

## RESEARCH ARTICLE

# A systematic review of dengue outbreak prediction models: Current scenario and future directions

Xing Yu Leung<sup>1</sup>, Rakibul M. Islam<sup>1</sup>, Mohammadmehdi Adhami<sup>1</sup>, Dragan Ilic<sup>1</sup>, Lara McDonald<sup>1</sup>, Shanika Palawaththa<sup>1</sup>, Basia Diug<sup>1</sup>, Saif U. Munshi<sup>2</sup>, Md Nazmul Karim<sup>1\*</sup>

**1** School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia,

**2** Department of Virology, Bangabandhu Sheikh Mujib Medical University, Dhaka, Bangladesh

\* [nazmul.karim@monash.edu](mailto:nazmul.karim@monash.edu)



## OPEN ACCESS

**Citation:** Leung XY, Islam RM, Adhami M, Ilic D, McDonald L, Palawaththa S, et al. (2023) A systematic review of dengue outbreak prediction models: Current scenario and future directions. *PLoS Negl Trop Dis* 17(2): e0010631. <https://doi.org/10.1371/journal.pntd.0010631>

**Editor:** Husain Poonawala, Tufts Medical Center, UNITED STATES

**Received:** July 5, 2022

**Accepted:** January 29, 2023

**Published:** February 13, 2023

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pntd.0010631>

**Copyright:** © 2023 Leung et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting Information](#) files.

## Abstract

Dengue is among the fastest-spreading vector-borne infectious disease, with outbreaks often overwhelm the health system and result in huge morbidity and mortality in its endemic populations in the absence of an efficient warning system. A large number of prediction models are currently in use globally. As such, this study aimed to systematically review the published literature that used quantitative models to predict dengue outbreaks and provide insights about the current practices. A systematic search was undertaken, using the Ovid MEDLINE, EMBASE, Scopus and Web of Science databases for published citations, without time or geographical restrictions. Study selection, data extraction and management process were devised in accordance with the 'Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies' ('CHARMS') framework. A total of 99 models were included in the review from 64 studies. Most models sourced climate (94.7%) and climate change (77.8%) data from agency reports and only 59.6% of the models adjusted for reporting time lag. All included models used climate predictors; 70.7% of them were built with only climate factors. Climate factors were used in combination with climate change factors (13.4%), both climate change and demographic factors (3.1%), vector factors (6.3%), and demographic factors (5.2%). Machine learning techniques were used for 39.4% of the models. Of these, random forest (15.4%), neural networks (23.1%) and ensemble models (10.3%) were notable. Among the statistical (60.6%) models, linear regression (18.3%), Poisson regression (18.3%), generalized additive models (16.7%) and time series/autoregressive models (26.7%) were notable. Around 20.2% of the models reported no validation at all and only 5.2% reported external validation. The reporting of methodology and model performance measures were inadequate in many of the existing prediction models. This review collates plausible predictors and methodological approaches, which will contribute to robust modelling in diverse settings and populations.

**Funding:** The author(s) received no specific funding for this work.

**Competing interests:** The authors declare that they have no conflicts of interest.

## Author summary

Dengue is considered as a major public health challenge and a life-threatening disease affecting people worldwide. Over the past decades, numerous forecast models have been developed to predict dengue incidence using various factors based on different geographical locations. Dengue transmission appears to be highly sensitive to climate variability and change, however quantitative models used to assess the relationship between climate change and dengue often differ due to their distribution assumptions, the nature of the relationship and the spatial and/or temporal dynamics of the response. We performed a systematic review to examine current literature surrounding existing quantitative models based on development methodology, predictor variable used and model performance. Our analysis demonstrates several shortcomings in current modelling practice, and advocates for the use of real time primary predictor data, the incorporation of non-climatic parameters as predictors and more comprehensive reporting of model development techniques and validation.

This review collates methodological approaches adopted in the modelling practices in the field across current literature. This will provide an evidence-based framework for upgrading future modelling practice to develop more accurate predictive models with robust techniques. In turn, this also provided an opportunity for the effective distribution of limited public health resources to prepare for demand.

## Introduction

Dengue fever is one of the fastest-spreading mosquito-borne disease primarily of tropical and subtropical regions and is caused by various dengue virus strains [1,2]. In 2017 alone, over 100 million people were estimated to have acquired the infection, contributing to a globally increasing burden of disease [3]. Although most infections are mild, dengue shock syndrome and dengue haemorrhagic fever are severe forms of infections and can be fatal [4,5]. The case-fatality rate can be as high as 20% in the absence of prompt diagnosis and lack of specific antiviral drugs or vaccines [6,7], particularly in resource-limited settings. When an outbreak is particularly large, the influx of severe dengue cases can overwhelm the health system and prevent optimal care. Dengue also imposes an enormous societal and economic burden on many of the tropical countries where the disease is endemic [8]. An accurate prediction of the size of the outbreak and trends in disease incidence early enough can limit further transmission [5], and is likely to facilitate planning the allocation of healthcare resources to meet the demand during an outbreak.

Vector-borne pathogens characteristically demonstrate spatial heterogeneity—a result of spatial variation in vector habitat, climate patterns and subsequent human control actions [9–11]. The interplay of human, climate and mosquito dynamics give rise to a complex system that determines the pattern of dengue transmission, which in turn influences the potential for outbreak [12]. These relationships have been explored over the decades in the development of predictive models worldwide. Models vary widely in their purposes [13–15] and settings [16–21]. Many of these models excel at different tasks, however for a prediction model to be efficient, it requires a systematic, self-adaptive and generalizable framework capable of identifying weather and population susceptibility patterns across geographic regions. The scientific community has not yet agreed upon a model that provides the best prediction. The selection of predictors for the existing models is also quite heterogeneous. Some models rely solely on climate variables [16], some include vector characteristics [17,18] others use population

characteristics [19–21]. A wide range of statistical techniques are used with varying degrees of accuracy and robustness among the existing models [16–21].

Clarity in the documentation of the model development processes and model performance are essential for ensuring the robustness of the prediction [22], which is scarce as many of the existing models have not yet been systematically appraised. Given the disparate approaches, a focused synthesis and appraisal of the existing models, along with their building techniques and factor catchments, is required. Carefully establishing these details will provide the foundation for updating and developing robust models in future. This study aimed to systematically review all published literature that reported quantitative models to predict dengue outbreaks, revealing several shortcomings in the usage of real time primary predictive data and non-climatic predictors in the development of models, as well as inadequate reporting of techniques, model and performance measure validation.

## Methods

### Search strategy and selection criteria

This systematic review's aim, search strategy and study selection process were devised in accordance with the seven items in the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies ('CHARMS') framework [23]. CHARMS framework is a systematic review tool, devised to facilitate and guide the methodological aspects the systematic review of prediction modelling studies, ranging from question development, appraisal of studies, and data extraction thereof. Detail of the CHARMS checklist can be found elsewhere [23]. The review followed the Preferred Reporting Items for Systematic Review and Meta-Analysis ('PRISMA') guidelines [24], and was registered in PROSPERO (CRD42018102100).

A literature search was conducted from inception until October 2022 using the electronic databases of Ovid MEDLINE, Embase, Scopus and Web of Science to obtain the information on the statistical models for predicting the number of dengue cases based on climatic factors. Google Scholar and the bibliography of included papers were also searched. The search strategies were developed under the guidance of an information specialist from Monash University Library. For the purposes of this study, dengue fever or dengue haemorrhagic fever or dengue shock syndrome were considered as a single entity "dengue". Search strategy included Medical Subject Headings ('MeSH') and keyword terms including "dengue", "severe dengue," "weather," "climate change," "model," "predict," and "forecast." The detailed search strategy and history are presented in [S1 Table](#).

The review included studies focused on (1) prognostic prediction models which aim to review models predicting future events, (2) incidence of dengue fever or dengue haemorrhagic fever cases, (3) models to be used to predict the number of cases prior to an outbreaks, (4) models intended to inform public health divisions of future dengue outbreaks, (5) models with no restrictions on the time span of prediction and (6) prediction model development studies without external validation, or with external validation in independent data. Peer-reviewed original articles that presented a model and were available as full-text articles were considered eligible if they focused on predicting the number of dengue cases or an outbreak based on number of dengue incidence. Articles that focused on updating previously developed models were only included if they presented an updated version of the model. Articles which dealt exclusively with dengue in international travellers, or which only analyse the correlation between climate parameters and dengue cases without presenting a prediction model were excluded. Furthermore, articles which used models for predicting the population of dengue vectors (e.g., *Aedes aegypti* or *Aedes albopictus*) as well as articles which only offer susceptible-

infected-recovered modelling stochastic or transmission rates modelling were excluded. Articles which presented a model only dealing with spatial or temporal components of dengue risk were considered ineligible. Conference proceedings, book chapters, abstracts or letters were also excluded. Titles and abstracts of the retrieved articles were screened independently by two reviewers (RMI, MMA). Two review team members (LM, XYL) then retrieved the full text of those potentially eligible studies and independently assessed their eligibility. Disagreements were resolved by a third reviewer (MNK). A detailed study selection process is illustrated in the PRISMA flow diagram (Fig 1) [24].

## Data analysis

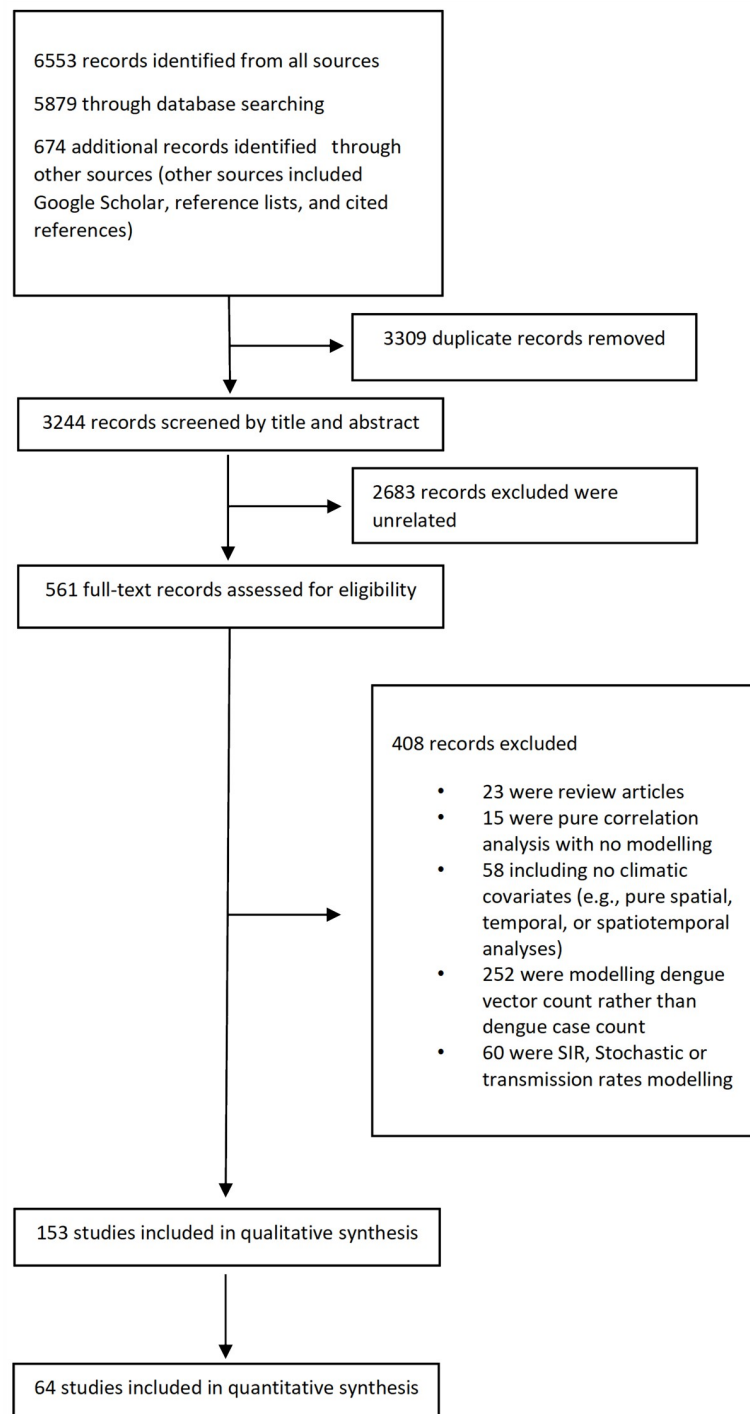
Based on the data extraction fields of the CHARMS framework [23], a standardised table was developed to extract data from the selected studies for assessment of quality and evidence synthesis. The data extraction table consists of eleven domains, each with a specific item, that extract data from the reports of the primary forecasting model. Key information extracted from the included articles were period and geographical region, sources of data, outcomes to be predicted, modelling covariates variables, sample size, statistical techniques, model performances, model evaluation, and key findings. Information regarding handling and/or reporting of missing data was also extracted. Each paper was independently reviewed by two reviewers (MMA, XYL) and discrepancies were resolved through discussion with each other or with a third reviewer (SP) where necessary.

Extracted data from the selected studies were summarised and the key information about the methodological characteristics of these models were tabulated. Descriptive statistics were generated based on model characteristics and comparative methodological features such as outcome types, target population, data sources and predictor selection techniques. All statistical analyses were performed using Stata (version 17.0).

## Results

The initial search yielded 6553 studies. After duplicates were removed, 3244 studies were screened for titles and abstracts. This led to 153 studies for full text review, and 64 that strictly met the inclusion criteria (Fig 1), 16 of these studies reported multiple models. A total of 99 models from 64 selected studies were identified. Characteristics of the models including, year, country and source of data used, predictors and outcome of the models, overall model development technique and model performance related variables are summarised in Table 1 [14,15,17–21,25–81].

Table 2 presents the sources of data used for modelling. Most of the models (90.7%) sourced dengue incidence data from surveillance and 44.3% used registry data, while 34.0% also used hospital or laboratory data. While most (94.7%) of the models used climate data from government agency reports, only around 22.1% of the models used data from the meteorological stations in real-time. Climate change data was also sourced mostly (77.8%) from government agency reports, only 11.1% used international environmental agency data and 22.2% used local environmental agency report. Half (50.0%) of the models used vector data from entomological surveillance and 25.0% used vector data from laboratory sources. Around 83.8% of the models were built based on the sample from general population, 16.2% used only urban samples. Around 46.9% of the models used monthly aggregate data, over a third (29.3%) used weekly aggregate data and 23.2% used daily aggregated data of the predictors. The majority (59.6%) of the models incorporated reporting time lag adjustment. Although 17.2% of the models addressed the missing data, 30.3% did not address the issue, while the majority (52.5%) did not specifically report the missing value. Around 80.8% of the models were intended to



**Fig 1. PRISMA flow diagram illustrating study selection process.**

<https://doi.org/10.1371/journal.pntd.0010631.g001>

Table 1. Characteristics of included predictive models.

Author and year	Country	Source of dengue data	Participant's recruitment	Outcome	Candidate predictors	Sample size	Missing data	Model development	Model techniques used	Model performance	Evaluation
Abualamrah et al. (2021) [25]	Saudi Arabia	Surveillance, notification	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	BIC	Internal only
Aburas et al. (2010) [17]	Singapore	Notification	Urban	Case	Climate	14209	Not reported	All variable	Machine learning	Not reported	Internal only
Adde et al. (2016) [26]	France	Surveillance, laboratory	Urban	Outbreak	Climate, climate change	Not reported	Not reported	Pre-selection	Statistical model	AIC	Internal only
Anggraeni et al. (2017) [27]	Indonesia	Notification, registry	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	MAPE	Internal only
Bal et al. (2020) [28]	India	Notification	Urban	Case	Climate	Not reported	Not reported	All variable	Statistical model	AIC	Internal and external
Banu et al. (2015) [29]	Bangladesh	Notification	Both	Outbreak	Climate, climate change	Not reported	Not reported	Pre-selection	Statistical model	AIC, R <sup>2</sup>	Internal
Baquero et al. (2018) [30]	Brazil	Notification, hospital, registry	Both	Case	Climate	Not reported	Reported	Pre-selection, all variable	Statistical model, machine learning	RMSE	Internal only
Betanzos-Reyes et al. (2018) [20]	Mexico	Notification	Both	Case	Climate, vector	2526	Not reported	Pre-selection	Statistical model	R <sup>2</sup>	Not reported
Bett et al. (2019) [31]	Vietnam	Notification	Urban	Case	Climate, demography, environment	Not reported	Not reported	Pre-selection	Statistical model	Theil's coefficient of inequality	Internal only
Bouzid et al. (2014) [32]	Mexico	Laboratory	Urban	Case	Climate, demography	Not reported	Not reported	All variable	Statistical model	GCV	Internal only
Buczak et al. (2018) [33]	Peru	Surveillance, laboratory	Both	Case	Climate, environment	Not reported	Not reported	All variable	Machine learning	RMSE, MARE	Not reported
Buczak et al. (2018) [33]	Puerto Rico	Surveillance, laboratory	Both	Case	Climate, environment	Not reported	Not reported	All variable	Machine learning	RMSE, MARE	Not reported
Carvajal et al. (2018) [34]	Philippines	Notification	Urban	Case	Climate, climate change	Not reported	Reported	Pre-selection	Statistical model, machine learning	AIC, R <sup>2</sup>	Internal only
Chang et al. (2015) [21]	Taiwan	Surveillance, laboratory	Both	Case	Climate, vector	8918	Not reported	All variable	Statistical model	AIC, AUC, ROC	Internal only
Chen et al. (2022) [35]	China	Surveillance, registry	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	R <sup>2</sup> , GCV	Internal only
Chen et al. (2022) [36]	China	Surveillance, registry	Both	Outbreak	Climate, vector	Not reported	Not reported	All variable	Machine learning	Retrospective forecasts	Internal only
Cheng et al. (2020) [37]	Taiwan	Notification, registry	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	AIC, BIC, MAE, RMSE	Internal only
Chuang et al. (2017) [38]	Taiwan	Notification	Urban	Outbreak	Climate, climate change	71793	Not reported	Pre-section	Statistical model	AIC	Internal only
Colón-González et al. (2013) [39]	Mexico	Surveillance	Urban	Case	Climate, demography, environment	Not reported	Not reported	All variable	Statistical model	Not reported	Internal only

(Continued)

Table 1. (Continued)

Author and year	Country	Source of dengue data	Participant's recruitment	Outcome	Candidate predictors	Sample size	Missing data	Model development	Model techniques used	Model performance	Evaluation
Depradine et al. (2004) [40]	Barbados	Notification, laboratory	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	R <sup>2</sup>	Not reported
Descloux et al. (2012) [41]	New Caledonia	Surveillance, laboratory, hospital	Both	Outbreak	Climate, climate change, vector	Not reported	Not reported	Pre-selection	Statistical model	AIC	Internal only
Dey et al. (2022) [42]	Bangladesh	Notification, registry	Both	Case	Climate	Not reported	Reported	All variable	Machine learning	MAE	Internal only
Dharmawardana et al. (2017) [43]	Sri Lanka	Notification	Both	Outbreak	Climate, environment	Not reported	Not reported	All variable	Machine learning	R <sup>2</sup>	Not reported
Earnest et al. (2012) [44]	Singapore	Notification, laboratory	Both	Outbreak	Climate, climate change	Not reported	Not reported	All variable	Statistical model	MAPE, R <sup>2</sup>	Not reported
Edussuriya et al. (2021) [45]	Sri Lanka	Notification	Both	Case	Climate, demography	Not reported	Reported	Pre-selection	Machine learning	RMSE	Not reported
Gharbi et al. (2011) [18]	France	Notification, laboratory	Both	Outbreak	Climate	Not reported	Not reported	All variable	Statistical model	AIC, RMSE	Not reported
Guo et al. (2017) [46]	China	Notification	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model, machine learning	RMSE, R <sup>2</sup>	Internal only
Jain et al. (2019) [47]	Thailand	Notification	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	RMSE, R <sup>2</sup> , AIC	Internal and external
Jayaraj et al. (2019) [48]	Malaysia	Notification, laboratory	Both	Both	Climate	Not reported	Not reported	Pre-selection	Statistical model	MAE, MSE	Internal only
Karim et al. (2012) [15]	Bangladesh	Notification	Urban	Case	Climate	22705	Not reported	All variable	Statistical model	ROC	Internal only
Lauer et al. (2018) [49]	Thailand	Notification, registry	Both	Case	Climate	Not reported	Not reported	Pre-selection	Statistical model	MAE	Internal only
Li et al. (2017) [50]	China	Notification	Both	Case	Climate	Not reported	Not reported	Pre-selection	Statistical model	Not reported	Internal only
Li et al. (2022) [51]	Brazil	Registry, hospital, laboratory	Both	Case	Climate, environment	Not reported	Not reported	Pre-selection	Machine learning	RMSE, MAE	Internal only
Liu et al. (2019) [52]	China	Notification, registry	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	RMSE, R <sup>2</sup> , AIC	Internal only
Lowe et al. (2013) [53]	Brazil	Notification, laboratory	Urban	Case	Climate, climate change, demography	Not reported	Not reported	Pre-selection	Statistical model	AUC	External only
Luz et al. (2008) [54]	Brazil	Notification, registry	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	AIC, ACF, PACF	Internal only
McGough et al. (2021) [55]	Brazil	Notification, registry	Both	Case	Climate	Not reported	Not reported	Pre-selection	Machine learning	Out-of-sample forecast	External only
Mincham et al. (2019) [56]	China	Notification, laboratory	Both	Case	Climate, vector	Not reported	Not reported	All variable	Machine learning	sensitivity, specificity, PPV, NPV	External only

(Continued)

Table 1. (Continued)

Author and year	Country	Source of dengue data	Participant's recruitment	Outcome	Candidate predictors	Sample size	Missing data	Model development	Model techniques used	Model performance	Evaluation
Nakhapakorn et al. (2005) [57]	Thailand	Notification	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	Not reported	Internal only
Nan et al. (2018) [58]	China	Notification, registry	Both	Case	Climate	Not reported	Not reported	Pre-selection	Machine learning	RMSE, MAE, R <sup>2</sup>	Internal only
Nguyen et al. (2022) [59]	Vietnam	Surveillance, registry	Both	Case	Climate	Not reported	Reported	All variable	Machine learning	MAE	Internal only
Nuraini et al. (2021) [60]	Indonesia	Notification	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	Least square method	Not reported
Olmoguez et al. (2019) [61]	Philippines	Notification	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model, machine learning	R <sup>2</sup> , MAPE	Not reported
Pham et al. (2018) [62]	Malaysia	Notification, registry	Both	Case	Climate	Not reported	Not reported	Pre-selection	Statistical model, machine learning	RMSE, MAE, Scatter plot	Not reported
Pham et al. (2020) [63]	Vietnam	Notification	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	Retrospective forecasts	Internal only
Phung et al. (2015) [64]	Vietnam	Notification, laboratory	Both	Case	Climate	13509	Not reported	All variable	Statistical model	AIC, BIC, ROC	Internal only
Phung et al. (2016) [14]	Vietnam	Surveillance/ notification	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	ROC	Internal only
Pineda et al. (2019) [65]	Philippines	Notification	Both	Case	Climate	Not reported	Not reported	All variable	Statistical model	Retrospective forecasts	Internal only
Pinto et al. (2011) [66]	Singapore	Notification	Urban	Case	Climate	Not reported	Not reported	All variable	Statistical model	Not reported	Internal only
Puengpreeda et al. (2020) [67]	Thailand	Notification, registry	Both	Case	Climate	Not reported	Addressed	All variable	Machine learning	MSE, MAE, MAPE, R <sup>2</sup>	Internal only
Qureshi et al. (2017) [19]	Pakistan	Notification	Urban	Outbreak	Climate, vector	Not reported	Not reported	All variable	Machine learning	Not reported	Not reported
Ramadona et al. (2016) [68]	Indonesia	Notification	Both	Case	Climate	7171	Not reported	All variable	Statistical model	AIC, RMSE	External only
Roster et al. (2021) [69]	Brazil	Notification	Both	Outbreak	Climate	Not reported	Not reported	All variable	Machine learning	MAE, RMSE	Internal only
Salim et al. (2021) [70]	Malaysia	Notification, registry, laboratory	Both	Outbreak	Climate, vector	Not reported	Not reported	All variable	Statistical model, machine learning	ROC	Internal only
Shi et al. (2016) [71]	Singapore	Notification	Urban	Case	Climate, vector	Not reported	Not reported	All variable	Statistical model	MAPE	Internal only
Siriyasatien et al. (2016) [72]	Thailand	Notification	Both	Case	Climate, vector, demography	Not reported	Not reported	All variable	Statistical model	AIC, BIC	Internal only
Withanage et al. (2018) [73]	Sri Lanka	Notification	Both	Case	Climate	56843	Not reported	All variable	Statistical model	MAPE, RMSE, MAE, PSS	Internal only

(Continued)



Table 1. (Continued)

Author and year	Country	Source of dengue data	Participant's recruitment	Outcome	Candidate predictors	Sample size	Missing data	Model development	Model techniques used	Model performance	Evaluation
Xu et al. (2020) [74]	China	Notification	Both	Case	Climate	Not reported	Not reported	Pre-selection	Machine learning	Not reported	Internal only
Yuan et al. (2019) [75]	Taiwan	Notification, registry	Both	Case	Climate	Not reported	Not reported	Pre-selection	Statistical model	AIC	Internal only
Yuan et al. (2020) [76]	China	Notification, laboratory	Both	Case	Climate	Not reported	Not reported	Pre-selection	Statistical model	MSE, NMSE	Internal only
Zafra, B (2020) [77]	Philippines	Notification, registry	Both	Case	Climate	Not reported	Reported	All variable	Machine learning	Not reported	Internal only
Zambrano et al. (2012) [78]	Honduras	Laboratory	Both	Case	Climate, climate change	3353	Not reported	All variable	Statistical model	Not reported	Not reported
Zhang et al. (2016) [79]	China	Notification, laboratory	Both	Case	Climate	38150	Not reported	All variable	Statistical model	AIC, ROC	Internal only
Zhao et al. (2020) [80]	Colombia	Notification	Both	Case	Climate, climate change, demography, environment	Not reported	Not reported	All variable	Machine learning	MAE, RMAE	Internal only
Zhu et al. (2019) [81]	China	Notification, laboratory	Both	Case	Climate	Not reported	Not reported	Pre-selection	Statistical model	R <sup>2</sup>	Not reported

AIC = Akaike information criterion, ACF = autocorrelation function, AUC = area under curve, BIC = Bayesian information criterion, GCV = generalized cross-validation score, MAE = mean absolute error, MAPE = mean absolute percentage error, MARE = mean absolute relative error, MSE = mean squared error, PACF = partial autocorrelation function, NMSE = normalised mean squared error, PSS = Pierce skill score, RMAE = relative mean absolute error, ROC = receiver operating characteristic, RMSE = root mean squared error, PPV = Positive predictive value, NPV = Negative predictive value

<https://doi.org/10.1371/journal.pntd.0010631.t001>

**Table 2. Source of data used for modelling.**

Sources of model data	N (%) *
<i>Dengue data source (n = 99)</i>	
Surveillance/notification	88(90.7)
Disease registry	43(44.3)
Hospital/laboratory	33(34.0)
<i>Climate data source (n = 99)</i>	
Government agency report	90 (94.7)
Meteorology station	21 (22.1)
Research Institute/centre	4 (4.2)
<i>Climate change data source (n = 18)</i>	
Government agency report	14 (77.8)
International environmental agency	2 (11.1)
Local environmental agency report	4 (22.2)
<i>Vector data source (n = 8)</i>	
Surveillance/monitoring data	4 (50.0)
Laboratory data	2 (25.0)
Government agency report	3 (37.5)
<i>Population source (n = 99)</i>	
General population	83 (83.8)
Metropolitan	16 (16.2)
<i>Data aggregation unit (n = 98)</i>	
Daily aggregate	23 (23.2)
Weekly aggregate	29 (29.3)
Monthly aggregate	46 (46.9)
<i>Lag time adjusted in model (n = 99)</i>	
No	40 (40.4)
Yes	59 (59.6)
<i>Treatment of missing data (n = 99)</i>	
Yes	17 (17.2)
No	30(30.3)
Not reported	52 (52.5)
<i>Prediction outcome (n = 99)</i>	
Dengue case count	80 (80.8)
Dengue outbreak	19 (19.2)

\* The percentages may not add up to 100, due to multiple responses

<https://doi.org/10.1371/journal.pntd.0010631.t002>

predict the number of dengue cases and 19.2% focused on predicting dengue outbreaks, based on predetermined case number threshold.

Table 3 summarises the statistical methods adopted by the prediction models. Modelling techniques were broadly categorised under two genres, statistical models (60.6%) and machine learning (39.4%). The statistical models were broadly comprised of linear regression models (18.3%), time series/autoregressive models (26.7%), Poisson regression models (18.3%) and generalized additive models (16.7%). Neural networks models (23.1%), random forest models (15.4%), and ensemble models (10.3%) were types of machine learning models used.

All theoretically plausible predictors were considered as candidate predictors in 71.7% of models and pre-selection of predictors based on unadjusted association with the model outcome was considered in 28.3% of models. Reporting of essential modelling techniques was

**Table 3. Statistical methods used among models (n = 99).**

Statistical methods	N (%) *
<i>Model building technique</i>	
<i>Statistical models (n = 60, 60.6%)</i>	
Linear regression model	11 (18.3)
Non-linear regression model	5 (8.3)
Time series/autoregressive model	16 (26.7)
Poisson regression model	11 (18.3)
Generalized linear model (GLM)	2 (3.3)
Generalized additive model (GAM)	10 (16.7)
Others	5 (8.3)
<i>Machine learning (n = 39, 39.4%)</i>	
Random forest	6 (15.4)
Neural network	9 (23.1)
Boosting algorithm	5 (12.8)
Support vector algorithm (SVA)	4 (10.3)
Ensemble models	4 (10.3)
Classification and regression tree (CART)	2 (5.1)
Long short-term memory (LSTM)	4 (10.3)
Others	5 (12.8)
<i>Predictor selection for model (n = 99)</i>	
All theoretically plausible predictors	71 (71.7)
Pre-selection (unadjusted association)	28 (28.3)
<i>Reporting model parameter (n = 99)</i>	
Model performance	84 (84.8)
Model calibration	58 (58.6)
Model discrimination	47 (47.5)
<i>Model validation (n = 99)</i>	
External and internal validation	5 (5.2)
Internal validation	75 (75.8)
No validation	20 (20.2)
<i>Model validation techniques (n = 79)</i>	
Split sample validation	16 (20.3)
Cross validation	32 (40.5)
Retrospective validation	5 (6.3)
Out of sample validation	3 (3.8)
Performance metrics only	23 (29.1)
<i>Performance metrics reported</i>	
MAE (Mean Absolute Error)	7 (7.1)
RMSE (Root Mean Squared Error)	11 (11.1)
AIC/BIC (Akaike / Bayesian Information Criterion)	1 (1.0)
ROC (Receiver Operating Characteristic)	5 (5.1)
MAPE (Mean Absolute Percentage Error)	5 (5.1)
GCV (Generalized Cross Validation score)	2 (2.0)
MSE (Mean Squared Error)	7 (7.3)

\* The percentages may not add up to 100 as studies used multiple methods

<https://doi.org/10.1371/journal.pntd.0010631.t003>

heterogeneous— 84.8% of models reported model performance, 58.6% reported model calibration and 47.5% reported model discrimination. Among the performance metrics, Root Mean Squared Error (‘RMSE’) (11.1%), Mean Squared Error (‘MSE’) (7.3%), Mean Absolute Percentage Error (‘MAPE’) (5.1%), and Receiver Operating Characteristic (‘ROC’) (5.1%) were notable. Of these models, most (75.8%) reported the internal validation alone, only 5.2% reported both internal and external validation and 20.2% reported no validation at all. The validation techniques included: split sample validation (development and validation) (20.3%), cross validation, which involves resampling of the derivation sample (40.5%) and performance metrics (29.1%).

Table 4 presents the factors used for prediction models. All of the models included in the review used climate predictors in their model. Among the climate predictors: humidity (77.4%), temperature (95.2%) and rainfall (81.0%), were used in most models. Windspeed and direction (27.4%), precipitation (15.5%) and sunshine (10.7%) were among other notable climate factors. Considering the similarity of the description of factors, climate change and environmental factors were collapsed in to one category under climate change. Overall, 18.2% of the models used climate change and/or environmental predictors. El Nino-Southern Oscillation (‘ENSO’), Southern Oscillation Index (‘SOI’), Oceanic Nino Index (‘ONI’), hydric balance and vegetation index were among the key climate change predictors. Vegetation Index and enhanced vegetation index were among the key environmental factors reported.

Vector-related predictors were included in 8.1% of models, and the key vector related predictors were container index, Breteau index, adult productivity index, breeding percentage and mosquito infection rate. Demographic predictors were included in 8.1% of models, and key demographic predictors were, population size, population density, area under the urban settlement, access to piped water, education coverage and GDP per capita (Table 4).

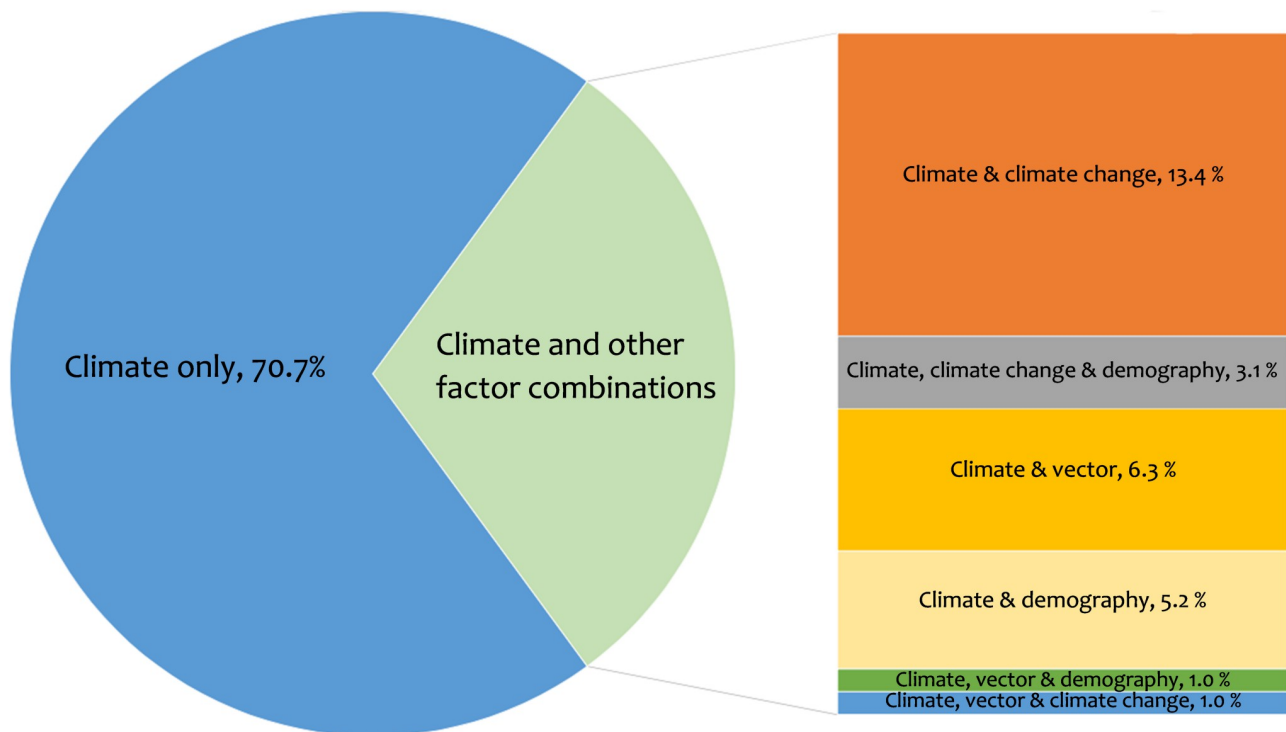
The combination of the predictors used in the model are depicted in Fig 2. While majority (70.7%) of the models were built solely on climate predictors, none of the models used the

**Table 4. Factors that appeared as predictors in the prediction models.**

Climate factors (100.0%)	Climate change and environmental factors (18.2%)	Entomological (Vector) factors (8.1%)	Demographic factors (8.1%)
<i>Temperature (95.2%)</i>	<i>El Nino-Southern Oscillation (ENSO)</i>	<i>Ades albopictus count</i>	<i>Population size</i>
Minimum temperature	<i>Southern Oscillation Index (SOI)</i>	<i>Container index</i>	<i>Population density</i>
Mean temperature	<i>Oceanic Nino Index (ONI)</i>	<i>Ades aegypti index</i>	<i>Access to piped water</i>
Maximum temperature	<i>Gini Index</i>	<i>Breteau index</i>	<i>Education coverage</i>
<i>Rainfall (81.0%)</i>	<i>Potential evapotranspiration</i>	<i>Adult productivity index</i>	<i>GDP per capita</i>
Average rainfall	<i>Azores high sea-level pressure</i>	<i>Weekly egg count in ovitrap</i>	<i>Area under urban settlement</i>
Accumulated rainfall	<i>Dipole mode index</i>	<i>Breeding percentage</i>	
Number of rainy days	<i>Hydric balance</i>	<i>Mosquito infection rate</i>	
<i>Humidity (77.4%)</i>	<i>Vegetation Index</i>	<i>Minimum infection rate</i>	
Relative humidity	<i>Enhanced vegetation index</i>	<i>Per man hour density (PMHD)</i>	
Absolute humidity	<i>Equatorial Pacific Ocean surface temperature</i>		
<i>Sunshine (10.7%)</i>			
Sunshine duration			
Insolation			
<i>Windspeed &amp; direction (27.4%)</i>			
<i>Precipitation (15.5%)</i>			
<i>Evaporation (8.3%)</i>			
<i>Atmospheric pressure (2.4%)</i>			

\* Figures in the parenthesis denotes the percentage of models included the predictor in the model and percentages may not add up to 100 as models used multiple categories of predictors in combination

<https://doi.org/10.1371/journal.pntd.0010631.t004>



**Fig 2. Combinations of predictors used in the prediction models.**

<https://doi.org/10.1371/journal.pntd.0010631.g002>

combination of all four (climate, climate change, vector and demography) categories of predictors. The combination of climate, climate change and demographic predictors was used in 3.1% of the models and the combination of climate and climate change predictors were used in 13.4% models. Among other notable combinations were, climate and vector predictors (6.3%) and climate and demographic predictor (5.2%).

## Discussion

This systematic review evaluated 99 dengue outbreak prediction models from 64 studies, predicting the number of dengue cases or outbreaks from a variety of settings and populations. Our review identified, three major area of inadequacy in the current modelling practices. Firstly, use of secondary predictor data—acquired from reports—were quite prevalent among models. Secondly, as data for other non-climate variables were not included in the majority of the models, they failed to capture a holistic view of dengue development in the prediction process. Lastly, inadequacy in the reporting of methodology, model validation and performance measures were quite prevalent in the existing prediction models. One positive aspect seen in the current modelling practice is the shift toward robust modelling technique, such as use of machine learning algorithm and autoregressive time series techniques.

While effective treatments and prevention measures are still being developed, an early warning system for an epidemic has the potential to reduce the toll of severe disease on the health system and population [82]. Developing a clearer understanding of the factors affecting dengue transmission is an important step towards mitigating the impact of the disease on health systems and on communities at large. Early prediction of dengue incidences or alerts regarding impending outbreak may contribute to the health system preparedness through

effective resource mobilization and creating public awareness. Such predictions also have policy implications, as epidemiological evidence generated through modelling feeds the policy making process and facilitates the prioritization of interventions, such as vector control and environmental modification particularly in regard to climate change [83]. Considering dengue is a mosquito-borne disease, the majority of outbreak prediction models focus on climate dependency of mosquito breeding and dengue transmission [4–7]. While many models have been successful in predicting relative cases of dengue in real settings, incorrect prediction results have been observed in several included studies. For example, a model by Adde et al. [26] was unable to forecast a dengue outbreak in 2001–2005 with the use of climate data from 1991–2000. One of the potential reasons for their inaccurate prediction is the geo-spatial variation of climate and environment within regions. In their study, the decision to include vastly heterogeneous geographical areas led to variation in model prediction, which—be due to exclusion of non-climatic factors—may be the explanation of the poor performance in many of the earlier prediction models could. An increasing number of later models appears to incorporate a wide range of vector parameters as well as demographic parameters. Chang et al. pointed out that, entomological (vector) factor combined with other meteorological (climate) factors, have better prediction performance, and their prediction accuracy is often higher than that of climate predictors alone [21].

For dengue incidence data, the majority of the models relied on reports from government organizations based on notifiable data. Notification involves passive surveillance, where there is potential for systematic underreporting along with varying time lag. Modelling with data from active surveillance or real-time study may minimize such limitations. A considerable number of models did not consider the time lag affecting the prediction, which may be responsible for possible delays in weather affecting mosquito vectors and subsequently viruses. Due to the nature of dengue disease dynamics, failure to address time lag in model development is likely to affect prediction accuracy. Critical points in the natural history of disease timeline those may generate time lag may start with mosquito development, and subsequently also during acquisition and amplification virus in mosquitoes, mosquito host behaviour (i.e. biting and feeding pattern) and the incubation period of the virus in the human body [12,48]. Some studies have found a positive correlation between climate variables with time-lags at several points in the natural history of disease timeline [48,53]. Therefore, the adjustment for the time lags while predicting dengue is indispensable, especially when meteorological data is used [12].

The majority of included models were built on conventional regression techniques. According to recent literature, the time series technique is particularly considered effective in predicting the highly auto-correlated nature of dengue infection [84,85]. Machine learning techniques are employed in around 40% of the included models, and is particularly prevalent among the recently developed models. Batista et al. confirmed superiority with ML techniques demonstrating a lower error rate compared to the conventional statistics-based model in predicting dengue cases. In the age of big data, this technique can leverage data availability and in addition to being non-parametric in nature, can also provide some leeway in terms of strict assumption [86]. Random forest, neural networks, gradient boosting and support vector algorithms are notable subsets of machine learning algorithms, which have made significant contributions to several areas of public health, particularly in the forecasting of infectious diseases like malaria [87] and COVID-19 [88], and may have similar utility for making dengue outbreak predictions. Although machine learning is gaining popularity, future modelling in this area may benefit from using mechanistic models [89]. This modelling technique have played an essential role in shaping public health policy over the past decades [90]. Mechanistic models have the potential to provide additional insight regarding precise dynamics of the transmission and infection of

dengue. As these models highlight underlying processes that drive the patterns. These models can particularly aid in the prediction through incorporating the observed trajectory of vectors.

In the modelling process, generating an algorithm or equation is only part of the process. It is not complete unless its performance has been assessed considering discrimination [91] and calibration [92], both internally and using the population outside of what it is developed from, respectively. Among the existing models examined, reporting of the discrimination and calibration is very low. Without knowledge of model performance through validation in both source populations and populations other than where it was developed, objective evaluation of models is difficult [93]. Predictive models can be of great value only if there is certainty of its accuracy, that is, how precisely the model can predict an outcome in real world [94]. In the majority of the published models, real-world validation has not been performed or reported. Generally, models are unlikely to predict as well in real-world samples as it would in the derivation sample; this validity shrinkage can often be quite substantial. Hence, future models should report a mechanism of estimating and reporting potential validity shrinkage as well as predictive validity in real world data [95, 96].

In a substantial proportion of the models that reported validation, the original dataset was randomly split into the development and validation subset. Although this approach is widely used in many model validation settings, there are some setbacks when using smaller operational sample sizes, as split-sample analyses give overly pessimistic estimates of model performance and are accompanied by large variability [97]. Bootstrapping is generally considered to be the preferred internal validation method in predictive models [98, 99]. Interestingly, bootstrapping was not used in any of the models in included studies, instead cross-validation technique was adopted in most of them. External validation, on the other hand, was used only in very few included studies. This is despite the fact that external validation is considered pivotal to model development and a key indicator of model performance through highlighting applicability to participants, centres, regions or environments [23]. The external validation is particularly essential for model redevelopment, where the original model is adjusted, updated, or recalibrated based on validation data to improve performance [100]. This update may include adjusting the baseline risk (interception or hazard) of the original model, adjusting the weight or regression coefficient of the predictor, adding new predictors, or removing existing predictors from the model.

This review has a number of strengths—specifically, the use of the CHARMS checklist [23], designed for the assessment of the applicability of the prediction models. In addition, inclusion and exclusion criteria were strictly followed, and database searches were conducted by an expert librarian. However, there are a few limitations of the review—the models in this review are not explicitly rated based on quality or performance due to the lack of accepted criteria for rating the quality of forecasting models. In addition, although calibration was reported in several studies, calibration measures lack clarification, which may impact the overall evaluation of the model performance. The model performance could not be compared across methodological approaches in quantitative synthesis because of a lack of model performance data, and those that did provide data are mostly generated from internal validation data which may result in overfitting.

## Conclusion

In conclusion, failure to use of real time primary predictor data, failing to incorporate non-climatic parameters as predictor and insufficient reporting of model development techniques, model validation and performance measure were the major inadequacies identified in the current modelling practice. The paradigm shift towards robust modelling techniques, such as the use of machine learning algorithms and autoregressive time series, is a significant positive

trend in contemporary model practices. The findings of this review have the potential to lay the groundwork for improved modelling practices in the future. These findings will contribute to robust modelling in different settings and populations and have important implications for the planning and decision-making process for early dengue intervention and prevention.

## Supporting information

**S1 Table. Search strategy for OVID Medline, as performed in October 2022.**  
(DOCX)

**S2 Table. PRISMA checklist.**  
(DOCX)

## Author Contributions

**Conceptualization:** Dragan Ilic, Saif U. Munshi, Md Nazmul Karim.

**Data curation:** Xing Yu Leung, Rakibul M. Islam, Mohammadmehdi Adhami, Lara McDonald, Shanika Palawaththa.

**Formal analysis:** Md Nazmul Karim.

**Methodology:** Md Nazmul Karim.

**Project administration:** Xing Yu Leung, Rakibul M. Islam, Mohammadmehdi Adhami, Lara McDonald, Shanika Palawaththa.

**Supervision:** Dragan Ilic, Basia Diug.

**Visualization:** Shanika Palawaththa, Basia Diug.

**Writing – original draft:** Xing Yu Leung, Saif U. Munshi, Md Nazmul Karim.

**Writing – review & editing:** Xing Yu Leung, Rakibul M. Islam, Mohammadmehdi Adhami, Dragan Ilic, Lara McDonald, Shanika Palawaththa, Basia Diug, Saif U. Munshi, Md Nazmul Karim.

## References

1. Simmons CP, Farrar JJ, Chau Nguyen vV, Wills B. Dengue. *N Engl J Med.* 2012; 366: 1423–1432. <https://doi.org/10.1056/NEJMra1110265> PMID: 22494122
2. Tatem AJ, Hay SI, Rogers DJ. Global traffic and disease vector dispersal. *Proc Natl Acad Sci USA.* 2006; 103: 6242–6247. <https://doi.org/10.1073/pnas.0508391103> PMID: 16606847
3. Zeng Z, Zhan J, Chen L, Chen H, Cheng S. Global, regional, and national dengue burden from 1990 to 2017: A systematic analysis based on the global burden of disease study 2017. *EClinicalMedicine.* 2021; 32: 100712. <https://doi.org/10.1016/j.eclinm.2020.100712> PMID: 33681736
4. Rigau-Pérez JG, Clark GG, Gubler DJ, Reiter P, Sanders EJ, Vorndam AV. Dengue and dengue haemorrhagic fever. *Lancet.* 1998; 352: 971–977. [https://doi.org/10.1016/s0140-6736\(97\)12483-7](https://doi.org/10.1016/s0140-6736(97)12483-7) PMID: 9752834
5. Chen HL, Hsiao WH, Lee HC, Wu SC, Cheng JW. Selection and Characterization of DNA Aptamers Targeting All Four Serotypes of Dengue Viruses. *PLoS One.* 2015; 10: e0131240. <https://doi.org/10.1371/journal.pone.0131240> PMID: 26110785
6. Capeding MR, Tran NH, Hadinegoro SRS, Ismail HI, Chopitayasunondh T, Chua MN, et al. Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial. *Lancet.* 2014; 384: 1358–1365. [https://doi.org/10.1016/S0140-6736\(14\)61060-6](https://doi.org/10.1016/S0140-6736(14)61060-6) PMID: 25018116
7. Halstead SB. Dengue vaccine development: a 75% solution? *Lancet.* 2012; 380: 1535–1536. [https://doi.org/10.1016/S0140-6736\(12\)61510-4](https://doi.org/10.1016/S0140-6736(12)61510-4) PMID: 22975339
8. Halstead SB. Dengue. *Curr Opin Infect Dis.* 2002; 15: 471–476. <https://doi.org/10.1097/00001432-200210000-00003> PMID: 12686878



9. World Health Organization. Dengue and severe dengue. 2022 Jan 10 [cited 2022 July 27]. In: WHO Fact Sheets [Internet]. Geneva: World Health Organization. Available from: [who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue](https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue).
10. Mbogo CM, Mwangangi JM, Nzovu J, Gu W, Yan G, Gunter JT, et al. Spatial and temporal heterogeneity of Anopheles mosquitoes and Plasmodium falciparum transmission along the Kenyan coast. *Am J Trop Med Hyg*. 2003; 68: 734–742. PMID: [12887036](https://pubmed.ncbi.nlm.nih.gov/12887036/)
11. Acevedo MA, Prosper O, Lopiano K, Ruktanonchai N, Caughlin TT, Martcheva M, et al. Spatial heterogeneity, host movement and mosquito-borne disease transmission. *PLoS One*. 2015; 10: e0127552. <https://doi.org/10.1371/journal.pone.0127552> PMID: [26030769](https://pubmed.ncbi.nlm.nih.gov/26030769/)
12. Teurlai M, Menkès CE, Cavarero V, Degallier N, Descloux E, Grangeon JP, et al. Socio-economic and Climate Factors Associated with Dengue Fever Spatial Heterogeneity: A Worked Example in New Caledonia. *PLoS Negl Trop Dis*. 2015; 9: e0004211. <https://doi.org/10.1371/journal.pntd.0004211> PMID: [26624008](https://pubmed.ncbi.nlm.nih.gov/26624008/)
13. Medlock JM, Leach SA. Effect of climate change on vector-borne disease risk in the UK. *Lancet Infect Dis*. 2015; 15: 721–730. [https://doi.org/10.1016/S1473-3099\(15\)70091-5](https://doi.org/10.1016/S1473-3099(15)70091-5) PMID: [25808458](https://pubmed.ncbi.nlm.nih.gov/25808458/)
14. Phung D, Talukder M, Rutherford S, Chu C. A climate-based prediction model in the high-risk clusters of the Mekong Delta region, Vietnam: Towards improving dengue prevention and control. *Trop Med Int Health*. 2016; 21: 1324–1333. <https://doi.org/10.1111/tmi.12754> PMID: [27404323](https://pubmed.ncbi.nlm.nih.gov/27404323/)
15. Karim MN, Munshi SU, Anwar N, Alam MS. Climatic factors influencing dengue cases in Dhaka city: a model for dengue prediction. *Indian J Med Res*. 2012; 136: 32–39. PMID: [22885261](https://pubmed.ncbi.nlm.nih.gov/22885261/)
16. Benedum CM, Seidahmed OME, Eltahir EAB, Markuzon N. Statistical modeling of the effect of rainfall flushing on dengue transmission in Singapore. *PLoS Negl Trop Dis*. 2018; 12: e0006935. <https://doi.org/10.1371/journal.pntd.0006935> PMID: [30521523](https://pubmed.ncbi.nlm.nih.gov/30521523/)
17. Aburas H, Cetiner B, Sari M. Dengue confirmed-cases prediction: A neural network model. *Expert Syst Appl*. 2010; 37: 4256–4260.
18. Gharbi M, Quenel P, Gustave J, Cassadou S, Ruche GL, Girdary L, et al. Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. *BMC Infect Dis*. 2011; 11: 166. <https://doi.org/10.1186/1471-2334-11-166> PMID: [21658238](https://pubmed.ncbi.nlm.nih.gov/21658238/)
19. Qureshi EMA, Tabinda AB, Vehra S. Predicting dengue outbreak in the metropolitan city Lahore, Pakistan, using dengue vector indices and selected climatological variables as predictors. *J Pak Med Assoc*. 2017; 67: 416–421. PMID: [28303992](https://pubmed.ncbi.nlm.nih.gov/28303992/)
20. Betanzos-Reyes AF, Rodríguez MH, Romero-Martínez M, Sesma-Medrano E, Rangel-Flores H, Santos-Luna R. Association of dengue fever with Aedes spp. abundance and climatological effects. *Salud Publica Mex*. 2018; 60: 12–20. <https://doi.org/10.21149/8141> PMID: [29689652](https://pubmed.ncbi.nlm.nih.gov/29689652/)
21. Chang F, Tseng Y, Hsu P, Chen C, Lian I, Chao D. Re-assess Vector Indices Threshold as an Early Warning Tool for Predicting Dengue Epidemic in a Dengue Non-endemic Country. *PLoS Negl Trop Dis*. 2015; 9: e0004043. <https://doi.org/10.1371/journal.pntd.0004043> PMID: [26366874](https://pubmed.ncbi.nlm.nih.gov/26366874/)
22. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med*. 2013; 10: e1001381. <https://doi.org/10.1371/journal.pmed.1001381> PMID: [23393430](https://pubmed.ncbi.nlm.nih.gov/23393430/)
23. Moons KGM, Groot AH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Med*. 2014; 11: e1001744. <https://doi.org/10.1371/journal.pmed.1001744> PMID: [25314315](https://pubmed.ncbi.nlm.nih.gov/25314315/)
24. Moher D, Liberati A, Etzclaff J, Altman D. Preferred reporting items for systematic reviews and Meta-analyses: The PRISMA statement. *Ann Intern Med*. 2009; 151: 264–269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135> PMID: [19622511](https://pubmed.ncbi.nlm.nih.gov/19622511/)
25. Abualamah WA, Akbar NA, Banni HS, Bafail MA. Forecasting the morbidity and mortality of dengue fever in KSA: A time series analysis (2006–2016). *Journal of Taibah University Medical Sciences*. 2021; 16(3): 448–55.
26. 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 Negl Trop Dis*. 2016; 10: e0004681. <https://doi.org/10.1371/journal.pntd.0004681> PMID: [27128312](https://pubmed.ncbi.nlm.nih.gov/27128312/)
27. Anggraeni W, Nurmasari R, Riksakomara E, Samopa F, Wibowo RP, Pujiadi. Modified Regression Approach for Predicting Number of Dengue Fever Incidents in Malang Indonesia. *Procedia Comput Sci*. 2017; 124: 142–150.
28. Bal S, Sodoudi S. Modeling and prediction of dengue occurrences in Kolkata, India, based on climate factors. *Int J Biometeorol*. 2020; 64: 1379–1391. <https://doi.org/10.1007/s00484-020-01918-9> PMID: [32328786](https://pubmed.ncbi.nlm.nih.gov/32328786/)

29. Banu S, Guo Y, Hu W, Dale P, Mackenzie JS, Mengersen K, et al. Impacts of El Niño Southern Oscillation and Indian Ocean Dipole on dengue incidence in Bangladesh. *Sci Rep*. 2015; 5: 16105.
30. 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(4): 1–12.
31. Bett B, Grace D, Lee HS, Lindahl J, Nguyen-Viet H, Phuc PD, et al. Spatiotemporal analysis of historical records (2001–2012) on dengue fever in Vietnam and development of a statistical model for forecasting risk. *PLoS One*. 2019; 14: e0224353. <https://doi.org/10.1371/journal.pone.0224353> PMID: 31774823
32. Bouzid M, Colón-González FJ, Lung T, Lake IR, Hunter PR. Climate change and the emergence of vector-borne diseases in Europe: Case study of dengue fever. *BMC Public Health*. 2014; 14: 781. <https://doi.org/10.1186/1471-2458-14-781> PMID: 25149418
33. Buczak AL, Baugher B, Moniz LJ, Bagley T, Babin SM, Guven E. Ensemble method for dengue prediction. *PLoS One*. 2018; 13: e0189988. <https://doi.org/10.1371/journal.pone.0189988> PMID: 29298320
34. Carvajal T, Viacrusis K, Hernandez L, Ho H, Amalin D, Watanabe K. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. *BMC Infect Dis*. 2018; 18: 183. <https://doi.org/10.1186/s12879-018-3066-0> PMID: 29665781
35. Chen J, Ding RL, Liu KK, Xiao H, Hu G, Xiao X, et al. Collaboration between meteorology and public health: Predicting the dengue epidemic in Guangzhou, China, by meteorological parameters. *FRONTIERS IN CELLULAR AND INFECTION MICROBIOLOGY*. 2022; 12: 881745. <https://doi.org/10.3389/fcimb.2022.881745> PMID: 36017372
36. Chen Y, Liu T, Yu X, Zeng Q, Cai Z, Wu H, et al. An ensemble forecast system for tracking dynamics of dengue outbreaks and its validation in China. *PLoS Computational Biology*. 2022; 18(6): 1–22. <https://doi.org/10.1371/journal.pcbi.1010218> PMID: 35759513
37. Cheng YC, Lee FJ, Hsu YT, Slud EV, Hsiung CA, Chen CH, et al. Real-time dengue forecast for outbreak alerts in Southern Taiwan. *PLoS Neglected Tropical Diseases*. 2020; 14(7): 1–18. <https://doi.org/10.1371/journal.pntd.0008434> PMID: 32716983
38. Chuang T, Luis Fernando C, Chen P. Effects of local and regional climatic fluctuations on dengue outbreaks in southern Taiwan. *PLoS One*. 2017; 12: e0178698. <https://doi.org/10.1371/journal.pone.0178698> PMID: 28575035
39. Colón-González FJ, Fezzi C, Lake IR, Hunter PR. The Effects of Weather and Climate Change on Dengue. *PLoS Negl Trop Dis*. 2013; 7: e2503. <https://doi.org/10.1371/journal.pntd.0002503> PMID: 24244765
40. Depradine C, Lovell E. Climatological variables and the incidence of Dengue fever in Barbados. *Int J Environ Health Res*. 2004; 14: 429–441. <https://doi.org/10.1080/09603120400012868> PMID: 15545038
41. Descloux E, Mangeas M, Menkes CE, Lengaigne M, Leroy A, Tehei T, et al. Climate-Based Models for Understanding and Forecasting Dengue Epidemics. *PLoS Negl Trop Dis*. 2012; 6: e1470. <https://doi.org/10.1371/journal.pntd.0001470> PMID: 22348154
42. Dey SK, Rahman MM, Howlader A, Siddiqi UR, Uddin KMM, Borhan R, et al. Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach. *PLoS ONE*. 2022; 17(7): 1–17. <https://doi.org/10.1371/journal.pone.0270933> PMID: 35857776
43. Dharmawardhana 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); 2017 Dec 15–16; Peradeniya, Sri Lanka. Piscataway, New Jersey: Institution of Electrical and Electronics Engineers; 2017. p. 1–6. <https://doi.org/10.1109/ICIINFS.2017.8300381>
44. Earnest A, Tan SB, Wilder-Smith A. Meteorological factors and El Niño Southern Oscillation are independently associated with dengue infections. *Epidemiol Infect*. 2012; 140: 1244–1251.
45. Edussuriya C, Deegalla S, Gawarammana I. An accurate mathematical model predicting number of dengue cases in tropics. *PLoS Neglected Tropical Diseases*. 2021; 15(11): 1–15. <https://doi.org/10.1371/journal.pntd.0009756> PMID: 34748566
46. Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: A case study in China. *PLoS Negl Trop Dis*. 2017; 11: e0005973. <https://doi.org/10.1371/journal.pntd.0005973> PMID: 29036169
47. Jain R, Sontisirikit S, Iamsirithaworn S, Prendering H. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. *BMC Infect Dis*. 2019; 19: 272. <https://doi.org/10.1186/s12879-019-3874-x> PMID: 30898092

48. Jayaraj V, Avoi R, Gopalakrishnan N, Raja D, Umasa Y. Developing a dengue prediction model based on climate in Tawau, Malaysia. *Acta Trop*. 2019; 197: 105055. <https://doi.org/10.1016/j.actatropica.2019.105055> PMID: 31185224
49. Lauer SA, Sakrejda K, Ray EL, Keegan LT, Bi Q, Suangtho P, et al. Prospective forecasts of annual dengue hemorrhagic fever incidence in Thailand, 2010–2014. *Proc Natl Acad Sci USA*. 2018; 115: E2175–E2182. <https://doi.org/10.1073/pnas.1714457115> PMID: 29463757
50. Li C, Wang X, Wu X, Liu J, Ji D, Du J. Modeling and projection of dengue fever cases in Guangzhou based on variation of weather factors. *Sci Total Environ*. 2017; 605–606: 867–873. <https://doi.org/10.1016/j.scitotenv.2017.06.181> PMID: 28683431
51. Li Z, Yang L, Dong J, Gurgel H, Xu L. Improving Dengue Forecasts by Using Geospatial Big Data Analysis in Google Earth Engine and the Historical Dengue Information-Aided Long Short Term Memory Modeling. *Biology*. 2022; 11(2). <https://doi.org/10.3390/biology11020169> PMID: 35205036
52. Liu D, Guo S, Zou M, Chen C, Deng F, Xie Z, et al. A dengue fever predicting model based on Baidu search index data and climate data in South China. *PLoS One*. 2019; 14: e0226841. <https://doi.org/10.1371/journal.pone.0226841> PMID: 31887118
53. Lowe R, Bailey T, Stephenson D, Jupp TE, Graham RJ, Barcellos C, et al. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Stat Med*. 2013; 32: 864–883. <https://doi.org/10.1002/sim.5549> PMID: 22927252
54. Luz PM, Mendes BVM, Codeco CT, Struchiner CJ, Galvani AP. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *Am J Trop Med Hyg*. 2008; 79: 933–939. PMID: 19052308
55. McGough SF, Clemente L, Kutz JN, Santillana M. A dynamic, ensemble learning approach to forecast dengue fever epidemic years in Brazil using weather and population susceptibility cycles. *J R Soc Interface*. 2021; 18: 20201006. <https://doi.org/10.1098/rsif.2020.1006> PMID: 34129785
56. Mincham G, Baldock KL, Rozilawati H, Williams CR. Development of a mechanistic dengue simulation model for Guangzhou. *Epidemiol Infect*. 2019; 147: e125. <https://doi.org/10.1017/S095026881900030X> PMID: 30869038
57. Nakhapakorn K, Tripathi N. An information value-based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence. *Int J Health Geogr*. 2005; 4: 13. <https://doi.org/10.1186/1476-072X-4-13> PMID: 15943863
58. Nan J, Liao X, Chen J, Chen X, Chen J, Dong G, et al. Using Climate Factors to Predict the Outbreak of Dengue Fever. In: 2018 7th International Conference on Digital Home (ICDH); 2018 Nov 30–Dec 1; Guilin, China. International Conference on Digital Home (ICDH); 2018.p. 213–8. <https://doi.org/10.1109/ICDH.2018.00045>
59. Nguyen V, Tran TT, Mulhall J, Minh HV, Duong TQ, Chien NV, et al. Deep learning models for forecasting dengue fever based on climate data in Vietnam. *PLoS Neglected Tropical Diseases*. 2022; 16(6): 1–22. <https://doi.org/10.1371/journal.pntd.0010509> PMID: 35696432
60. Nuraini N, Fauzi IS, Fakhruddin M, Sopaheluwakan A, Soewono E. Climate-based dengue model in Semarang, Indonesia: Predictions and descriptive analysis. *Infect Dis Model*. 2021; 6: 598–611. <https://doi.org/10.1016/j.idm.2021.03.005> PMID: 33869907
61. Olmoguez I, Catindig M, Fel M, Fatima G. Developing a Dengue Forecasting Model: A Case Study in Iligan City. *Int J Adv Comput Sci Appl*. 2019; 10: 281–286.
62. 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); 2018 Oct 26–28; Subang Jaya, Malaysia. Piscataway, New Jersey: Institution of Electrical and Electronics Engineers; 2018. p. 1–6. <https://doi.org/10.1109/ICACCAF.2018.8776790>
63. Pham NTT, Nguyen CT, Pineda-Cortel MRB. Time-series modelling of dengue incidence in the Mekong Delta region of Viet Nam using remote sensing data. *Western Pacific Surveillance and Response*. 2020; 11(1): 13–21. <https://doi.org/10.5365/wpsar.2018.9.2.012> PMID: 32963887
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. <https://doi.org/10.1016/j.actatropica.2014.10.005> PMID: 25447266
65. Pineda-Cortel MRB, Clemente BM, Pham NTT. Modeling and predicting dengue fever cases in key regions of the Philippines using remote sensing data. *Asian Pacific Journal of Tropical Medicine*. 2019; 12(2): 60–6.
66. Pinto E, Coelho M, Oliver L, Massad E. The influence of climate variables on dengue in Singapore. *Int J Environ Health Res*. 2011; 21: 415–426. <https://doi.org/10.1080/09603123.2011.572279> PMID: 21557124

67. Puengpreeda A, Yhusumrarn S, Sirikulvadhana S. Weekly Forecasting Model for Dengue Hemorrhagic Fever Outbreak in Thailand. *Eng J*. 2020; 24: 71–87.
68. Ramadona AL, Lazuardi L, Hii YL, Holmner Å, Kusnanto H, Rocklöv J. Prediction of Dengue Outbreaks Based on Disease Surveillance and Meteorological Data. *PLoS One*. 2016; 11: e0152688. <https://doi.org/10.1371/journal.pone.0152688> PMID: 27031524
69. Roster K, Connaughton C, Rodrigues FA. Predicting Dengue Fever in Brazilian Cities. *bioRxiv*:430949 [Preprint]. 2021 [cited 2022 June 27]. Available from: <https://www.biorxiv.org/content/10.1101/2021.02.17.430949v1>
70. Salim NAM, Wah YB, Reeves C, Smith M, Yaacob WFW, Mudin RN, et al. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Sci Rep*. 2021; 11: 939. <https://doi.org/10.1038/s41598-020-79193-2> PMID: 33441678
71. Shi Y, Liu X, Kok S, Rajarethinam J, Liang S, Yap G, et al. Three-Month Real-Time Dengue Forecast Models: An Early Warning System for Outbreak Alerts and Policy Decision Support in Singapore. *Environ Health Perspect*. 2016; 124: 1369–1375. <https://doi.org/10.1289/ehp.1509981> PMID: 26662617
72. Sิริyasatien P, Phumee A, Ongruk P, Jampachaisri K, Kesorn K. Analysis of significant factors for dengue fever incidence prediction. *BMC Bioinformatics*. 2016; 17: 166. <https://doi.org/10.1186/s12859-016-1034-5> PMID: 27083696
73. Withanage GP, Hapugoda MD, Gunawardena YINS, Viswakula SD. A forecasting model for dengue incidence in the District of Gampaha, Sri Lanka. *Parasit Vectors* 2018; 11: 262. <https://doi.org/10.1186/s13071-018-2828-2> PMID: 29690906
74. 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. <https://doi.org/10.3390/ijerph17020453> PMID: 31936708
75. Yuan HY, Wen TH, Kung YH, Tsou HH, Chen CH, Chen LW, et al. Prediction of annual dengue incidence by hydro-climatic extremes for southern Taiwan. *International Journal of Biometeorology*. 2019; 63(2):259–68. <https://doi.org/10.1007/s00484-018-01659-w> PMID: 30680621
76. Yuan H, Liang J, Lin P, Sucipto K, Tsegaye MM, Wen TH, et al. The effects of seasonal climate variability on dengue annual incidence in Hong Kong: A modelling study. *Sci Rep*. 2020; 10: 4297. <https://doi.org/10.1038/s41598-020-60309-7> PMID: 32152334
77. Zafra B. Predicting dengue in the Philippines using artificial neural network. *medRxiv*:20209718 [Preprint]. 2021 [cited 2022 July 3]. Available from: <https://www.medrxiv.org/content/10.1101/2020.10.08.20209718v3>
78. Zambrano LI, Sevilla C, Reyes-Garcia SZ, Sierra M, Kafati R, Rodriguez-Morales AJ, et al. Potential impacts of climate variability on Dengue Hemorrhagic Fever in Honduras, 2010. *Trop Biomed*. 2012; 29: 499–507. PMID: 23202593
79. 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 Negl Trop Dis*. 2016; 10: e0004473. <https://doi.org/10.1371/journal.pntd.0004473> PMID: 26894570
80. 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 Negl Trop Dis*. 2020; 14: e0008056. <https://doi.org/10.1371/journal.pntd.0008056> PMID: 32970674
81. Zhu B, Wang L, Wang H, Cao Z, Zha L, Li Z. Prediction model for dengue fever based on interactive effects between multiple meteorological factors in Guangdong, China (2008–2016). *PLoS One*. 2019; 14: e0225811. <https://doi.org/10.1371/journal.pone.0225811> PMID: 31815950
82. Drake JM. Fundamental limits to the precision of early warning systems for epidemics of infectious diseases. *PLoS Med*. 2005; 2: e144. <https://doi.org/10.1371/journal.pmed.0020144> PMID: 15916474
83. Scott TW, Morrison AC, Lorenz LH, Clark GG, Strickman D, Kittayapong P. Longitudinal studies of *Aedes aegypti* (Diptera: Culicidae) in Thailand and Puerto Rico: population dynamics. *J Med Entomol*. 2000; 14: 77–88. <https://doi.org/10.1603/0022-2585-37.1.77> PMID: 15218910
84. Johansson MA, Reich NG, Hota A, Brownstein JS, Santillana M. Evaluating the performance of infectious disease forecasts: A comparison of climate-driven and seasonal dengue forecasts for Mexico. *Sci Rep*. 2016; 6: 33707. <https://doi.org/10.1038/srep33707> PMID: 27665707
85. Riley P, Ben-Nun M, Turtle J, Bacon D, Riley S. SARIMA Forecasts of Dengue Incidence in Brazil, Mexico, Singapore, Sri Lanka, and Thailand: Model Performance and the Significance of Reporting Delays. *medRxiv*: 20141093 [Preprint]. 2021 [cited 2022 July 3]. Available from: <https://www.medrxiv.org/content/10.1101/2020.06.26.20141093v1>
86. Batista EDA, Araújo WC, Lira RV, Batista LIA. Predicting dengue cases through Machine Learning and Deep Learning: a systematic review. *Res Soc Dev*. 2021; 10: e33101119347.

87. Zinszer K, Verma AD, Charland K, Brewer TF, Brownstein JS, Sun Z, et al. A scoping review of malaria forecasting: past work and future directions. *BMJ Open*. 2012; 2: e001992. <https://doi.org/10.1136/bmjopen-2012-001992> PMID: 23180505
88. Bullock J, Luccioni A, Pham KH, Lam CSN, Luengo-Oroz M. Mapping the Landscape of Artificial Intelligence Applications against COVID-19. *J Artif Intell Res*. 2020; 69: 807–845.
89. Lessler J, Cummings DA. Mechanistic Models of Infectious Disease and Their Impact on Public Health. *Am J Epidemiol*. 2016 Mar 1; 183(5): 415–22. <https://doi.org/10.1093/aje/kww021> PMID: 26893297
90. Baker RE, Peña J, Jayaratnam J, Jérusalem A. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*. 2018; 14(5): 20170660. <https://doi.org/10.1098/rsbl.2017.0660> PMID: 29769297
91. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010; 21: 128–138. <https://doi.org/10.1097/EDE.0b013e3181c30fb2> PMID: 20010215
92. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019; 17: 230. <https://doi.org/10.1186/s12916-019-1466-7> PMID: 31842878
93. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014; 35: 1925–1931. <https://doi.org/10.1093/eurheartj/ehu207> PMID: 24898551
94. Johansson MA, Apfeldorf KM, Dobson S, Devita J, Buczak AL, Baugher B, et al. An open challenge to advance probabilistic forecasting for dengue epidemics. *Proceedings of the National Academy of Sciences of the United States of America*. 2019 Nov 26; 116(48): 24268–74. <https://doi.org/10.1073/pnas.1909865116> PMID: 31712420
95. Ivanescu AE, Li P, George B, Brown AW, Keith SW, Raju D, et al. The importance of prediction model validation and assessment in obesity and nutrition research. *Int J Obes (Lond)*. 2016 Jun; 40(6):887–94. <https://doi.org/10.1038/ijo.2015.214> PMID: 26449421
96. Steyerberg EW, Lingsma HF. Validating prediction models. *BMJ*. 2008; 336: 789.
97. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *Journal of clinical epidemiology*. 2001 Aug 1; 54(8): 774–81. [https://doi.org/10.1016/s0895-4356\(01\)00341-9](https://doi.org/10.1016/s0895-4356(01)00341-9) PMID: 11470385
98. Efron B, Tibshirani R. Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*. 1997 Jun 1; 92(438): 548–60.
99. Cavanaugh JS. Bootstrap Cross-Validation Improves Model Selection in Pharmacometrics. *Stat Biopharm Res*. 2020; 10: 1–36.
100. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012. 98: 691–698. <https://doi.org/10.1136/heartjnl-2011-301247> PMID: 22397946