In the first, logistic regression was used to determine the presence or absence of larvae in the sites of interest using climatic attributes as explanatory variables, and then used a bootstrap approach in an administrative division. In the second, with the negative binomial model inflated to zero, an estimate of the larvae count of the positive division predicted in the first stage is made, and then positive larvae sites are identified and the number of larvae is predicted. Splitting the model into two stages increases the accuracy of identifying positive larval locations. A benefit for risk prediction in non-homogeneous regions.

da Cruz Ferreira et al. (19) developed a temporal prediction of mosquito infestation based on climatic data and monitoring data from *Aedes*. The climatic variables used were daily rain, temperature (minimum, average, and maximum), and relative humidity, and dengue data were obtained from the Health Department of Porto Alegre. The Generalized Additive Model (GAM) and Logistic Regression methods were used. The first method was used for two models, one was fitted with climatic variables, and the other with climatic variables and mosquito abundance as an explanatory variable. Additionally, the second method was used to assess the effect of adult mosquito infestation on the probability of dengue incidence. The second GAM model predicted the data better than the first. The researchers stated that if the population of *Aedes* is continuously monitored the predictions of the infestation rate will be more reliable. And monitoring this population is important for dengue control in Brazilian cities.

The studies presented here brought several different perspectives to control the *A. albopictus* and *A. aegypti* mosquitoes. Some of the variables considered in these studies were: stratification by type of water reservoir; neglected environments, such as garages that contain tires and other potential breeding sites; local and comprehensive assessment of breeding sites; evaluation of mosquito larvae stages; and the seasonality of the mosquito cycle dynamics. Furthermore, in the construction of the prediction models, different machine learning techniques and statistical methods were used. The models with a broad and more restricted evaluation of the study regions proved to be good and robust in terms of evaluation metrics. Many of these works present scalability and reproducibility for prediction at the national level, relating the magnitude of the population of *Aedes* mosquitoes, the incidence of arboviruses, and the monitoring of this vector. Thus, these approaches can be used to support the implementation of epidemiological surveillance policies. However, some of these studies had limitations, the lack

of clarity and uniformity regarding the evaluation metrics and the number of tests, in order to obtain results with statistical significance. Some of these studies also omitted the complete description of the configurations of the adopted classifiers.

3.5. Clustering, Spatiotemporal Modeling

Prediction models involving clustering and spatiotemporal prediction presented relatively few studies when compared to the other approaches presented in this systematic review (Table 2). For studies with only these types of approaches, the year with the highest production was 2015, when 5 articles were published on the theme (Figure 2). An important point to highlight is the fact that, in the studies included in this group, the authors achieved the highest scores regarding the quality criterion involving the discussion about the limitations presented by the models (Figure 3).

Mathur et al. (114) brought a spatiotemporal prediction of dengue. This study also discussed and implemented dengue modeling with clustered incidence map visualization in Selangor, Malaysia. The spatiotemporal mapping was performed using the clustering technique with the k-mean algorithm. Thus were generated the incidence clusters. Then, the Gaussian mixture model was applied, finding the incidence density of dengue. Next, the K-means (K-NN algorithm) was used to find the centroid of the incidence. The Expectation-Maximization (EM) Algorithm was used to relate the clusters. The Bayesian Information Criteria (BIC) is then used to optimize the EM. Finally, with the Geographic Information System (GIS) technique, it is possible to accurately visualize the mapping of dengue incidence vulnerability in Selangor. The latter was used in the prediction. To create the proposed model, the R studio software was used, and to measure the vulnerability index, the K-means grouping was used. This study brings a spatiotemporal approach that can be used to implement health promotion policies. Another study that brought the spatiotemporal approach was the work of (115). In his study, Andersson et al. (115) made use of street images (Google Street) to implement a model to predict dengue hemorrhagic rates in the city of Rio de Janeiro, Brazil. In order to create this model, a siamese convolutional neural network technique was used. First, to create the models, dengue data in Rio de Janeiro were obtained and normalized. Next, street images were labeled according to latitude and longitude. The capacity of convolutional neural networks was analyzed with two approaches Simple-4CSCNN and ResNet-4CSCNN. The proposed models were implemented in the PyTorch framework. Simple-4CSCNN proved faster, with better loss of rating, but exhibited worse results in the validation test set. ResNet-4CSCNN generalized the training data well and reasonable results in the test set. The advantage of this approach is the use of street images to predict dengue cases, and the lack of work on the same line makes comparisons difficult.

The study of (116) was aimed at mapping the probability of an epidemic outbreak of Zika in the world. For this, three models were implemented, reverse propagation neural network (BPNN) (with sigmoid activation function), gradient increase machine (GBM), and random forest (RF). High-dimensional multidisciplinary covariate layers were combined

with comprehensive localization data on Zika virus infection in humans. In addition to the demographic distribution data of the *Aedes* mosquitoes, global climate data, socioeconomic data, night light data, and human movement data were used. To create the models, the R language (version 3.3.3) was used. Models were trained with cross-validation 10 times. To assess the performance of the prediction models, the ROC curve was used. The models created were robust and capable of simulating the global probability of transmission risk of ZIKV and also quantified the uncertainty of the accuracy of the prediction models. The models created provided reference information for model selection in the area of epidemiological cartography. However, the study only uses the AUC as a metric for evaluating the models.

In the study of (117), the authors developed a model for clustering and mapping dengue risk susceptibility. In his model, Ghosh et al. (117) used as variables epidemiological data, temperature (maximum and minimum), precipitation, relative humidity and Earth Surface Temperature (LST) images, demographic, socioeconomic, vegetation, and water index data. Two statistical methods were used to create the models: Poisson Models (to form the clusters) and Multiple Logistic Regression. This first was used to estimate the incidence of dengue. Moran location and weighting function I based on the specific spatial distance of the outlier were also used. This second function was used to estimate the probability of dengue occurrence using climatic variables as attributes. The researchers observed a strong association between monthly dengue cases and monthly mean rainfall and an association between monthly mean air humidity and disease cases. The model takes into account a spatiotemporal approach for predicting dengue risk. In addition, it considers the social and demographic aspects of predicting dengue.

In (118)'s approach, the authors created spatiotemporal prediction models for dengue cases taking into account population density. As variables, dengue cases in the city of Khyber Pakhtunkhwa, transmission vector records (*A. aegypti* and *A. albopictus*), population density and distance to roads and rivers were considered. As methods, logistic regression, variogram function, and binomial kriging with a binary logistic drift were used. Logistic regression was used to assess the correlation between dengue cases and other variables (covariants). Then the variogram function (spherical, Gaussian, circular, and Matérn) is calculated for the city under study and its subregions. Additionally, at the end, the estimation of the weights of the kriging equations is done using the weights of the variogram model. The researchers claim that the "presence" of the mosquito and population density affect the dynamics of the disease. And the models performed well in cities with high population density. However, the study did not make clear the databases used as well as the periods chosen for modeling, testing, and validation.

Phanitchat et al. (119) developed an identification of sub-district level dengue clusters in Thailand. For this purpose, data on weekly dengue cases (by gender), population density per Km^2 temperature, and rain in the same period for Khon Kaem province were used. The models used were Bayesian Poisson Regression and Local Indicators of Spatial Association

(LISA). The first was to assess the relationship between the number of monthly dengue cases in the 199 sub-districts. The metric for evaluating the fit of the model was the Wantabe-Akaike Information Criterion (WAIC). Finally, LISA was used to identify hot and cold spots and outliers in the incidence of dengue. The article concludes that dengue outbreaks are more frequent in the rainy season. With the analysis by hotspots, it is observed that there is a cluster of cases around the urban areas of Khon Kaem and in rural areas in the southwest of the region. The spatiotemporal approach is useful for application in health promotion strategies. However, there is an inherent limitation regarding the collection of public data, such as underreporting of cases, errors in reporting symptomatic cases, and absence of asymptomatic cases. In addition, the use of data is a little out of date.

Chen et al. (120) developed a new framework for producing spatiotemporal prediction at the neighborhood level. Various data were used, such as dengue incidence data (with home address data and start date), movement patterns, construction age of buildings in a neighborhood, meteorological data (maximum and minimum temperature and average relative humidity), number of national weekly cases, index by Normalized Difference (NDVI) among others. The separate prediction models and submodels created were based on LASSO for each prediction window. Climatic variables and their effects have a greater effect when analyzing longer time intervals. The fact of having less vegetation, older buildings, greater connectivity to other areas, and more travelers arriving in the area causes the number of cases to increase. The proposed model brings a spatiotemporal approach at the neighborhood level up to 3 months in advance. The system proved to be robust to changes in baseline incidence over time.

Jat and Mala (121), in turn, brought an approach to the use of digital geospatial technologies to identify potential sources of dengue incidence. For this purpose, the spatiotemporal grouping of dengue incidences was performed using the Kulldorff scanning method. With the help of Getis-OrdGi statistics, high-risk areas were identified and then implemented in the GIS. And the data obtained was correlated with meteorological parameters, such as wind speed, humidity and demographic factors, such as age and gender. This work shows that the occurrence of dengue is not random, it is directly linked to meteorological phenomena. Thus, this study serves as a warning and to use actions to group regions that may be focuses on dengue spread.

The studies cited here brought interesting approaches to dengue and Zika, considering both epidemiological and climatic data (such as precipitation and temperature), as well as population density, age of construction of buildings, socioeconomic and demographic data, and cases of the disease by patient gender. One of the works is the first, as far as is known, to use street view images to predict dengue cases, something quite innovative. The models created had good results regarding their evaluation metrics. Both machine learning models and statistical models were used. These surveys also rely on algorithms that do not have a great computational weight, which makes their use by the public service viable. The models made were both at the sub-district level and the global level. The spatio-temporal approach

brought by the studies in this section helps health managers in directing public resources to areas that need more attention.

3.6. Other Approaches

The articles included in group 6 are articles that combined more than one approach in their predictions, or that had a very different approach from the rest of the articles evaluated in this review. According to **Figure 2**, three of the studies were published in the year 2019. In the years 2016, 2017, and 2020, 1, 1, and 2 studies were published, respectively (**Figure 2**). Analyzing the scores referring to the quality criteria, the average scores in most QC were above 0.7, except QC5 and QC6 (**Figure 3**).

Among the seven articles, three of them simultaneously addressed numerical prediction models of arboviruses cases and also prediction of the risk of epidemiological outbreaks (122–124). In the study of (125), the authors addressed risk production models as well as clustering models to identify regions with similar patterns of disease transmission. Harumy et al. (126), on the other hand, the authors investigated prediction models of the area as the greatest potential to suspend arboviruses and case prediction. Yamamoto et al. (127), in turn, brought an approach to detecting the importation of arboviruses into a country. As for arboviruses groups, the publications were mostly concentrated on dengue (122, 124, 125). Only Yamamoto et al. (127) brought a study considering Zika virus disease cases.

It is important to highlight the variety of models that were covered. Among them, we can mention Random Forest, RF-USA, Logistic Regression (124), and Naïve Bayes (125) for the classification steps. Both (123) and (122) considered a threshold value for identifying an epidemic or outbreak. In the steps involving regression, probabilistic models (127), LASSO, ARIMA, SARIMA (124), Generalized Linear Regression (123), Artificial Neural Network (122, 126), SEIR model (122), and multiple variate regression (125) models were used.

For studies that presented a mixed approach, models for the numerical prediction of cases are essential for the analysis of the epidemiological curve of the disease. In this way, health authorities may have indications that combat policies are or are not effective in combating arboviruses. On the other hand, predictions with spatial approaches can indicate regions with more or less intensity of cases, which can help the distribution of financial and human resources to the most critical regions. Therefore, a mixed approach to the prediction of arboviruses is shown to be robust to assist in decision-making on arbovirus prevention policies.

4. CONCLUSION

Arboviruses have a major impact on populations affected by seasonal outbreaks of these diseases. In addition to the impact caused by the number of deaths and infections, the socioeconomic impacts tend to remain until the next outbreak. The prevention and control of the occurrence of these diseases are directly associated with the monitoring/control of their transmitting vector.

In this sense, this systematic review aimed to identify predictive models of diseases transmitted by *A. aegypti*, as well as

identify existing models for modeling vector dynamics. For this, we defined a review protocol that was followed throughout the process. We obtained 429 publications retrieved from scientific databases using a predefined search string. After filtering through the exclusion and inclusion criteria, 139 studies remained in the review for analysis and evaluation of quality criteria.

The remaining studies after the entire analysis process were grouped according to their similar characteristics. Arboviruses' prediction studies are mostly linked to the numerical prediction of cases. According to the results obtained, we observed that among the arboviruses transmitted by *A. aegypti*, most of the studies are aimed at predicting dengue. Both in numerical prediction models, as a prediction of outbreaks, epidemics, and disease diagnosis. Studies regarding predictions with a spatiotemporal approach are also more focused on dengue rather than on Zika and chikungunya. An important point to highlight is the fact that few studies were focused on the spatiotemporal prediction of diseases, as well as the prediction of models related to mosquito dynamics.

Another point that can be highlighted in the studies in this review is in relation to the variables selected for the generation of arbovirus models. In the case of modeling taking into account numerical prediction, prediction of outbreaks and epidemics, and spatiotemporal prediction, we observed that most studies consider climatic variables as model parameters. Among them, the most common are the historical series temperature, rain and relative humidity. However, parameters related to natural phenomena and also variables obtained by remote sensing also gained prominence, as well as data from social networks and search queries. Furthermore, data related to vector monitoring have also been included both in numerical prediction models of arboviruses and in models related to the dynamics of the *A. aegypti* itself. On the other hand, in arbovirus models that prioritize the detection of infection in the individual, we note that the most used parameters are symptomatological parameters. The use of models based only on symptomatological parameters can cause fever-like diseases to be confused with dengue, Zika, and chikungunya. However, we also found studies that use hematological parameters to detect infection.

In this study, we analyzed that, for prediction problems involving arboviruses and also involving mosquito dynamics, a large part of the data is obtained through local health and climatology agencies. Missing data and cases of underreporting by health agencies are one of the most reported problems in the studies evaluated.

Furthermore, this systematic review also demonstrated that there is a range of models that are widely used in prediction problems. Poisson models and moving average models (ARIMA, SARIMA) are widely used to predict historical series. However, we observe that artificial neural networks, support vector

machines, and decision tree-based models are widely explored by the studies in this review. It is important to highlight that in many of the works that use Artificial Intelligence models, the authors often do not describe the configurations of the evaluated models and how the models were validated. In other words, although the models have good evaluation metrics, there is no way to guarantee their statistical relevance.

Finally, the arboviruses dynamics is a very heterogeneous problem that involves the interaction of various factors such as climatic and environmental factors, mosquitoes, and human beings. The heterogeneity of arbovirus dynamics is precisely what makes the prediction problem a very complex problem. Therefore, from this systematic review, we hope to provide a theoretical foundation regarding the state-of-the-art of dengue, Zika, and chikungunya prediction models, as well as the breeding sites of its main urban transmitter vector. Hence, we believe that there is great potential for exploring models with a spatiotemporal approach. These models can be an important tool in the fight against arbovirus-borne diseases, as they contain spatial information of epidemiological interest that will be able to more effectively direct human and financial resources, especially in more vulnerable countries.

DATA AVAILABILITY STATEMENT

The data and materials for all experiments reviewed in this study are publicly available as open datasets and are cited in the body of the text.

AUTHOR CONTRIBUTIONS

CL, AnS, GM, CC, AbS, and WS designed the research protocol. CL, AnS, GM, and CC wrote the document review. AM, AA, LD, IB, MT, EB, and SB supported the research. AbS, TM, TA, LC, OY, PK, KJ, and WS supervised and supported all the work. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Brazilian research agencies FACEPE, CAPES, CNPq, and the University College London held UKRI research grant number NE/T013664/1.

ACKNOWLEDGMENTS

The authors are grateful to the Brazilian research agencies FACEPE, CAPES, CNPq, and to UKRI, research grant number NE/T013664/1, for the partial financial support of this research.

REFERENCES

- Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. (2013) 496:504. doi: 10.1038/nature12060
- WHO. *Ending the Neglect to Attain the Sustainable Development Goals: A Road Map for Neglected Tropical Diseases 2021-2030*. (2021). Available online at: https://www.who.int/neglected_diseases/resources/who-ucn-ntd-2020.01/en/ (accessed April 6, 2021).

3. de Lima TFM, Lana RM, de Senna Carneiro TG, Codeço CT, Machado GS, Ferreira LS, et al. Dengueme: a tool for the modeling and simulation of dengue spatiotemporal dynamics. *Int J Environ Res Public Health*. (2016) 13:920. doi: 10.3390/ijerph13090920
4. Figueiredo R, Paiva C, Morato M. *Arboviroses*. Rio de Janeiro: Fundacao Oswaldo Cruz (2017). Available online at: <http://www.canal.fiocruz.br/video/index.php?v=arboviroses1les1-1924>
5. Musso D, Stramer SL, Busch MP. Zika virus: a new challenge for blood transfusion. *Lancet*. (2016) 387:1993–4. doi: 10.1016/S0140-6736(16)30428-7
6. Musso D, Gubler DJ. Zika virus. *Clin Microbiol Rev*. (2016) 29:487–524. doi: 10.1128/CMR.00072-15
7. PAHO. *Vector-Borne Diseases*. (2019). Available online at: https://www.paho.org/bra/index.php?option=com_content&view=article&id=5796:doencas-transmissivveis-analise-de-situacao-de-saude&Itemid=0 (accessed April 6, 2021).
8. WHO. *Chikungunya*. (2020).
9. WHO. *Zika virus*. (2018).
10. Wilder-Smith A, Gluber DJ, Weaver SC, Monath TP, Heymann DL, Scott TW. Epidemic arboviral diseases: priorities for research and public health. *Lancet Infect Dis*. (2016) 17:E101–6. doi: 10.1016/S1473-3099(16)30518-7
11. Gubler DJ. Dengue, urbanization and globalization: the unholy trinity of the 21st century. *Trop Med Health*. (2011) 39:S3–11. doi: 10.2149/tmh.2011-S05
12. Mohammed A, Chadee DD. Effects of different temperature regimens on the development of *Aedes aegypti* (L.) (Diptera: Culicidae) mosquitoes. *Acta Trop*. (2011) 119:38–43. doi: 10.1016/j.actatropica.2011.04.004
13. Valotto CFB, Carvasin G, Silva HHG, Geris R, Silva IGd. Alterações morfo-histológicas em larvas de *Aedes aegypti* (Linnaeus, 1762) (Diptera, Culicidae) causadas pelo tanino catéuico isolado da planta do cerrado *Magonia pubescens* (Sapindaceae). *Rev Patol Trop*. (2010) 39:309–21. Available online at: <http://repositorio.bc.ufg.br/handle/ri/182>
14. Aburas HM, Cetiner BG, Sari M. Dengue confirmed-cases prediction: a neural network model. *Expert Syst Appl*. (2010) 37:4256–60. doi: 10.1016/j.eswa.2009.11.077
15. Ibrahim F, Taib MN, Abas WABW, Guan CC, Sulaiman S. A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN). *Comput Methods Prog Biomed*. (2005) 79:273–81. doi: 10.1016/j.cmpb.2005.04.002
16. Yusof Y, Mustafa Z. Dengue outbreak prediction: a least squares support vector machines approach. *Int J Comput Theory Eng*. (2011) 3:489. doi: 10.7763/IJCTE.2011.V3.355
17. Chakraborty T, Chattopadhyay S, Ghosh I. Forecasting dengue epidemics using a hybrid methodology. *Phys A Stat Mech Appl*. (2019) 527:121266. doi: 10.1016/j.physa.2019.121266
18. Stolerman LM, Maia PD, Kutz JN. Forecasting dengue fever in Brazil: an assessment of climate conditions. *PLoS ONE*. (2019) 14:e0220106. doi: 10.1371/journal.pone.0220106
19. da Cruz Ferreira DA, Degener CM, de Almeida Marques-Toledo C, Bendati MM, Fetzer LO, Teixeira CP, et al. Meteorological variables and mosquito monitoring are good predictors for infestation trends of *Aedes aegypti*, the vector of dengue, chikungunya and Zika. *Parasites Vect*. (2017) 10:78. doi: 10.1186/s13071-017-2025-8
20. Padmanabhan P, Seshaiyer P, Castillo-Chavez C. Mathematical modeling, analysis and simulation of the spread of Zika with influence of sexual transmission and preventive measures. *Lett Biomath*. (2017) 4:148–66. doi: 10.30707/LiB4.1Padmanabhan
21. Jindal A, Rao S. Agent-based modeling and simulation of mosquito-borne disease transmission. In: *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. São Paulo: International Foundation for Autonomous Agents and Multiagent Systems (2017). p. 426–35.
22. Saptarini NGAPH, Dillak RY, Pakan PD. Dengue haemorrhagic fever outbreak prediction using Elman Levenberg neural network and genetic algorithm. In: *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*. Makassar (2018). p. 188–91. doi: 10.1109/EIConCIT.2018.8878529
23. Polwiang S. The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003–2017). *BMC Infect Dis*. (2020) 20:208. doi: 10.1186/s12879-020-4902-6
24. Siregar FA, Makmur T. Climate risk and environmental determinants on dengue transmission. *Indian J Public Health Res Dev*. (2019) 10:e0009761. doi: 10.5958/0976-5506.2019.00241.9
25. Sukama Y, Hertono GF, Handari BD, Aldila D. Comparing activation functions in predicting dengue hemorrhagic fever cases in DKI Jakarta using recurrent neural networks. In: *AIP Conference Proceedings*. Surakarta: AIP Publishing LLC (2020). p. 020059. doi: 10.1063/5.0030456
26. Siregar F, Makmur T. Time series analysis of dengue hemorrhagic fever cases and climate: a model for dengue prediction. In: *Journal of Physics: Conference Series*. Medan: IOP Publishing (2019). p. 012072. doi: 10.1088/1742-6596/1235/1/012072
27. Puengpreeda A, Yhusumrarn S, Sirikulvadhana S. Weekly forecasting model for dengue hemorrhagic fever outbreak in Thailand. *Eng J*. (2020) 24:71–87. doi: 10.4186/ej.2020.24.3.71
28. Soliman M, Lyubchich V, Gel YR. Ensemble forecasting of the Zika space-time spread with topological data analysis. *Environmetrics*. (2020) 31:e2629. doi: 10.1002/env.2629
29. Teng Y, Bi D, Xie G, Jin Y, Huang Y, Lin B, et al. Dynamic forecasting of Zika epidemics using Google Trends. *PLoS ONE*. (2017) 12:e0165085. doi: 10.1371/journal.pone.0165085
30. Morsy S, Dang TN, Kamel MG, Zayan AH, Makram OM, Elhady M, et al. Prediction of Zika-confirmed cases in Brazil and Colombia using Google Trends. *Epidemiol Infect*. (2018) 146:1625–7. doi: 10.1017/S0950268818002078
31. Ogden NH, Fazil A, Safronetz D, Drebot MA, Wallace J, Rees EE, et al. Risk of travel-related cases of Zika virus infection is predicted by transmission intensity in outbreak-affected countries. *Parasites Vect*. (2017) 10:1–9. doi: 10.1186/s13071-017-1977-z
32. Dhaka A, Singh P. Comparative analysis of epidemic alert system using machine learning for dengue and chikungunya. In: *2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence)*. Noida (2020). p. 798–804. doi: 10.1109/Confluence47617.2020.9058048
33. Verma S, Sharma N. Statistical models for predicting chikungunya incidences in India. In: *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*. Jalandhar (2018). p. 139–42. doi: 10.1109/ICSCCC.2018.8703218
34. Kuruge DA, Granmo OC, Goodwin M. A novel tsetlin automata scheme to forecast dengue outbreaks in the Philippines. In: *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*. Volos (2018). p. 680–5. doi: 10.1109/ICTAI.2018.00108
35. Lopez-Montenegro LE, Pulecio-Montoya AM, Marcillo-Hernandez GA. Dengue cases in Colombia: mathematical forecasts for 2018–2022. *MEDICC Rev*. (2019). 21:38–45. doi: 10.37757/MR2019.V21.N2-3.8
36. Espina K, Estuar MRJE. Infodemiology for syndromic surveillance of dengue and typhoid fever in the Philippines. *Proc Comput Sci*. (2017) 121:554–61. doi: 10.1016/j.procs.2017.11.073
37. Damayanti A, Hidayati N, Pratiwi A. Model identification for prediction of dengue fever disease spreading using Bat Algorithm and backpropagation. In: *Journal of Physics: Conference Series*. Purwokerto: IOP Publishing (2020). p. 012002. doi: 10.1088/1742-6596/1494/1/012002
38. Hasanah H, Hertono GF, Sarwinda D. Prediction of dengue incidence in DKI Jakarta using adaptive neuro-fuzzy inference system. *AIP Conf Proc*. (2020) 2296:020024. doi: 10.1063/5.0030455
39. Roziqin MC, Basuki A, Harsono T. A comparison of Montecarlo linear and dynamic polynomial regression in predicting dengue fever case. In: *2016 International Conference on Knowledge Creation and Intelligent Computing (KCIC)*. Manado (2016). p. 213–8. doi: 10.1109/KCIC.2016.7883649
40. Halim S, Octavia T, Felecia, Handoyo A. Dengue fever outbreak prediction in Surabaya using a geographically weighted regression. In: *2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*. Bangkok (2019). p. 1–5. doi: 10.1109/TIMES-iCON47539.2019.9024438
41. Nakvisut A, Phienthrakul T. Two-step prediction technique for dengue outbreak in Thailand. In: *2018 International Electrical Engineering Congress (IEECON)*. Krabi (2018). p. 1–4. doi: 10.1109/IEECON.2018.8712258
42. Mishra VK, Tiwari N, Ajaymon SL. Dengue disease spread prediction using twofold linear regression. In: *2019 IEEE 9th International*

- Conference on Advanced Computing (IACC)*. (2019). p. 182–7. doi: 10.1109/IACC48062.2019.8971567
43. Pineda-Cortel MRB, Clemente BM, Nga PTT. Modeling and predicting dengue fever cases in key regions of the Philippines using remote sensing data. *Asian Pac J Trop Med*. (2019) 12:60–6. doi: 10.4103/1995-7645.250838
 44. Kerdprasop N, Kerdprasop K, Chuaybamroong P. Computational intelligence and statistical learning performances on predicting dengue incidence using remote sensing data. *Adv Sci Technol Eng Syst J*. (2020) 5:344–50. doi: 10.25046/aj050440
 45. Dharmawardana KGS, Lokuge JN, Dassanayake PSB, Sirisena ML, Fernando ML, Perera AS, et al. Predictive model for the dengue incidences in Sri Lanka using mobile network big data. In: *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*. Peradeniya (2017). p. 1–6. doi: 10.1109/ICIINFS.2017.8300381
 46. Zhao N, Charland K, Carabali M, Nsoesie EO, Maheu-Giroux M, Rees E, et al. Machine learning and dengue forecasting: Comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia. *PLoS Neglect Trop Dis*. (2020) 14:e0008056. doi: 10.1371/journal.pntd.0008056
 47. Abbas S, Ilyas M. Assessing the impact of El Niño southern oscillation index and land surface temperature fluctuations on dengue fever outbreaks using ARIMAX (p)-PARX (p)-NBARX (p) models. *Arab J Geosci*. (2018) 11:1–12. doi: 10.1007/s12517-018-4119-9
 48. Liao CM, Huang TL, You SH, Cheng YH, Hsieh NH, Chen WY. Regional response of dengue fever epidemics to interannual variation and related climate variability. *Stochast Environ Res Risk Assess*. (2015) 29:947–58. doi: 10.1007/s00477-014-0948-6
 49. Chang FS, Tseng YT, Hsu PS, Chen CD, Lian IB, Chao DY. Re-assess vector indices threshold as an early warning tool for predicting dengue epidemic in a dengue non-endemic country. *PLoS Neglect Trop Dis*. (2015) 9:e0004043. doi: 10.1371/journal.pntd.0004043
 50. Shi Y, Kok SY, Rajarethinam J, Liang S, Yap G, Chong CS, et al. Three-month real-time dengue forecast model: an early warning system for outbreaks alerts and policy decision in Singapore. *Environ Health Perspect*. (2016) 124:1369–75. doi: 10.1289/ehp.1509981
 51. Liu KK, Wang T, Huang XD, Wang GL, Xia Y, Zhang YT, et al. Risk assessment of dengue fever in Zhongshan, China: a time-series regression tree analysis. *Epidemiol Infect*. (2017). 145:451–61. doi: 10.1017/S095026881600265X
 52. Jing QL, Cheng Q, Marshall JM, Hu WB, Yang ZC, Lu JH. Imported cases and minimum temperature drive dengue transmission in Guangzhou, China: evidence from ARIMAX model. *Epidemiol Infect*. (2018) 146:1226–35. doi: 10.1017/S0950268818001176
 53. Siriysatien P, Phumee A, Ongruk P, Jampachaisri K, Kesorn K. Analysis of significant factors for dengue fever incidence prediction. *BMC Bioinform*. (2016) 17:166. doi: 10.1186/s12859-016-1034-5
 54. Guo P, Zhang Q, Chen Y, Xiao J, He J, Zhang Y, et al. An ensemble forecast model of dengue in Guangzhou, China using climate and social media surveillance data. *Sci Tot Environ*. (2019) 647:752–62. doi: 10.1016/j.scitotenv.2018.08.044
 55. Jayaraj VJ, Avoi R, Navindran G, Dhese BR, Yusri U. Developing a dengue prediction model based on climate in Tawau, Malaysia. *Acta Trop*. (2019) 197:105055. doi: 10.1016/j.actatropica.2019.105055
 56. 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
 57. Mustafa Z, Sulaiman MH, Mohsin MFM, Yusof Y, Ernawan F, Rosli KAM. An application of hybrid swarm intelligence algorithms for dengue outbreak prediction. In: *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEET)*. Amman (2019). p. 731–5. doi: 10.1109/JEET.2019.8717436
 58. Mustafa Z, Sulaiman MH, Emawan F, Yusof Y, Mohsin MFM. Dengue outbreak prediction: hybrid meta-heuristic model. In: *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. Busan (2018). p. 271–4. doi: 10.1109/SNPD.2018.8441095
 59. Doni AR, Sasipraba T. LSTM-RNN based approach for prediction of dengue cases in India. *Ingénierie des Systèmes d'Information*. (2020) 25:327–35. doi: 10.18280/isi.250306
 60. Husin NA, Mustapha N, Sulaiman N, Yaacob R, Hamdan H, Hussin M. Performance of hybrid GANN in comparison with outbreaks standalone models on dengue outbreak prediction. *J Comput Sci*. (2016) 12:300–6. doi: 10.3844/jcssp.2016.300.306
 61. Shshvat K, Basu R, Bhondekar P, Kaur A. An ensemble model for forecasting infectious diseases in India. *Trop Biomed*. (2019) 36:822–32.
 62. Rahman KM, Sharker Y, Rumi RA, Khan MUI, Shomik MS, Rahman MW, et al. An Association between rainy days with clinical dengue fever in Dhaka, Bangladesh: findings from a hospital based study. *Int J Environ Res Public Health*. (2020) 17:9506. doi: 10.3390/ijerph17249506
 63. Chumpu R, Khamsemanan N, Nattee C. The association between dengue incidences and provincial-level weather variables in Thailand from 2001 to 2014. *PLoS ONE*. (2019) 14:e0226945. doi: 10.1371/journal.pone.0226945
 64. Phung D, Huang C, Rutherford S, Chu C, Wang X, Nguyen M, et al. Identification of the prediction model for dengue incidence in Can Tho city, a Mekong Delta area in Vietnam. *Acta Trop*. (2015) 141:88–96. doi: 10.1016/j.actatropica.2014.10.005
 65. Baquero OS, Santana LMR, Chiaravalloti-Neto F. Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. *PLoS ONE*. (2018) 13:e0195065. doi: 10.1371/journal.pone.0195065
 66. Eljorde FI, Clarite DS, Gerardo BD, Byun Y. Tracking and prediction of dengue outbreak using cloud-based services and artificial neural network. *Int J Multimedia Ubiquit Eng*. (2016) 11:355–66. doi: 10.14257/ijmue.2016.11.5.33
 67. Thiruchelvam L, Asirvadam VS, Dass SC, Daud H, Gill BS. K-step ahead prediction models for dengue occurrences. In: *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. Kuching (2017). p. 541–6. doi: 10.1109/ICSIPA.2017.8120671
 68. Pham DN, Aziz T, Kohan A, Nellis S, Jamil JbA, Khoo JJ, et al. How to efficiently predict dengue incidence in Kuala Lumpur. In: *2018 Fourth International Conference on Advances in Computing, Communication Automation (ICACCA)*. Subang Jaya (2018). p. 1–6. doi: 10.1109/ICACCAF.2018.8776790
 69. Mussumeci E, Codeso Coelho F. Large-scale multivariate forecasting models for Dengue - LSTM versus random forest regression. *Spatial Spatio Temp Epidemiol*. (2020) 35:100372. doi: 10.1016/j.sste.2020.100372
 70. Xu J, Xu K, Li Z, Meng F, Tu T, Xu L, et al. Forecast of dengue cases in 20 Chinese cities based on the deep learning method. *Int J Environ Res Public Health*. (2020) 17:453. doi: 10.3390/ijerph17020453
 71. Anggraeni W, Pramudita G, Riksakomara E, P W R, Samopa F, Pujiadi, et al. Artificial neural network for health data forecasting, case study: number of dengue hemorrhagic fever cases in Malang Regency, Indonesia. In: *2018 International Conference on Electrical Engineering and Computer Science (ICECOS)*. South Kuta (2018). p. 207–12. doi: 10.1109/ICECOS.2018.8605254
 72. Datoc HI, Caparas R, Caro J. Forecasting and data visualization of dengue spread in the Philippine Visayas island group. In: *2016 7th International Conference on Information, Intelligence, Systems Applications (IISA)*. Chalkidiki (2016). p. 1–4. doi: 10.1109/IISA.2016.7785420
 73. Jarrin EP, Cordeiro FB, Medranda WC, Barrett M, Zambrano M, Regato M. A Machine Learning-Based algorithm for the assessment of clinical metabolomic fingerprints in Zika virus disease. In: *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. Guayaquil: IEEE (2019). p. 1–6. doi: 10.1109/LA-CCI47412.2019.9037029
 74. Melo CFOR, Navarro LC, De Oliveira DN, Guerreiro TM, Lima EdO, Delafiori J, et al. A machine learning application based in random forest for integrating mass spectrometry-based metabolomic data: a simple screening method for patients with Zika virus. *Front Bioeng Biotechnol*. (2018) 6:31. doi: 10.3389/fbioe.2018.00031
 75. Mahalakshmi B, Suseendran G. Prediction of Zika virus by multilayer perceptron neural network (MLPNN) using cloud. *Int J Recent Technol Eng*. (2019) 8:1–6. doi: 10.35940/ijrte.B1041.098251119
 76. Mello-Román JD, Mello-Román JC, Gomez-Guerrero S, Garcia-Torres M. Predictive models for the medical diagnosis of dengue: a case

- study in Paraguay. *Comput Math Methods Med.* (2019) 2019:7307803. doi: 10.1155/2019/7307803
77. Sarma D, Hossain S, Mittra T, Bhuiya MAM, Saha I, Chakma R. Dengue prediction using machine learning algorithms. In: *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*. Kuching: IEEE (2020). p. 1–6. doi: 10.1109/R10-HTC49770.2020.9357035
 78. Ho TS, Weng TC, Wang JD, Han HC, Cheng HC, Yang CC, et al. Comparing machine learning with case-control models to identify confirmed dengue cases. *PLoS Neglect Trop Dis.* (2020) 14:e0008843. doi: 10.1371/journal.pntd.0008843
 79. Alam M, Sethi S, Shakil KA. Distributed machine learning based biocloud prototype. *Int J Appl Eng Res.* (2015) 10:37578–83.
 80. Ganthimathi M, Thangamani M, Mallika C, Balaji VP. Prediction of dengue fever using intelligent classifier. *Int J Emerg Trends Eng Res.* (2020) 8:1338–41. doi: 10.30534/ijeter/2020/65842020
 81. Kapoor R, Kadyan V, Ahuja S. Weight based-artificial neural network (W-ANN) for predicting dengue using machine learning approach with Indian perspective. *Int J Sci Technol Res.* (2020). 9:3290–8.
 82. Ariffin AM, Aris NM. Data-driven neural network model for early self-diagnosis of dengue symptoms. *J Theoret Appl Inform Technol.* (2020) 98:4228–38.
 83. Dharap P, Raimbault S. Performance evaluation of machine learning-based infectious screening flags on the HORIBA Medical Yumizen H550 Haematology Analyzer for vivax malaria and dengue fever. *Malaria J.* (2020) 19:1–10. doi: 10.1186/s12936-020-03502-3
 84. Srivastava S, Soman S, Rai A, Cheema AS. An online learning approach for dengue fever classification. In: *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. Rochester, NY: IEEE (2020). p. 163–8. doi: 10.1109/CBMS49503.2020.00038
 85. Sasongko PS, Wibawa HA, Maulana F, Bahtiar N. Comparacao de desempenho de modelos de redes neurais artificiais para deteccao de dengue. In: *2017 Conferencia Internacional sobre Informatica e Ciencias Computacionais (ICICoS)*. Semarang: IEEE (2017). p. 183–8.
 86. Iqbal N, Islam M. Machine learning for dengue outbreak prediction: a performance evaluation of different prominent classifiers. *Informatica.* (2019) 43:363–71. doi: 10.31449/inf.v43i3.1548
 87. Balamurugan SAA, Mallick MM, Chinthana G. Improved prediction of dengue outbreak using combinatorial feature selector and classifier based on entropy weighted score based optimal ranking. *Inform Med Unlocked.* (2020) 20:100400. doi: 10.1016/j.imu.2020.100400
 88. Abeyrathna KD, Granmo OC, Zhang X, Goodwin M. Adaptive continuous feature binarization for tsetlin machines applied to forecasting dengue incidences in the Philippines. In: *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. Canberra, ACT (2020). p. 2084–92. doi: 10.1109/SSCI47803.2020.9308291
 89. Jongmuenwai B, Lowanichchai S, Jabjone S. Comparison using data mining algorithm techniques for predicting of dengue fever data in northeastern of Thailand. In: *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. Chiang Rai (2018). p. 532–5. doi: 10.1109/ECTICon.2018.8619953
 90. Najar AM, Irawan MI, Adzkiya D. Extreme learning machine method for dengue hemorrhagic fever outbreak risk level prediction. In: *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*. Kuala Lumpur (2018). p. 1–5. doi: 10.1109/ICSCEE.2018.8538409
 91. Zhu G, Hunter J, Jiang Y. Improved prediction of dengue outbreak using the delay permutation entropy. In: *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. Chengdu (2016). p. 828–32. doi: 10.1109/iThings-GreenCom-CPSCom-SmartData.2016.172
 92. Anggraeni W, Sumpeno S, Yuniarno EM, Rachmadi RF, Gumelar AB, Purnomo MH. Prediction of dengue fever outbreak based on climate factors using fuzzy-logistic regression. In: *2020 International Seminar on Intelligent Technology and Its Applications (ISITIA)*. (2020). p. 199–204. doi: 10.1109/ISITIA49792.2020.9163708
 93. Rahmawati D, Huang YP. Using C-support vector classification to forecast dengue fever epidemics in Taiwan. In: *2016 International Conference on System Science and Engineering (ICSSE)*. (2016). p. 1–4. doi: 10.1109/ICSSE.2016.7551552
 94. Chan TC, Hu TH, Hwang JS. Daily forecast of dengue fever incidents for urban villages in a city. *Int J Health Geograph.* (2015) 14:1–11. doi: 10.1186/1476-072X-14-9
 95. Brett TS, Rohani P. Dynamical footprints enable detection of disease emergence. *PLoS Biol.* (2020) 18:e3000697. doi: 10.1371/journal.pbio.3000697
 96. Zhang Y, Wang T, Liu K, Xia Y, Lu Y, Jing Q, et al. Developing a time series predictive model for dengue in Zhongshan, China based on weather and Guangzhou dengue surveillance data. *PLoS Neglect Trop Dis.* (2016) 10:e0004473. doi: 10.1371/journal.pntd.0004473
 97. Adde A, Roucou P, Mangeas M, Ardillon V, Desenclos JC, Rousset D, et al. Predicting dengue fever outbreaks in french guiana using climate indicators. *PLoS Neglect Trop Dis.* (2016) 10:e0004681. doi: 10.1371/journal.pntd.0004681
 98. Bowman LR, Tejada GS, Coelho GE, Sulaiman LH, Gill BS, McCall PJ, et al. Alarm variables for dengue outbreaks: a multi-centre study in Asia and Latin America. *PLoS ONE.* (2016) 11:e0157971. doi: 10.1371/journal.pone.0157971
 99. Zainudin Z, Shamsuddin SM. Predictive analytics in Malaysian Dengue data from 2010 until 2015 using BigML. *Int J Advance Soft Compu Appl.* (2016) 8:18–30.
 100. Teng Y, Bi D, Xie G, Jin Y, Huang Y, An X, et al. Model-informed risk assessment for Zika virus outbreaks in the Asian-Pacific regions. *J Infect.* (2017) 74:484–91. doi: 10.1016/j.jinf.2017.01.015
 101. Akhtar M, Kraemer MU, Gardner LM. A dynamic neural network model for predicting risk of Zika in real time. *BMC Med.* (2019) 17:171. doi: 10.1186/s12916-019-1389-3
 102. Raizada S, Mala S, Shankar A. Vector borne disease outbreak prediction by machine learning. In: *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*. Bengaluru (2020). p. 213–8. doi: 10.1109/ICSTCEE49637.2020.9277286
 103. Hassani H, Yeganegi MR, Silva ES, Ghodsi F. Risk management, signal processing and econometrics: a new tool for forecasting the risk of disease outbreaks. *J Theoret Biol.* (2019) 467:57–62. doi: 10.1016/j.jtbi.2019.01.032
 104. Anno S, Hara T, Kai H, Lee MA, Chang Y, Oyoshi K, et al. Spatiotemporal dengue fever hotspots associated with climatic factors in Taiwan including outbreak predictions based on machine-learning. *Geospatial Health.* (2019) 14. doi: 10.4081/gh.2019.771
 105. Haddawy P, Wettayakorn P, Nonthaleerak B, Su Yin M, Wiratsudakul A, Schöning J, et al. Large scale detailed mapping of dengue vector breeding sites using street view images. *PLoS Neglect Trop Dis.* (2019) 13:e0007555. doi: 10.1371/journal.pntd.0007555
 106. Raja DB, Mallol R, Ting CY, Kamaludin F, Ahmad R, Ismail S, et al. Artificial intelligence model as predictor for dengue outbreaks. *Malaysian J Public Health Med.* (2019) 19:103–8. doi: 10.37268/mjphm/vol.19/no.2/art.176
 107. Asmai SA, Abidin ZZ, Nizam AFNAR, Mohd Ali MH. Aedes mosquito larvae recognition with a mobile app. *Int J Adv Trends Comput Sci Eng.* (2020) 9:5059–65. doi: 10.30534/ijatcse/2020/126942020
 108. Lee KY, Chung N, Hwang S. Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. *Ecol Inform.* (2016) 36:172–80. doi: 10.1016/j.ecoinf.2015.08.011
 109. Rubio-Solis A, Musah A, P Dos Santos W, Massoni T, Birjovanu G, Kostkova P. Zika virus: prediction of Aedes mosquito larvae occurrence in Recife (Brazil) using online extreme learning machine and neural networks. In: *Proceedings of the 9th International Conference on Digital Public Health.* (2019). p. 101–10. doi: 10.1145/3357729.3357738
 110. Bennett KL, Martinez CG, Almanza A, Rovira JR, McMillan WO, Enriquez V, et al. High infestation of invasive Aedes mosquitoes in used tires along the local transport network of Panama. *Parasites Vect.* (2019) 12:1–10. doi: 10.1186/s13071-019-3522-8
 111. Estallo EL, Benitez EM, Lanfri MA, Scavuzzo CM, Almiron WR. dados ambientais MODIS para avaliar doencas de Chikungunya, Dengue e Zika por meio da estimativa da atividade de oviposicao de Aedes (Stegomyia) aegypti. *IEEE J Select Top Appl Earth Observ Remote Sens.* (2016) 9:5461–6. doi: 10.1109/JSTARS.2016.2604577
 112. Estallo EL, Luduena-Almeida FF, Introini MV, Zaidenberg M, Almiron WR. Weather variability associated with Aedes (Stegomyia)

- aegypti* (Dengue vector) oviposition dynamics in Northwestern Argentina. *PLoS ONE*. (2015) 10:e0127820. doi: 10.1371/journal.pone.0127820
113. Hettiarachchige C, von Cavallar S, Lynar T, Hickson RI, Gambhir M. Risk prediction system for dengue transmission based on high resolution weather data. *PLoS ONE*. (2018) 13:e0208203. doi: 10.1371/journal.pone.0208203
 114. Mathur N, Asirvadam VS, Dass SC. Spatial-temporal visualization of dengue incidences using gaussian kernel. In: *2018 International Conference on Intelligent and Advanced System (ICIAS)*. Kuala Lumpur: IEEE (2018). p. 1–6. doi: 10.1109/ICIAS.2018.8540593
 115. Andersson VO, Ferreira Birck MA, Araujo RM. Towards predicting dengue fever rates using convolutional neural networks and street-level images. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. Rio de Janeiro (2018). p. 1–8. doi: 10.1109/IJCNN.2018.8489567
 116. Jiang D, Hao M, Ding F, Fu J, Li M. Mapping the transmission risk of Zika virus using machine learning models. *Acta Trop*. (2018) 185:391–9. doi: 10.1016/j.actatropica.2018.06.021
 117. Ghosh S, Dinda S, Chatterjee ND, Das K, Riya M. The spatial clustering of dengue disease and risk susceptibility mapping: an approach towards sustainable health management in Kharagpur, India. *Spatial Inform Res*. (2019) 27:187–204. doi: 10.1007/s41324-018-0224-9
 118. Ahmad H, Ali A, Fatima SH, Zaidi F, Khisroon M, Rasheed SB, et al. Spatial modeling of Dengue prevalence and kringing prediction of Dengue outbreak in Khyber Pakhtunkhwa (Pakistan) using presence only data. *Stochast Environ Res Risk Assess*. (2020) 34:1023–36. doi: 10.1007/s00477-020-01818-9
 119. Phanitchat T, Zhao B, Haque U, Pientong C, Ekalaksananan T, Aromseree S, et al. Spatial and temporal patterns of dengue incidence in northeastern Thailand 2006–2016. *BMC Infect Dis*. (2019) 19:743. doi: 10.1186/s12879-019-4379-3
 120. Chen Y, Ong JHY, Rajarethinam J, Yap G, Ng LC, Cook AR. Neighbourhood level real-time forecasting of dengue cases in tropical urban Singapore. *BMC Med*. (2018) 16:129. doi: 10.1186/s12916-018-1108-5
 121. Jat MK, Mala S. Application of GIS and space-time scan statistic for vector born disease clustering. In: *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance*. New Delhi (2017). p. 329–38. doi: 10.1145/3047273.3047361
 122. Bomfim R, Pei S, Shaman J, Yamana T, Makse HA, Andrade JS Jr., et al. Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas. *J R Soc Interface*. (2020) 17:20200691. doi: 10.1098/rsif.2020.0691
 123. Ramadona AL, Lazuardi L, Hii YL, Holmner, Kusnanto H, Rocklöv. Prediction of dengue outbreaks based on disease surveillance and meteorological data. *PLoS ONE*. (2016) 11:e0152688. doi: 10.1371/journal.pone.0152688
 124. Benedum CM, Shea KM, Jenkins HE, Kim LY, Markuzon N. Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore. *PLoS Neglect Trop Dis*. (2020) 14:e0008710. doi: 10.1371/journal.pntd.008710
 125. Agarwal N, Koti SR, Saran S, Senthil Kumar A. Data mining techniques for predicting dengue outbreak in geospatial domain using weather parameters for New Delhi, India. *Curr Sci*. (2018) 114:2281–91. doi: 10.18520/cs/v114/i11/2281-2291
 126. Harumy T, Chan H, Sodhy G. Prediction for dengue fever in Indonesia using neural network and regression method. In: *Journal of Physics: Conference Series*. Medan: IOP Publishing (2020). p. 012019. doi: 10.1088/1742-6596/1566/1/012019
 127. Yamamoto N, Lee H, Nishiura H. Exploring the mechanisms behind the country-specific time of Zika virus importation. *Math Biosci Eng*. (2019) 16:3272–84. doi: 10.3934/mbe.2019163
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Lima, da Silva, Moreno, Cordeiro da Silva, Musah, Aldosery, Dutra, Ambrizzi, Borges, Tunali, Basibuyuk, Yenigün, Massoni, Browning, Jones, Campos, Kostkova, Silva Filho and dos Santos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.