RESEARCH ARTICLE

Data-driven methods for dengue prediction and surveillance using real-world and Big Data: A systematic review

Emmanuelle Sylvestre[1,2]*, Clarisse Joachim[3,4], Elsa Cécilia-Joseph[2], Guillaume Bouzillé[1], Boris Campillo-Gimenez[1,5], Marc Cuggia[1], André Cabié[6,7,8]

1 Université de Rennes, CHU Rennes, INSERM, LTSI – UMR 1099, Rennes, France, **2** CHU Martinique, Centre de Données Cliniques, Martinique, France, **3** CHU Martinique, Pôle de Cancérologie Hématologie Urologie, Registre Général des Cancers de la Martinique, Martinique, France, **4** CHU Martinique, Pôle de Cancérologie Hématologie Urologie, Martinique Cancer Data Hub, Martinique, France, **5** Centre de Lutte Contre le Cancer Eugène Marquis, Rennes, France, **6** CHU Martinique, Infectious and Tropical Diseases Unit, Martinique, France, **7** CHU Martinique, INSERM, CIC-1424, Martinique, France, **8** PCCEI, Université de Montpellier, INSERM, EFS, Université Antilles, Montpellier, France

* emmanuelle.sylvestre@chu-martinique.fr

Abstract

Background

Traditionally, dengue surveillance is based on case reporting to a central health agency. However, the delay between a case and its notification can limit the system responsiveness. Machine learning methods have been developed to reduce the reporting delays and to predict outbreaks, based on non-traditional and non-clinical data sources. The aim of this systematic review was to identify studies that used real-world data, *Big Data* and/or machine learning methods to monitor and predict dengue-related outcomes.

Methodology/Principal findings

We performed a search in PubMed, Scopus, Web of Science and grey literature between January 1, 2000 and August 31, 2020. The review (ID: CRD42020172472) focused on data-driven studies. Reviews, randomized control trials and descriptive studies were not included. Among the 119 studies included, 67% were published between 2016 and 2020, and 39% used at least one novel data stream. The aim of the included studies was to predict a dengue-related outcome (55%), assess the validity of data sources for dengue surveillance (23%), or both (22%). Most studies (60%) used a machine learning approach. Studies on dengue prediction compared different prediction models, or identified significant predictors among several covariates in a model. The most significant predictors were rainfall (43%), temperature (41%), and humidity (25%). The two models with the highest performances were Neural Networks and Decision Trees (52%), followed by Support Vector Machine (17%). We cannot rule out a selection bias in our study because of our two main limitations: we did not include preprints and could not obtain the opinion of other international experts.

Conclusions/Significance

Combining real-world data and *Big Data* with machine learning methods is a promising approach to improve dengue prediction and monitoring. Future studies should focus on how to better integrate all available data sources and methods to improve the response and dengue management by stakeholders.

Author summary

Dengue is one of the most important arbovirus infections in the world and its public health, societal and economic burden is increasing. Although the majority of dengue cases are asymptomatic or mild, severe disease forms can lead to death. For this reason, early diagnosis and monitoring of dengue are crucial to decrease mortality. However, most endemic regions still rely on traditional monitoring methods, despite the growing availability of novel data sources and data-driven methods based on real-world data, *Big Data*, and machine learning algorithms. In this systematic review, we identified and analyzed studies that used these novel approaches for dengue monitoring and/or prediction. We found that novel data streams, such as Internet search engines and social media platforms, and machine learning methods can be successfully used to improve dengue management, but are still vastly ignored in real life. These approaches should be combined with traditional methods to help stakeholders better prepare for each outbreak and improve early responsiveness.

Introduction

Dengue virus (DENV) is an arbovirus transmitted to humans by *Aedes aegypti* or *Aedes albopictus* female mosquitoes [1]. The incidence of dengue, the disease caused by DENV, has rapidly increased around the world in recent decades [2] due to population growth, urbanization, increased travel, and insufficient vector control [3]. The World Health Organization (WHO), considers dengue a major global public health challenge in the tropical and subtropical regions [4]. Today, dengue is one of the most important vector-borne diseases in the world and recent studies on its prevalence estimate that 3.9 billion people are at risk of transmission, with 390 million infections and 96 million symptomatic cases per year [1,5]. Although most infections are asymptomatic or are characterized by intense flu-like symptoms that last up to 10 days [6], severe forms of dengue hemorrhagic fever/dengue shock syndrome can also occur [7] and might lead to death. Mortality due to dengue can be greatly reduced by early diagnosis, appropriate clinical management [3,7].

Most dengue-endemic regions (mainly South-East Asia, the Americas, and the Pacific region) rely on traditional surveillance, based on hospital syndromic reporting and laboratory confirmation of a subset of cases to a central health agency [3,8]. The method is very accurate, but is hampered by its lack of responsiveness with substantial delays between a case and its notification [8], which can limit the health system ability/rapidity to put in place appropriate measures to avoid drastic consequences. Moreover, this traditional surveillance system is expensive, due to the time needed to aggregate and manually validate data [9]. These limitations have prompted researchers to investigate other solutions. Many studies have described alternative methods, such as mobile, digital and Internet-based systems, to efficiently crowdsource data from the community [3]. However, these approaches have not been translated yet

into the standard dengue management practice. Yet, they are relevant for all dimensions of dengue management, such as monitoring, clinical management, and dengue outbreak forecasting [3,8]. Over the years, scientists have developed statistical and machine learning models to reduce the reporting delays and monitor new cases in almost real-time, but also to accurately use non-traditional and non-clinical data sources (e.g. Internet search engines and social media platforms) to predict communicable disease outbreaks [10–13], including dengue. Many studies have proposed new strategies based on *Big Data* and machine learning models to improve dengue outbreak management. However, recent systematic reviews only examined the relevance and usefulness of Internet-based surveillance systems in emerging tropical disease management [8,14], and they did not focus specifically on dengue management. Furthermore, recent systematic reviews on dengue analyzed monitoring [15], vaccine efficacy [16], epidemiological trends [17,18], the overall disease burden [19–21] and clinical prognosis models [22], but they did not discuss these new methods to improve dengue management.

Therefore, the first aim of this systematic review was to identify and describe all real-world and *Big Data*-based methods used to monitor and predict/forecast dengue-related outcomes, regardless of the region and/or population. The second aim was to analyze several features of these studies, such as the data sources and their origin, the different outcome types (e.g. epidemiological and clinical outcomes), the chosen statistical methods, and their performance and variability based on the population and location.

Methods

This systematic review was performed following the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" (PRISMA) guidelines [23]. Four reviewers (ES, CJ, AC and MC) developed the systematic review protocol. The literature search was performed in September 2020. The study protocol was registered on the PROSPERO registry of systematic reviews (ID: CRD42020172472).

Eligibility criteria

The review focused on studies that used real-world data, *Big Data* and/or machine learning methods to monitor, predict and/or forecast dengue outbreaks or dengue-related outcomes (clinical or epidemiological). Studies from any country (also regions outside endemic regions) were included, without any language filter. Analyses could be performed on past or future data.

Inclusion criteria

- Dengue diagnosis based on the standard WHO definition [7] valid at the time of the study

- Studies on humans, regardless of age, sex, and disease severity

- Studies using real-world data (including *Big Data)* (i.e. data not collected in experimental conditions) [24] for surveillance and/or prediction of dengue outbreaks.

Exclusion criteria

- Studies without original data, such as reviews, editorials, guidelines and perspectives articles

- Randomized control trials, case series, and case reports

- Descriptive epidemiological studies without any modeling

- Studies on other arbovirus types (e.g. chikungunya, Zika virus disease)

- Studies exclusively on mosquitoes (without any human data) and *in vitro* studies

- Studies only on incidence using geographic information systems

Search methodology

Information sources and search strategy.   The literature search was carried out in MEDLINE (PubMed), Scopus and Web of Science between January 1, 2000 and August 31, 2020 to identify potentially eligible studies. MeSH terms and keywords were used to perform the queries. First, the MeSH term "Dengue" was combined with several other MeSH terms (e.g. Data mining, Big Data, Forecasting, Social media), using the Boolean operator AND. Then, a more specific combination of keywords was used for all databases: i) Dengue AND [Monitoring OR Surveillance] AND [Big Data OR Data mining OR Instagram OR Facebook OR Twitter OR Tweets OR Google OR Baidu OR Google Trends OR Social media OR Social network OR Internet], ii) Dengue AND [Prediction OR Forecasting OR Modeling OR Modelling] AND [Big Data OR Data mining OR Instagram OR Facebook OR Twitter OR Tweets OR Google OR Baidu OR Google Trends OR Social media OR Social network OR Internet], iii) Dengue AND [Big Data OR Data mining OR Instagram OR Facebook OR Twitter OR Tweets OR Google OR Baidu OR Google Trends OR Social media OR social network OR Internet].

Relevant articles were also searched in the grey literature, including French-language studies on HAL (Hyper Articles en Lignes) [25], which is an open archive where authors can deposit scholarly documents from all academic fields, *theses.fr* [26], which is the French open database for all ongoing and defended PhD theses in France, and the WHO Dengue Bulletin.

Finally, the references of the retained studies and of major dengue epidemiological review articles were screened to identify studies overlooked by the previous search strategies.

Selection process.   Two independent authors (ES and CJ) screened the title and abstract to select relevant studies for the review. They read the full text of all studies that seemed to meet the eligibility criteria, or if the abstract was not explicit enough to make a decision. In case of disagreement, a third reviewer helped to reach a consensus (AC).

Quality assessment, data collection, extraction, and analysis.   Two reviewers (ES and CJ) extracted data from the selected articles, including first and last authors, year of publication, study period, objectives, study population, methodology, model performance and evaluation, study site (S1 Text).

As reporting guidelines for machine learning models and real-world data studies are not available, each reviewer independently performed a quality assessment using quality assessment criteria described in previous review articles on these topics [27–29] (S1 Table). A narrative synthesis of all eligible studies was prepared using the following framework: i) data sources and outcomes, ii) statistical and machine learning methods, iii) evaluation metrics, and iv) study results.

All descriptive analyses from the extracted articles were performed using R version 3.6.3 [30].

Results

Among the 2064 studies identified, 119 articles were included in this systematic review (Fig 1) [31–148]. Although the search time window was from January 1, 2000, the first included studies were published in 2008, and 67% of the eligible articles were published between 2016 and 2020 (Fig 2). The study populations were predominantly from South-East Asia (37%) and South America (22%). Among the 119 papers included, 77 (65%) were articles, and 42 (35%) were conference papers. On the basis of the Web of Science "Research Area" and the Scopus "Subject Area" classification, the topic of the selected articles were aggregated into eight
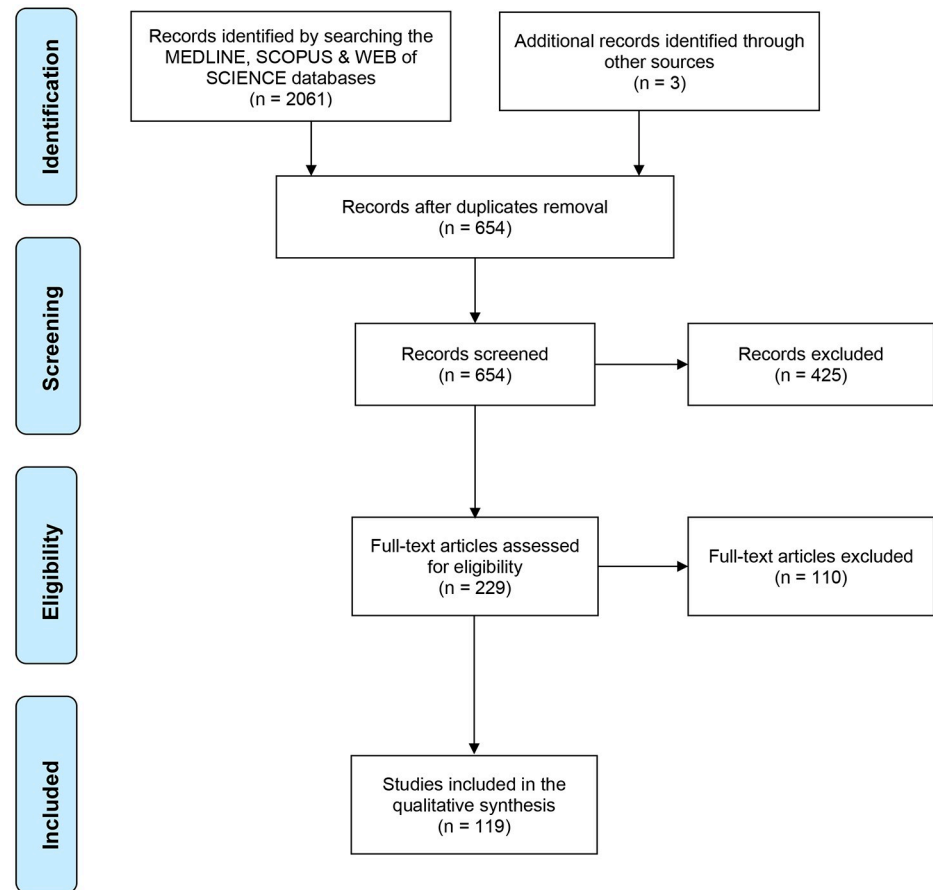
Fig 1. PRISMA Flow Diagram describing the screening process for the systematic review.

https://doi.org/10.1371/journal.pntd.0010056.g001

categories and three main themes: i) Information Technology & Science (52% of all articles), ii) Medicine (24%), and iii) Health Informatics, Public Health & Biology (24%) (Table 1). Conference papers were mainly classified in the "Information Technology & Science" category (39/42; 93%), whereas articles were more evenly distributed in the "Medicine" (28/77; 36%), "Health Informatics, Public Health & Biology" (26/77; 34%) and "Information Technology & Science" (23/77; 30%) themes (S2 Table). The complete list of all selected studies and their characteristics are in S3 Table.

Data sources

All included studies, except one [68], used only retrospective data. Most articles had multiple and heterogeneous data sources. The most conventional data sources were: government agencies (n = 72, 46%) and medical institutions (e.g. hospitals/laboratories) (n = 30, 19%). The data retrieved from these sources included epidemiological data, climate and environmental data from meteorological departments, and clinical and biological data. Some studies also used open access data from the WHO or from databases of published studies (S3 Table).

Among the included studies, 47/119 (39%) used at least one novel data stream, such as Internet search engines and social networks [14]. Most of these studies (n = 41, 87%) were published after 2015. Google was the most frequently used Internet search engine (n = 19
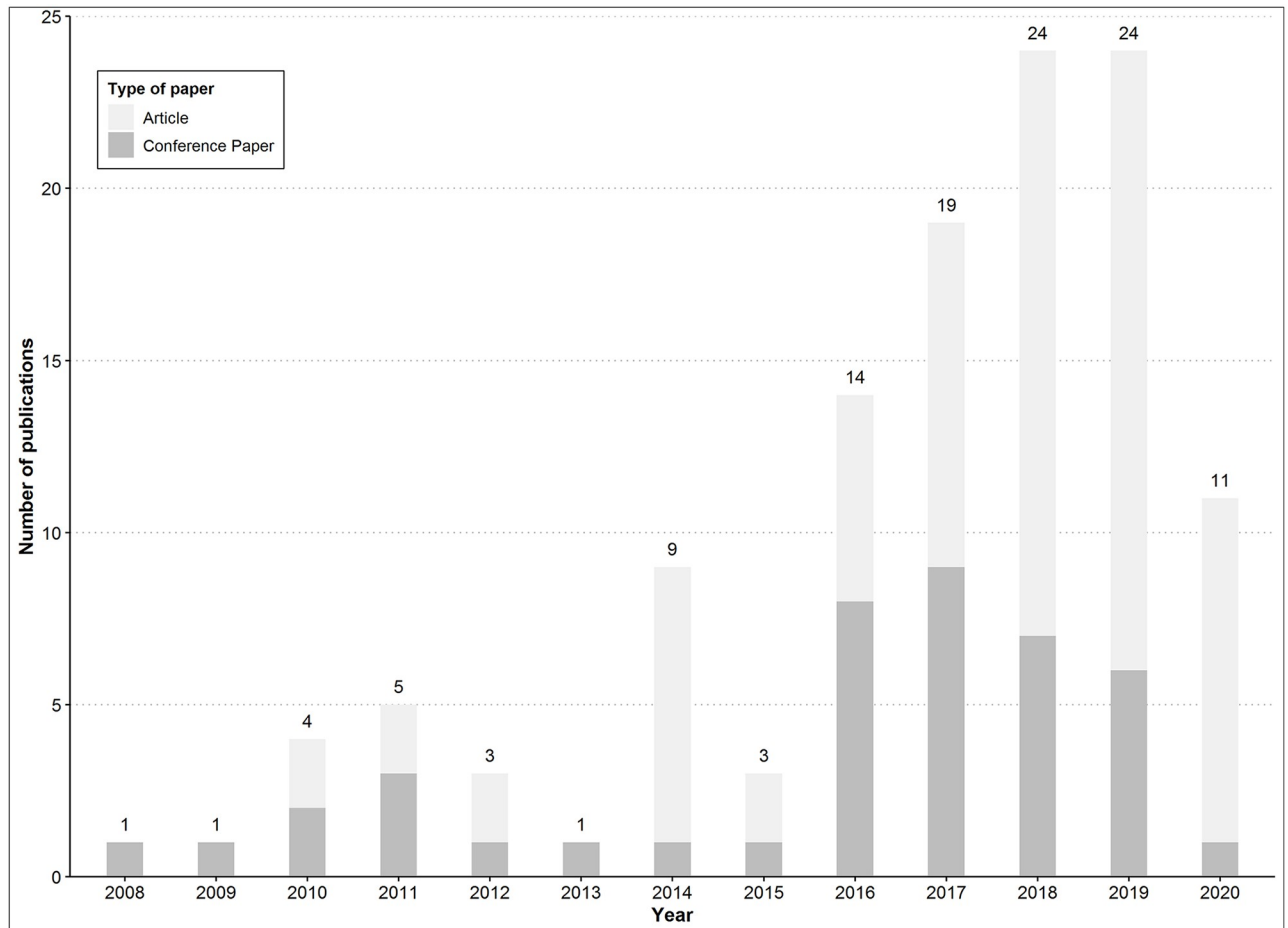
Fig 2. Number of publications on dengue prediction and/or surveillance published between January 1, 2000 and August 31, 2020.

studies) and Twitter the most frequently used social network (n = 18). Many studies based on novel data streams were research articles (n = 33, 70%), but the main theme, regardless of the study type (Conference paper or Article) varied depending on the data. Specifically, studies based on Google data were classified homogeneously into the three main themes. Conversely, studies that exploited social networks as data source were evenly distributed between Conference papers (n = 9) and Articles (n = 10), but only few of them were classified into the Medicine theme (Table 2).

Most studies used structured data, but 41 (34%) studies had an unstructured data source, such as Internet search-based queries or Twitter (Table 2). Among the 41 studies that used unstructured data, 28 (68%) did not develop their own pre-processing methods for these data sources, but simply used keywords related to their research. However, when studies used Natural-Language Processing (NLP)-based methods, they had a full pre-processing framework based on the NLP state-of-the-art recommendations.

Overall, studies that used non-conventional data relied less frequently on clinical data. Conversely, studies that used human data relied mostly on traditional sources, such as weather and environmental data. Moreover, genomic and vector data were vastly underused in combination with other sources, because only five studies using at least one of these sources were included in this systematic review. Data sources are detailed in Table 2.

Table 1. Type, study population and themes of the selected studies.

	n	%
Study type	**119**	
Article	77	65
Conference paper	42	35
Geographic region[*]		
Americas		
Caribbean	3	2
North America	3	2
South America	28	22
Asia		
East Asia	16	13
South-East Asia	47	37
South Asia	27	21
Australia	1	1
Worldwide	2	2
Study main theme		
Information Technology & Science	**62**	**52**
Computer Science	42	35
Engineering	10	8
Science & Technology—Other Topics	10	8
Medicine	**28**	**24**
Infectious Diseases & Tropical Medicine	20	17
Medicine—Other Topics	8	7
Health Informatics, Public Health & Biology	**29**	**24**
Biology	7	6
Medical Informatics	16	13
Public Health	6	5

[*]Some studies were carried out in more than one geographic regions

https://doi.org/10.1371/journal.pntd.0010056.t001

Statistical methods

The main aim of the included studies was to predict a dengue-related outcome (n = 65, 55%), to assess the validity of data sources for dengue surveillance (n = 29, 24%), or both (n = 25, 21%). The most frequently chosen outcomes (for prediction and monitoring) were dengue incidence rate (n = 58, 49%), dengue diagnosis based on symptoms (n = 20, 17%), and dengue outbreaks (n = 18, 15%) (S4 Table).

Only one study [48] used NLP-based methods for dengue prediction or surveillance, but as a pre-treatment step to extract and format data for modelling.

The model choice was related to the study objectives (prediction/forecasting or validity of a data source for dengue monitoring). Overall, most studies compared the performances of different models and statistical methods. The most frequently used models, regardless of the study aim(s), were regression-based models (25%), followed by decision-tree models (18%), and artificial neural networks (15%). Most studies on dengue monitoring used correlation analyses to identify relevant variables and/or data sources. Correlation methods (Pearson correlation or Spearman correlation) were especially useful to assess the validity of novel data streams, such as Twitter and Internet search engines. Most studies that included machine-learning algorithms used supervised learning methods (69%). The models' characteristics are detailed in Table 3.

Table 2. Data sources for dengue monitoring and prediction depending on the main theme.

Number of studies n(%)[a]		Study main theme n (%)		
		IT	Med	PH
	119	**62**	**28**	**29**
Traditional data sources				
Epidemiological and demographic data	86 (72)	42 (68)	24 (86)	20 (69)
Clinical and biological data	33 (27)	20 (32)	3 (11)	10 (34)
Genomic sequence data	2 (1)	1 (2)	0 (0)	1 (3)
Climate, environmental and geographic data	45 (37)	26 (42)	12 (43)	7 (24)
Vector data	4 (3)	1 (2)	3 (11)	0 (0)
Novel data streams				
Internet search engine data	25 (21)	8 (13)	11 (39)	6 (21)
Baidu	6 (5)	2 (3)	4 (14)	0 (0)
Google	19 (15)	6 (9)	7 (25)	6 (20)
Social media data	21 (17)	14 (22)	4 (14)	3 (10)
Twitter	18 (14)	12 (19)	4 (14)	2 (6)
Other	3 (2)	2 (3)	0 (0)	1 (3)
Other data sources	10	2 (3)	3 (11)	5 (17)
Cellphone	2	2 (3)	0 (0)	0 (0)
HealthMap	2	0 (0)	1 (3)	1 (3)
LeXisNexis	2	0 (0)	1 (3)	1 (3)
Political stability	1	0 (0)	0	1 (3)
Wikipedia	1	0 (0)	1 (3)	0 (0)

[a] As most studies used several data sources, some articles are present several times.

IT: Information Technology & Science; Med: Medicine; PH: Health Informatics, Public Health & Biology

To evaluate and assess the performance of the chosen statistical methods and/or models, 71 studies (60%) used a machine learning approach and partitioned their data into a training set and a test set. Like for the models, the choice of evaluation metrics was closely related to the study aim(s). All articles used at least one metric, and most of them more than one. Overall, the most common metrics were based on a Confusion Matrix (53%), with Accuracy as the most used metric, followed by Recall or Sensitivity. Correlation-based metrics were used in 37% of studies, especially correlation coefficients (Pearson or Spearman, depending on the data source). The aim of most studies that used correlation metrics was to assess a data source for dengue monitoring (n = 37, 84% of the 44 studies with correlation metric). Error-based metrics were also commonly used (n = 35, 29% of all studies). Few studies used other metrics (n = 22, 18% of all studies) and only 9 studies (8%) did not used at least one metrics falling into the above categories. (Table 4).

Study results

Among the 54 studies on surveillance, 37 (68%) assessed novel data streams, such as Internet search engines and social media, particularly Google (n = 16, 30%) and Twitter (n = 16, 30%). The most common traditional data source evaluated was climate, environmental and geographic data (n = 13/54; 24%) (S5 Table). All studies found a statistically significant association between the data source and the dengue-related outcome.

The aim of the studies on prediction (n = 90) could be categorized in two main groups: i) comparing different models to predict a dengue-related outcome, and ii) finding the

Table 3. Statistical methods and models used in the selected studies depending on the study aim[*].

Statistical methods	Prediction n (%)	Surveillance [a] n (%)	Prediction and surveillance n (%)	Totaln (%)
Methods for statistical analysis	**153**	**59**	**68**	**280**
Machine learning methods	126 (82)	27 (46)	51 (75)	**204 (73)**
Supervised learning	121 (79)	21 (36)	50 (74)	**192 (69)**
Unsupervised learning	5 (3)	6 (10)	1 (1)	**12 (4)**
Other model types (including time series models)	25 (16)	9 (15)	4 (6)	**38 (14)**
Correlation	2 (1)	23 (39)	13 (19)	**38 (14)**
Models for analyses	**151**	**36**	**55**	**242**
Artificial neural networks	31 (21)	2 (6)	3 (5)	**36 (15)**
Association rules	3 (2)	1 (3)	0 (0)	**4 (2)**
Bayesian models	12 (8)	5 (14)	3 (5)	**20 (8)**
Clustering	5 (3)	5 (14)	1 (2)	**11 (5)**
Decision tree	35 (23)	2 (6)	6 (11)	**43 (18)**
Regression model	20 (13)	9 (25)	31 (56)	**60 (25)**
Support-vector machine	17 (11)	3 (8)	7 (13)	**27 (11)**
Time series	12 (8)	1 (3)	3 (5)	**16 (7)**
Other [b]	16 (11)	8 (22)	1 (2)	**25 (10)**

[*]As most studies used several models and/or statistical methods, some are listed several times.

[a] Studies evaluating a data source (traditional or novel data streams) for dengue monitoring

[b] Some models classified as "Other" are also included in the "Supervised learning" category

https://doi.org/10.1371/journal.pntd.0010056.t003

significant predictors among several covariates in a model. Twenty-two studies (24%) included tried to respond to both aims.

The most significant predictors were rainfall (22 models, 43% of 51 studies), temperature (21 models, 41% of 51 studies), and humidity (13 models, 25% of 51 studies). These predictors were also the most frequent in studies to predict dengue incidence rates or dengue outbreaks. Conversely, in studies on dengue diagnosis prediction, the most frequent predictors were fever (4 models, 66% of 6 studies), arthralgia/myalgia (3 models, 50% of 6 studies), platelet count (2 models, 33% of 6 studies), and white blood cell count (2 models, 33% of 6 studies) (Table 5).

Overall, in studies comparing different models, neural networks and decision trees gave the best performances and were the best models in 13 studies (52% of 54 studies), followed by support vector machine (9/54 studies, 17%). In studies to predict dengue incidence rates, regression-based models showed the highest performance (5/24 studies, 21%) (Table 6). The full list of models and predictors, depending on the outcome, is provided in S5 Table.

Discussion

This systematic review showed that in the last 20 years, data-driven methods for dengue monitoring and prediction have become very popular, particularly in Asia where 72% of the included studies were performed. Very few studies were carried out outside Asia or the Americas, which is to be expected, because these are the two biggest dengue-endemic regions and 70% of the actual dengue burden is in Asia [149–151]. Studies in African countries were noticeably absent, although this continent also is a dengue-endemic region.

The most frequent data sources were conventional data traditionally used in dengue-related studies, such as case counts, climate, environmental, and clinical data. However, this review also highlighted the growing interest by the scientific community for novel Big Data streams for dengue surveillance and prediction [14,33,39–41,43,49,51–53,56,60,65,66,69–71,75–77,79–

Table 4. Evaluation metrics used in the selected articles depending on their aim(s)*.

Evaluation metrics	Prediction n (%)	Surveillance [a] n(%)	Prediction and surveillance n(%)	Total n(%)
Correlation metrics	**8**	**22**	**18**	**48**
Correlation coefficient	3 (38)	16 (73)	9 (50)	**28 (58)**
R-squared	4 (50)	5 (23)	9 (50)	**18 (38)**
Other correlation metric	1 (12)	1 (5)	0 (0)	**2 (4)**
Error-based metrics	**34**	**2**	**21**	**57**
Root mean square error	14 (41)	0 (0)	9 (43)	**23 (40)**
Mean absolute error	7 (21)	0 (0)	4 (19)	**11 (19)**
Mean absolute percentage error	4 (12)	0 (0)	3 (14)	**7 (12)**
Mean squared error	3 (9)	0 (0)	3 (14)	**6 (11)**
Other	6 (18)	2 (100)	2 (10)	**10 (18)**
Confusion matrix-based metrics	**147**	**13**	**17**	**177**
Accuracy	38 (26)	6 (46)	7 (41)	**51 (29)**
Recall/Sensitivity	32 (22)	2 (15)	3 (18)	**37 (21)**
Specificity	20 (14)	0 (0)	3 (18)	**23 (13)**
Precision/Positive predictive value	17 (12)	1 (8)	1 (6)	**19 (11)**
F-score	12 (8)	2 (15)	0 (0)	**14 (8)**
AUC and/or ROC curve[b]	16 (11)	1 (8)	3 (18)	**20 (11)**
Kappa statistic	5 (3)	0 (0)	0 (0)	**5 (3)**
Other	7 (5)	1 (8)	0 (0)	**8 (5)**
Other evaluation metrics	**10**	**6**	**11**	**27**
Number of articles using the evaluation metric	**65**	**29**	**25**	**119**
Correlation metrics	7 (11)	19 (66)	18 (72)	**44 (37)**
Error-based metrics	19 (29)	1 (3)	15 (60)	**35 (29)**
Confusion matrix-based metrics	47 (72)	9 (31)	8 (32)	**64 (54)**
Other evaluation metrics	8 (12)	6 (21)	8 (32)	**22 (18)**

* As studies used several metrics, some articles are listed more than once.

[a] Studies evaluating a data source (traditional data or novel data streams) for dengue monitoring

[b] AUC: Area Under the ROC Curve. ROC: Receiver Operating Characteristic

https://doi.org/10.1371/journal.pntd.0010056.t004

Table 5. Most significant predictors for the three most frequently studied outcomes.

Number of studies n (%)	Dengue incidence rates n = 27	Dengue outbreaks n = 9	Dengue diagnosis n = 6
Significant predictors*			
Rainfall	14 (52)	7 (78)	0 (0)
Temperature	14 (52)	6 (67)	0 (0)
Humidity	9 (33)	1 (11)	0 (0)
Mosquito-related predictor	0 (0)	2 (22)	0 (0)
Google search index	4 (15)	0 (0)	0 (0)
Baidu search index	3 (11)	0 (0)	0 (0)
Tweets	3 (11)	0 (0)	0 (0)
Fever	0 (0)	0 (0)	4 (66)
Arthralgia/myalgia	0 (0)	0 (0)	3 (50)
Platelet count	0 (0)	0 (0)	2 (33)
White blood cell count	0 (0)	0 (0)	2 (33)
Other	13 (48)	6 (67)	5 (83)

* Most studies found several significant predictors

https://doi.org/10.1371/journal.pntd.0010056.t005

Table 6. Model with the best performance for the three most frequently studied outcomes.

Number of studies	Dengue incidence rates n = 24	Dengue outbreaks n = 9	Dengue diagnosis n = 14
Best model			
Artificial neural network	4 (17)	1 (11)	4 (29)
Decision tree	4 (17)	2 (22)	4 (29)
Support vector machine	4 (17)	1 (11)	4 (29)
Regression model	5 (21)	1 (11)	0 (0)
Time series	3 (12)	2 (22)	0 (0)
Bayesian models	2 (8)	0 (0)	0 (0)
Association rules	1 (4)	1 (11)	0 (0)
Clustering	0 (0)	0 (0)	1 (7)
Other	1 (4)	1 (11)	1 (7)

https://doi.org/10.1371/journal.pntd.0010056.t006

81,84,85,91,92,98,100,102,105,110–112,114,115,126,127,130,135–138]. Indeed, social media and Internet search engines have become widely accessible worldwide, and therefore they represented the most popular novel data streams in the included studies. The easy access to these sources facilitates the assessment of their influence on infectious disease surveillance and prediction [152–154]. This is particularly true for neglected tropical diseases, such as dengue, Zika virus disease and chikungunya, because of their reoccurrence and the massive increase of their incidence in recent years [155,156]. Moreover, harnessing these novel data streams can improve traditional dengue surveillance systems, because they allow the early detection of an outbreak, and thus can decrease delays between the actual dengue outbreak onset and the official case notifications [157,158]. In the case of dengue control, early response is especially important because it can influence the outbreak severity.

Our analysis also identified the underutilization of some data sources. Genomic data and vector-based data were exploited only in 6 of the 119 included studies [35,42,50,57,75,131], despite the importance of vector surveillance in dengue. Moreover, studies using genomic data were based only on human genome data, although scientists could easily access viral genome sequencing data, for instance via the European Virus Archive—GLOBAL (EVAg) [159]. EVAg aim is to offer access to viruses and to virus sequencing data (including dengue) to scientists, government agencies and academic institutions. None of the included studies made use of data provided by this archive. The lack of vector data is surprising because this type of information is crucial in dengue monitoring studies[160,161]. However, we could not evaluate publication bias, especially in the case of underused data sources. As all included studies on the pertinence of a data source found a significant association between the source and a dengue-related outcome, we cannot exclude that some data sources were not underused, but rather not relevant for dengue management. However, the nature of the underused data sources could suggest that there is a dichotomy between data sources and the objectives of dengue studies: the studies focus either on techniques for vector monitoring/prediction or on techniques for human surveillance/prediction, but rarely on both. This dichotomy was also observed within human surveillance and prediction studies. Specifically, health scientists seemed to rely mainly on traditional data, whereas information technology researchers focused more on non-traditional data (especially social networks). Thus, studies using hospital data for dengue prediction rarely leveraged other data sources, such as climate data. Conversely, studies based on non-traditional data sources rarely used human data, besides the official number of dengue case counts. This might be explained by the fact that clinical data are often hard to access for researchers, particularly outside the medical community, for legal and ethical reasons. Furthermore, a substantial number of the selected papers were conference papers from

Information Technology & Sciences Conferences rather than Medicine Conferences. This might reflect the lack of interactions between research teams focused on prediction and/or informatics and physicians and/or government agencies focused on infectious disease monitoring and management. Yet, this research field would greatly benefit from combining their complementary approaches/expertise. Nevertheless, the most commonly studied outcomes in these articles based on real-world data were dengue incidence rate, dengue outbreaks and dengue diagnosis because they need to assess the reliability of novel data streams compared with traditional data sources. As most studies could demonstrate that these sources and methods can complete traditional surveillance and prediction methods, stakeholders should be more aware of these alternative methodologies and novel data streams, and reach out to these highly specialized teams to optimize outbreak dynamic tracking and to improve data completeness and prediction model accuracy.

Most of the included studies relied on machine learning methods, particularly supervised learning models, to assess traditional and also novel data streams. These models were useful also for the analysis of traditional data sources, and allowed scientists to harness non-structured data with NLP methods [40,43,48,49,51–53,56,60,65,66,69–71,73,76,77,79–81,84,85,92,98,100,102,105,110–112,114,115,126,127,130,134–139]. Unsupervised learning models were not the method of choice in most studies, possibly because these studies wanted to identify relevant data sources and/or indicators for dengue monitoring and prediction. Indeed, unsupervised learning tends to be used to identify clusters with similar characteristics [162,163]. Studies that used these methods wanted to predict dengue diagnosis based on the patient clinical profiles or to assess the validity of novel data sources, such as Twitter. Moreover, this approach for dengue research is fairly recent: with the exception of one conference paper from 2011, all studies using unsupervised learning models were published after 2016. Similarly, most studies relying on NLP methods were published rather recently, especially after 2017 (35 of the 42 studies with NLP methods). These two observations suggest that unsupervised learning and NLP might become more prominent in dengue research. It is important to note that despite the use of real-world data, these statistical methods were employed to analyze only retrospective data (but for one study), making their pertinence in real conditions difficult to assess.

Evaluation metrics are crucial in real-world data studies because they help to determine whether the collected data are fit for the purpose (here, dengue surveillance and prediction) and to assess data quality and bias [164]. Although most of the included prediction studies used at least one of the gold standard metrics for information retrieval, such as precision (or positive predictive value) and recall (or sensitivity) [165], several articles employed only error-based metrics, such as root mean square error and mean absolute error. The choice of evaluation metrics is obviously related to the study objective, but even studies where information retrieval metrics could be calculated did not necessarily use them. Again, these methodological choices might be explained by the discrepancy between health scientists who prefer "traditional" modeling evaluation metrics and information technology scientists who focus on information retrieval metrics.

This study also highlighted that despite the variety of approaches to predict dengue outcomes, some factors are constantly relevant, regardless of the study period or country, such as weather-based predictors, artificial neural networks, and decision tree models. However, a consensus on universal models and data sources has not been reached and will probably be difficult to attain due to the complex nature of dengue transmission.

This review has two main weaknesses despite the systematic approach. First, we only searched for published articles and did not look for preprints. Second, besides the experts involved in this review, we could not obtain the opinion of other international experts due to

the infectious disease context of 2020 (COVID-19 and dengue outbreaks in many regions). Therefore, we may have missed relevant studies for the review. Finally, the definition of real-world data can vary according to the stakeholders' view. We had to choose one single definition for the reviewing process, but other definitions do exist. Therefore, we cannot rule out a selection bias in our study.

Overall, this review showed that combining novel real-world and *Big Data* sources with machine learning methods is a promising approach to improve dengue prediction and outbreak monitoring. These new approaches are especially relevant because they can help government agencies and experts to better prepare for each resurgence and better manage outbreaks. Their aim is not to replace existing systems, but to complement them, especially for reducing delays between outbreaks and reporting. Future studies should focus on better integrating all available data sources and methods to improve the stakeholders' response and to better understand dengue outbreaks.

Supporting information

S1 Checklist. PRISMA Checklist.
(DOCX)

S2 Checklist. PRISMA for Abstracts Checklist.
(DOCX)

S1 Text. Data extraction sheet.
(DOCX)

S2 Text. PROSPERO Protocol.
(PDF)

S1 Table. Quality assessment criteria.
(DOCX)

S2 Table. Themes associated with the included studies.
(DOCX)

S3 Table. Characteristics of studies included in the systematic review.
(DOCX)

S4 Table. Studied outcomes in dengue fever surveillance and prediction.
(DOCX)

S5 Table. Detailed study outcomes and results.
(XLSX)

Author Contributions

Writing – review & editing: Emmanuelle Sylvestre, Clarisse Joachim, Elsa Cécilia-Joseph, Guillaume Bouzillé, Boris Campillo-Gimenez, Marc Cuggia, André Cabié.

References

1. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013; 496: 504–507. https://doi.org/10.1038/nature12060 PMID: 23563266

2. Waggoner JJ, Gresh L, Vargas MJ, Ballesteros G, Tellez Y, Soda KJ, et al. Viremia and Clinical Presentation in Nicaraguan Patients Infected With Zika Virus, Chikungunya Virus, and Dengue Virus. Clin Infect Dis Off Publ Infect Dis Soc Am. 2016; 63: 1584–1590. https://doi.org/10.1093/cid/ciw589 PMID: 27578819

3. Katzelnick LC, Coloma J, Harris E. Dengue: knowledge gaps, unmet needs, and research priorities. Lancet Infect Dis. 2017; 17: e88–e100. https://doi.org/10.1016/S1473-3099(16)30473-X PMID: 28185868

4. World Health Organization. Global strategy for dengue prevention and control, 2012–2020. Geneva, Switzerland: World Health Organization; 2012. http://apps.who.int/iris/bitstream/10665/75303/1/9789241504034_eng.pdf

5. Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl Trop Dis. 2012; 6: e1760. https://doi.org/10.1371/journal.pntd.0001760 PMID: 22880140

6. Chan M, Johansson MA. The incubation periods of Dengue viruses. PloS One. 2012; 7: e50972. https://doi.org/10.1371/journal.pone.0050972 PMID: 23226436

7. World Health Organization. Dengue Guidelines for Diagnosis, Treatment, Prevention and Control. Special Programme for Research and Training in Tropical Diseases, editor. Geneva: World Health Organization; 2009.

8. Milinovich GJ, Williams GM, Clements ACA, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis. 2014; 14: 160–168. https://doi.org/10.1016/S1473-3099(13)70244-5 PMID: 24290841

9. Madoff LC, Fisman DN, Kass-Hout T. A new approach to monitoring dengue activity. PLoS Negl Trop Dis. 2011; 5: e1215. https://doi.org/10.1371/journal.pntd.0001215 PMID: 21647309

10. Samaras L, Sicilia M-A, García-Barriocanal E. Predicting epidemics using search engine data: a comparative study on measles in the largest countries of Europe. BMC Public Health. 2021; 21: 100. https://doi.org/10.1186/s12889-020-10106-8 PMID: 33472589

11. Lu FS, Hattab MW, Clemente CL, Biggerstaff M, Santillana M. Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches. Nat Commun. 2019; 10: 147. https://doi.org/10.1038/s41467-018-08082-0 PMID: 30635558

12. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis Off Publ Infect Dis Soc Am. 2009; 49: 1557–1564. https://doi.org/10.1086/630200 PMID: 19845471

13. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. CMAJ Can Med Assoc J. 2009; 180: 829–831. https://doi.org/10.1503/cmaj.090215 PMID: 19364791

14. Gianfredi V, Bragazzi NL, Nucci D, Martini M, Rosselli R, Minelli L, et al. Harnessing Big Data for Communicable Tropical and Sub-Tropical Disorders: Implications From a Systematic Review of the Literature. Front Public Health. 2018; 6. https://doi.org/10.3389/fpubh.2018.00090 PMID: 29619364

15. Runge-Ranzinger S, McCall PJ, Kroeger A, Horstick O. Dengue disease surveillance: an updated systematic literature review. Trop Med Int Health TM IH. 2014; 19: 1116–1160. https://doi.org/10.1111/tmi.12333 PMID: 24889501

16. da Silveira LTC, Tura B, Santos M. Systematic review of dengue vaccine efficacy. BMC Infect Dis. 2019; 19: 750. https://doi.org/10.1186/s12879-019-4369-5 PMID: 31455279

17. Gutierrez-Barbosa H, Medina-Moreno S, Zapata JC, Chua JV. Dengue Infections in Colombia: Epidemiological Trends of a Hyperendemic Country. Trop Med Infect Dis. 2020; 5.

18. Ramos-Castañeda J, Barreto Dos Santos F, Martínez-Vega R, Galvão de Araujo JM, Joint G, Sarti E. Dengue in Latin America: Systematic Review of Molecular Epidemiological Trends. PLoS Negl Trop Dis. 2017; 11: e0005224. https://doi.org/10.1371/journal.pntd.0005224 PMID: 28068335

19. Ahmed AM, Mohammed AT, Vu TT, Khattab M, Doheim MF, Ashraf Mohamed A, et al. Prevalence and burden of dengue infection in Europe: A systematic review and meta-analysis. Rev Med Virol. 2020; 30: e2093. https://doi.org/10.1002/rmv.2093 PMID: 31833169

20.   Simo FBN, Bigna JJ, Kenmoe S, Ndangang MS, Temfack E, Moundipa PF, et al. Dengue virus infection in people residing in Africa: a systematic review and meta-analysis of prevalence studies. Sci Rep. 2019; 9: 13626. https://doi.org/10.1038/s41598-019-50135-x PMID: 31541167

21.   Cafferata ML, Bardach A, Rey-Ares L, Alcaraz A, Cormick G, Gibbons L, et al. Dengue Epidemiology and Burden of Disease in Latin America and the Caribbean: A Systematic Review of the Literature and Meta-Analysis. Value Health Reg Issues. 2013; 2: 347–356. https://doi.org/10.1016/j.vhri.2013.10.002 PMID: 29702769

22.   Dao Phuoc T, Khuong Quynh L, Vien Dang Khanh L, Ong Phuc T, Le Sy H, Le Ngoc T, et al. Clinical prognostic models for severe dengue: a systematic review protocol. Wellcome Open Res. 2019; 4: 12. https://doi.org/10.12688/wellcomeopenres.15033.2 PMID: 31448337

23.   Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009; 339: b2535–b2535. https://doi.org/10.1136/bmj.b2535 PMID: 19622551

24.   Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence—What Is It and What Can It Tell Us? N Engl J Med. 2016; 375: 2293–2297. https://doi.org/10.1056/NEJMsb1609216 PMID: 27959688

25.   Baruch P. Open Access Developments in France: the HAL Open Archives System. Learn Publ. 2007; 20: 267–282. https://doi.org/10.1087/095315107X239636

26.   Agence bibliographique de l'enseignement. Thèses. Agence bibliographique de l'enseignement supérieur (ABES); [cited 1 Apr 2021]. http://www.theses.fr

27.   Aswi A, Cramb SM, Moraga P, Mengersen K. Bayesian spatial and spatio-temporal approaches to modelling dengue fever: a systematic review. Epidemiol Infect. 2019; 147. https://doi.org/10.1017/S0950268818002807 PMID: 30369335

28.   Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015; 162: W1–W73. https://doi.org/10.7326/M14-0698 PMID: 25560730

29.   Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLOS ONE. 2020; 15: e0234722. https://doi.org/10.1371/journal.pone.0234722 PMID: 32530947

30.   R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/index.html

31.   Polwiang S. The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003–2017). BMC Infect Dis. 2020; 20: 208. https://doi.org/10.1186/s12879-020-4902-6 PMID: 32164548

32.   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. https://doi.org/10.3390/ijerph17020453 PMID: 31936708

33.   Rangarajan P, Mody SK, Marathe M. Forecasting dengue and influenza incidences using a sparse representation of Google trends, electronic health records, and time series data. PLoS Comput Biol. 2019; 15: e1007518. https://doi.org/10.1371/journal.pcbi.1007518 PMID: 31751346

34.   Anno S, Hara T, Kai H, Lee M-A, 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. https://doi.org/10.4081/gh.2019.771 PMID: 31724367

35.   Romero D, Olivero J, Real R, Guerrero JC. Applying fuzzy logic to assess the biogeographical risk of dengue in South America. Parasit Vectors. 2019; 12: 428. https://doi.org/10.1186/s13071-019-3691-5 PMID: 31488198

36.   Mello-Román JD, Mello-Román JC, Gómez-Guerrero S, García-Torres M. Predictive Models for the Medical Diagnosis of Dengue: A Case Study in Paraguay. Comput Math Methods Med. 2019; 2019: 7307803. https://doi.org/10.1155/2019/7307803 PMID: 31485259

37.   Stolerman LM, Maia PD, Kutz JN. Forecasting dengue fever in Brazil: An assessment of climate conditions. PloS One. 2019; 14: e0220106. https://doi.org/10.1371/journal.pone.0220106 PMID: 31393908

38.   Macedo Hair G, Fonseca Nobre F, Brasil P. Characterization of clinical patterns of dengue patients using an unsupervised machine learning approach. BMC Infect Dis. 2019; 19: 649. https://doi.org/10.1186/s12879-019-4282-y PMID: 31331271

39.   Husnayain A, Fuad A, Lazuardi L. Correlation between Google Trends on dengue fever and national surveillance report in Indonesia. Glob Health Action. 2019; 12: 1552652. https://doi.org/10.1080/16549716.2018.1552652 PMID: 31154985

40. Souza RCSNP, Assunção RM, Oliveira DM, Neill DB, Meira W. Where did I get dengue? Detecting spatial clusters of infection risk with social network data. Spat Spatio-Temporal Epidemiol. 2019; 29: 163–175. https://doi.org/10.1016/j.sste.2018.11.005 PMID: 31128626

41. Ramadona AL, Tozan Y, Lazuardi L, Rocklöv J. A combination of incidence data and mobility proxies from social media predicts the intra-urban spread of dengue in Yogyakarta, Indonesia. PLoS Negl Trop Dis. 2019; 13: e0007298. https://doi.org/10.1371/journal.pntd.0007298 PMID: 30986218

42. Davi C, Pastor A, Oliveira T, Neto FB de L, Braga-Neto U, Bigham AW, et al. Severe Dengue Prognosis Using Human Genome Data and Machine Learning. IEEE Trans Biomed Eng. 2019; 66: 2861–2868. https://doi.org/10.1109/TBME.2019.2897285 PMID: 30716030

43. 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 Total Environ. 2019; 647: 752–762. https://doi.org/10.1016/j.scitotenv.2018.08.044 PMID: 30092532

44. Koh Y-M, Spindler R, Sandgren M, Jiang J. A model comparison algorithm for increased forecast accuracy of dengue fever incidence in Singapore and the auxiliary role of total precipitation information. Int J Environ Health Res. 2018; 28: 535–552. https://doi.org/10.1080/09603123.2018.1496234 PMID: 30016117

45. Carvajal TM, Viacrusis KM, Hernandez LFT, Ho HT, Amalin DM, 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

46. 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. https://doi.org/10.1371/journal.pone.0195065 PMID: 29608586

47. Chen Y, Chu CW, Chen MIC, Cook AR. The utility of LASSO-based models for real time forecasts of endemic infectious diseases: A cross country comparison. J Biomed Inform. 2018; 81: 16–30. https://doi.org/10.1016/j.jbi.2018.02.014 PMID: 29496631

48. Villanes A, Griffiths E, Rappa M, Healey CG. Dengue Fever Surveillance in India Using Text Mining in Public Media. Am J Trop Med Hyg. 2018; 98: 181–191. https://doi.org/10.4269/ajtmh.17-0253 PMID: 29141718

49. 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

50. Chatterjee S, Dey N, Shi F, Ashour AS, Fong SJ, Sen S. Clinical application of modified bag-of-features coupled with hybrid neural-based classifier in dengue fever classification using gene expression data. Med Biol Eng Comput. 2018; 56: 709–720. https://doi.org/10.1007/s11517-017-1722-y PMID: 28891000

51. Guo P, Wang L, Zhang Y, Luo G, Zhang Y, Deng C, et al. Can internet search queries be used for dengue fever surveillance in China? Int J Infect Dis IJID Off Publ Int Soc Infect Dis. 2017; 63: 74–76. https://doi.org/10.1016/j.ijid.2017.08.001 PMID: 28797591

52. Yang S, Kou SC, Lu F, Brownstein JS, Brooke N, Santillana M. Advances in using Internet searches to track dengue. PLoS Comput Biol. 2017; 13: e1005607. https://doi.org/10.1371/journal.pcbi.1005607 PMID: 28727821

53. Marques-Toledo C de A, Degener CM, Vinhal L, Coelho G, Meira W, Codeço CT, et al. Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting Dengue at country and city level. PLoS Negl Trop Dis. 2017; 11: e0005729. https://doi.org/10.1371/journal.pntd.0005729 PMID: 28719659

54. Premaratne MK, Perera SSN, Malavige GN, Jayasinghe S. Mathematical Modelling of Immune Parameters in the Evolution of Severe Dengue. Comput Math Methods Med. 2017; 2017: 2187390. https://doi.org/10.1155/2017/2187390 PMID: 28293273

55. Jayasundara SDP, Perera SSN, Malavige GN, Jayasinghe S. Mathematical modelling and a systems science approach to describe the role of cytokines in the evolution of severe dengue. BMC Syst Biol. 2017; 11: 34. https://doi.org/10.1186/s12918-017-0415-3 PMID: 28284213

56. Li Z, Liu T, Zhu G, Lin H, Zhang Y, He J, et al. Dengue Baidu Search Index data can improve the prediction of local dengue epidemic: A case study in Guangzhou, China. PLoS Negl Trop Dis. 2017; 11: e0005354. https://doi.org/10.1371/journal.pntd.0005354 PMID: 28263988

57. Kesorn K, Ongruk P, Chompoosri J, Phumee A, Thavara U, Tawatsin A, et al. Morbidity Rate Prediction of Dengue Hemorrhagic Fever (DHF) Using the Support Vector Machine and the Aedes aegypti Infection Rate in Similar Climates and Geographical Areas. PloS One. 2015; 10: e0125049. https://doi.org/10.1371/journal.pone.0125049 PMID: 25961289

58.   Dayama P, Sampath K. Dengue disease outbreak detection. Stud Health Technol Inform. 2014; 205: 1105–1109. PMID: 25160360

59.   Sampath K, Dayama P. Predicting the operations alert levels for dengue surveillance and control. Stud Health Technol Inform. 2014; 205: 1100–1104. PMID: 25160359

60.   Gluskin RT, Johansson MA, Santillana M, Brownstein JS. Evaluation of Internet-based dengue query data: Google Dengue Trends. PLoS Negl Trop Dis. 2014; 8: e2713. https://doi.org/10.1371/journal.pntd.0002713 PMID: 24587465

61.   Flamand C, Fritzell C, Prince C, Abboud P, Ardillon V, Carvalho L, et al. Epidemiological assessment of the severity of dengue epidemics in French Guiana. PloS One. 2017; 12: e0172267. https://doi.org/10.1371/journal.pone.0172267 PMID: 28196111

62.   Torres C, Barguil S, Melgarejo M, Olarte A. Fuzzy model identification of dengue epidemic in Colombia based on multiresolution analysis. Artif Intell Med. 2014; 60: 41–51. https://doi.org/10.1016/j.artmed.2013.11.008 PMID: 24388398

63.   Buczak AL, Koshute PT, Babin SM, Feighner BH, Lewis SH. A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data. BMC Med Inform Decis Mak. 2012; 12: 124. https://doi.org/10.1186/1472-6947-12-124 PMID: 23126401

64.   Hoen AG, Keller M, Verma AD, Buckeridge DL, Brownstein JS. Electronic event-based surveillance for monitoring dengue, Latin America. Emerg Infect Dis. 2012; 18: 1147–1150. https://doi.org/10.3201/eid1807.120055 PMID: 22709430

65.   Althouse BM, Ng YY, Cummings DAT. Prediction of dengue incidence using search query surveillance. PLoS Negl Trop Dis. 2011; 5: e1258. https://doi.org/10.1371/journal.pntd.0001258 PMID: 21829744

66.   Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Negl Trop Dis. 2011; 5: e1206. https://doi.org/10.1371/journal.pntd.0001206 PMID: 21647308

67.   Faisal T, Taib MN, Ibrahim F. Neural network diagnostic system for dengue patients risk classification. J Med Syst. 2012; 36: 661–676. https://doi.org/10.1007/s10916-010-9532-x PMID: 20703665

68.   Ibrahim F, Faisal T, Salim MIM, Taib MN. Non-invasive diagnosis of risk in dengue patients using bio-electrical impedance analysis and artificial neural network. Med Biol Eng Comput. 2010; 48: 1141–1148. https://doi.org/10.1007/s11517-010-0669-z PMID: 20683676

69.   Syamsuddin M, Fakhruddin M, Sahetapy-Engel JTM, Soewono E. Causality Analysis of Google Trends and Dengue Incidence in Bandung, Indonesia With Linkage of Digital Data Modeling: Longitudinal Observational Study. J Med Internet Res. 2020; 22: e17633. https://doi.org/10.2196/17633 PMID: 32706682

70.   Romero-Alvarez D, Parikh N, Osthus D, Martinez K, Generous N, Del Valle S, et al. Google Health Trends performance reflecting dengue incidence for the Brazilian states. BMC Infect Dis. 2020; 20: 252. https://doi.org/10.1186/s12879-020-04957-0 PMID: 32228508

71.   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

72.   Musa SS, Zhao S, Chan H-S, Jin Z, He DH. A mathematical model to study the 2014–2015 large-scale dengue epidemics in Kaohsiung and Tainan cities in Taiwan, China. Math Biosci Eng MBE. 2019; 16: 3841–3863. https://doi.org/10.3934/mbe.2019190 PMID: 31499639

73.   Messina JP, Brady OJ, Golding N, Kraemer MUG, Wint GRW, Ray SE, et al. The current and future global distribution and population at risk of dengue. Nat Microbiol. 2019; 4: 1508–1515. https://doi.org/10.1038/s41564-019-0476-8 PMID: 31182801

74.   Titus Muurlink O, Stephenson P, Islam MZ, Taylor-Robinson AW. Long-term predictors of dengue outbreaks in Bangladesh: A data mining approach. Infect Dis Model. 2018; 3: 322–330. https://doi.org/10.1016/j.idm.2018.11.004 PMID: 30839927

75.   Marques-Toledo CA, Bendati MM, Codeço CT, Teixeira MM. Probability of dengue transmission and propagation in a non-endemic temperate area: conceptual model and decision risk levels for early alert, prevention and control. Parasit Vectors. 2019; 12: 38. https://doi.org/10.1186/s13071-018-3280-z PMID: 30651125

76.   Verma M, Kishore K, Kumar M, Sondh AR, Aggarwal G, Kathirvel S. Google Search Trends Predicting Disease Outbreaks: An Analysis from India. Healthc Inform Res. 2018; 24: 300–308. https://doi.org/10.4258/hir.2018.24.4.300 PMID: 30443418

77.   Ho HT, Carvajal TM, Bautista JR, Capistrano JDR, Viacrusis KM, Hernandez LFT, et al. Using Google Trends to Examine the Spatio-Temporal Incidence and Behavioral Patterns of Dengue Disease: A

Case Study in Metropolitan Manila, Philippines. Trop Med Infect Dis. 2018; 3. https://doi.org/10.3390/tropicalmed3040118 PMID: 30423898

78. Phakhounthong K, Chaovalit P, Jittamala P, Blacksell SD, Carter MJ, Turner P, et al. Predicting the severity of dengue fever in children on admission based on clinical features and laboratory indicators: application of classification tree analysis. BMC Pediatr. 2018; 18: 109. https://doi.org/10.1186/s12887-018-1078-y PMID: 29534694

79. Strauss RA, Castro JS, Reintjes R, Torres JR. Google dengue trends: An indicator of epidemic behavior. The Venezuelan Case. Int J Med Inf. 2017; 104: 26–30. https://doi.org/10.1016/j.ijmedinf.2017.05.003 PMID: 28599813

80. Nsoesie EO, Flor L, Hawkins J, Maharana A, Skotnes T, Marinho F, et al. Social Media as a Sentinel for Disease Surveillance: What Does Sociodemographic Status Have to Do with It? PLoS Curr. 2016; 8. https://doi.org/10.1371/currents.outbreaks.cc09a42586e16dc7dd62813b7ee5d6b6 PMID: 28123858

81. Liu K, Wang T, Yang Z, Huang X, Milinovich GJ, Lu Y, et al. Using Baidu Search Index to Predict Dengue Outbreak in China. Sci Rep. 2016; 6: 38040. https://doi.org/10.1038/srep38040 PMID: 27905501

82. Ximenes R, Amaku M, Lopez LF, Coutinho FAB, Burattini MN, Greenhalgh D, et al. The risk of dengue for non-immune foreign visitors to the 2016 summer olympic games in Rio de Janeiro, Brazil. BMC Infect Dis. 2016; 16: 186. https://doi.org/10.1186/s12879-016-1517-z PMID: 27129407

83. Mohamad Mohsin MF, Abu Bakar A, Hamdan AR. Outbreak detection model based on danger theory. Appl Soft Comput. 2014; 24: 612–622. https://doi.org/10.1016/j.asoc.2014.08.030 PMID: 32362801

84. Puengpreeda A, Yhusumrarn S, Sirikulvadhana S. Weekly Forecasting Model for Dengue Hemorrhagic Fever Outbreak in Thailand. Eng J-Thail. 2020; 24: 71–87. https://doi.org/10.4186/ej.2020.24.3.71

85. Amin S, Uddin MI, Hassan S, Khan A, Nasser N, Alharbi A, et al. Recurrent Neural Networks With TF-IDF Embedding Technique for Detection and Classification in Tweets of Dengue Disease. Ieee Access. 2020; 8: 131522–131533. https://doi.org/10.1109/ACCESS.2020.3009058

86. Manogaran G, Lopez D, Chilamkurti N. In-Mapper combiner based MapReduce algorithm for processing of big climate data. Future Gener Comput Syst- Int J Escience. 2018; 86: 433–445. https://doi.org/10.1016/j.future.2018.02.048

87. Agarwal N, Koti SR, Saran S, Kumar AS. Data mining techniques for predicting dengue outbreak in geospatial domain using weather parameters for New Delhi, India. Curr Sci. 2018; 114: 2281–2291. https://doi.org/10.18520/cs/v114/i11/2281-2291

88. Manogaran G, Lopez D. A Gaussian process based big data processing framework in cluster computing environment. Clust Comput- J Netw Softw Tools Appl. 2018; 21: 189–204. https://doi.org/10.1007/s10586-017-0982-5

89. Jahangir I, Abdul-Basit, Hannan A, Javed S. Prediction of Dengue Disease Through Data Mining by Using Modified Apriori Algorithm. Proceedings of the 4th Acm International Conference of Computing for Engineering and Sciences (icces'2018). New York: Assoc Computing Machinery; 2018.

90. Husin NA, Alharogi A, Mustapha N, Hamdan H, Husin UA. Early Self-Diagnosis of Dengue Symptoms Using Fuzzy and Data Mining Approach. In: Nifa F a. A, Lin CK, Hussain A, editors. Proceedings of the 3rd International Conference on Applied Science and Technology (icast'18). Melville: Amer Inst Physics; 2018. p. 020048.

91. Anggraeni W, Pramudita G, Riksakomara E, Radityo PW, Samopa F, Pujiadi, et al. Artificial Neural Network for Health Data Forecasting, Case Study: Number of Dengue Hemorrhagic Fever Cases in Malang Regency, Indonesia. 2018 International Conference on Electrical Engineering and Computer Science (icecos). New York: Ieee; 2018. pp. 207–212.

92. Dennison Livelo E, Cheng C. Intelligent Dengue Infoveillance Using Gated Recurrent Neural Learning and Cross-Label Frequencies. 2018 Ieee International Conference on Agents (ica). New York: Ieee; 2018. pp. 2–7.

93. Wiratmadja II, Salamah SY, Govindaraju R. Healthcare Data Mining: Predicting Hospital Length of Stay of Dengue Patients. J Eng Technol Sci. 2018; 50: 110–126. https://doi.org/10.5614/j.eng.technol.sci.2018.50.1.8

94. Arafiyah R, Hermin F. Data mining for dengue hemorrhagic fever (DHF) prediction with naive Bayes method. 1st International Conference of Education on Sciences, Technology, Engineering, and Mathematics (ice-Stem). Bristol: Iop Publishing Ltd; 2018. p. 012077.

95. Abuhamad HIS, Abu Bakar A, Zainudin S, Sahani M, Ali ZM. Feature Selection Algorithms for Malaysian Dengue Outbreak Detection Model. Sains Malays. 2017; 46: 255–265. https://doi.org/10.17576/jsm-2017-4602-10

96. Manivannan P, Devi PI. Dengue Fever Prediction Using K-Means Clustering Algorithm. 2017 Ieee International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (incos). New York: Ieee; 2017.

97. 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. 2017 Ieee International Conference on Industrial and Information Systems (iciis). New York: Ieee; 2017. pp. 278–283.

98. Espina K, Estuar MRJE. Infodemiology for Syndromic Surveillance of Dengue and Typhoid Fever in the Philippines. In: CruzCunha MM, Varajao JEQ, Rijo R, Martinho R, Peppard J, SanCristobal JR, et al., editors. Centeris 2017—International Conference on Enterprise Information Systems / Projman 2017—International Conference on Project Management / Hcist 2017—International Conference on Health and Social Care Information Systems and Technologies, Centeri. Amsterdam: Elsevier Science Bv; 2017. pp. 554–561.

99. Rahim NF, Taib SM, Abidin AIZ. Dengue Fatality Prediction Using Data Mining. J Fundam Appl Sci. 2017; 9: 671–683. https://doi.org/10.4314/jfas.v9i6s.52

100. Klein GH, Neto PG, Tezza R. Big Data and social media: surveillance of networks as management tool. Saude E Soc. 2017; 26: 208–217. https://doi.org/10.1590/S0104-12902017164943

101. Kerdprasop N, Kerdprasop K. Remote Sensing Based Modeling of Dengue Outbreak with Regression and Binning Classification. 2016 2nd Ieee International Conference on Computer and Communications (iccc). New York: Ieee; 2016. pp. 46–49.

102. Anggraeni W, Aristiani L. Using Google Trend Data in Forecasting Number of Dengue Fever Cases with ARIMAX Method Case Study: Surabaya, Indonesia. Proceedings of 2016 International Conference on Information & Communication Technology and Systems (icts). New York: Ieee; 2016. pp. 114–118.

103. Mathulamuthu SS, Asirvadam VS, Dass SC, Gill BS, Loshini T. Predicting Dengue Incidences Using Cluster Based Regression on Climate Data. 2016 6th Ieee International Conference on Control System, Computing and Engineering (iccsce). New York: Ieee; 2016. pp. 245–250.

104. Rahmawati D, Huang Y-P. Using C-support Vector Classification to Forecast Dengue Fever Epidemics in Taiwan. In: Wang WJ, Lee PJ, Er MJ, Jeng JT, editors. 2016 International Conference on System Science and Engineering (icsse). New York: Ieee; 2016.

105. Missier P, Romanovsky A, Miu T, Pal A, Daniilakis M, Garcia A, et al. Tracking Dengue Epidemics Using Twitter Content Classification and Topic Modelling. In: Casteleyn S, Dolog P, Pautasso C, editors. Current Trends in Web Engineering, Icwe 2016 International Workshops. Cham: Springer International Publishing Ag; 2016. pp. 80–92.

106. Abeyrathna MP a. R, Abeygunawrdane DA, Wijesundara R a. a. V, Mudalige VB, Bandara M, Perera S, et al. Dengue Propagation Prediction using Human Mobility. 2nd International Mercon 2016 Moratuwa Engineering Research Conference. New York: Ieee; 2016. pp. 156–161.

107. Fathima AS, Manimeglai D. Analysis of Significant Factors for Dengue Infection Prognosis Using the Random Forest Classifier. Int J Adv Comput Sci Appl. 2015; 6: 240–245.

108. Tazkia RAK, Narita V, Nugroho AS. Dengue Outbreak Prediction for GIS based Early Warning System. 2015 International Conference on Science in Information Technology (ICSITech). New York: Ieee; 2015. pp. 121–125.

109. Wu Y, Lee G, Fu X, Hung T. Detect climatic factors contributing to dengue outbreak based on wavelet, support vector machines and genetic algorithm. In: Ao SI, Gelman L, Hukins DWL, Hunter A, Korsunsky AM, editors. World Congress on Engineering 2008, Vols I-Ii. Hong Kong: Int Assoc Engineers-Iaeng; 2008. pp. 303-+.

110. Salam N, Deeba F, Qadir F, Al-Hijli F, Al-Otaibi YN. Analysis of Correlation between Google Search Trends and Dengue Outbreaks from India. J Clin Diagn Res. 2019; 13: LC13–LC15. https://doi.org/10.7860/JCDR/2019/42611.13304

111. Chire Saire JE. Building Intelligent Indicators to Detect Dengue Epidemics in Brazil using Social Networks. OrjuelaCanon AD, editor. 2019 Ieee Colombian Conference on Applications in Computational Intelligence (colcaci). New York: Ieee; 2019.

112. Swain S, Seeja KR. Analysis of Epidemic Outbreak in Delhi Using Social Media Data. In: Kaushik S, Gupta D, Kharb L, Chahal D, editors. Information, Communication and Computing Technology. Singapore: Springer-Verlag Singapore Pte Ltd; 2017. pp. 25–34.

113. Saravanan N, Gayathri V. Classification of Dengue Dataset Using J48 Algorithm and Ant Colony Based Aj48 Algorithm. Proceedings of the International Conference on Inventive Computing and Informatics (icici 2017). New York: Ieee; 2017. pp. 1062–1067.

114. Carlos MA, Nogueira M, Machado RJ. Analysis of Dengue Outbreaks Using Big Data Analytics and Social Networks. 2017 4th International Conference on Systems and Informatics (icsai). New York: Ieee; 2017. pp. 1592–1597.

115. Ye X, Li S, Yang X, Qin C. Use of Social Media for the Detection and Analysis of Infectious Diseases in China. Isprs Int J Geo-Inf. 2016; 5: 156. https://doi.org/10.3390/ijgi5090156

116. Li W, Chen Y. Risk Factor Identification and Spatiotemporal Diffusion Path During the Dengue Outbreak. In: Weng Q, Gamba P, Xian G, Chen JM, Liang S, editors. 2016 4rth International Workshop on Earth Observation and Remote Sensing Applications (EORSA). New York: Ieee; 2016.

117. Srilekha G, Anupama B. Prediction of Dengue Outbreaks with Big Data using Machine Learning. GEDRAG Organ Rev. 2020; 33. https://doi.org/10.37896/GOR33.02/073

118. Ganthimathi M, Thangamani M, Mallika C, Prasanna Balaji V. Prediction of dengue fever using intelligent classifier. Int J Emerg Trends Eng Res. 2020; 8: 1338–1341. https://doi.org/10.30534/ijeter/2020/65842020

119. Kumar NK, Sikamani KT. Prediction of chronic and infectious diseases using machine learning classifiers-A systematic approach. Int J Intell Eng Syst. 2020; 13: 11–20. https://doi.org/10.22266/IJIES2020.0831.02

120. Guiyab RB. Development of prediction models for the dengue survivability prediction: An integration of data mining and decision support system. Int J Innov Technol Explor Eng. 2019; 8: 2199–2205. https://doi.org/10.35940/ijitee.J9411.0881019

121. Chovatiya M, Dhameliya A, Deokar J, Gonsalves J, Mathur A. Prediction of dengue using recurrent neural network. 2019. pp. 926–929. https://doi.org/10.1109/icoei.2019.8862581

122. Kerdprasop K, Kerdprasop N, Chansilp K, Chuaybamroong P. The Use of Spaceborne and Oceanic Sensors to Model Dengue Incidence in the Outbreak Surveillance System. Lect Notes Comput Sci Subser Lect Notes Artif Intell Lect Notes Bioinforma. 2019;11619 LNCS: 447–460.

123. Link H, Richter SN, Leung VJ, Brost RC, Phillips CA, Staid A. Statistical models of dengue fever. Commun Comput Inf Sci. 2019; 996: 175–186. https://doi.org/10.1007/978-981-13-6661-1_14

124. Arafiyah R, Hermin F, Kartika IR, Alimuddin A, Saraswati I. Classification of Dengue Haemorrhagic Fever (DHF) using SVM, naive bayes and random forest. 2018.

125. Mishra S, Tripathy HK, Panda AR. An improved and adaptive attribute selection technique to optimize Dengue fever prediction. Int J Eng Technol. 2018; 7: 480–486. https://doi.org/10.14419/ijet.v7i2.29.13802

126. Wu C-H, Kao S-C, Kan M-H. Knowledge discovery in open data of dengue epidemic. 2017.

127. Albinati J, Meira W Jr, Pappa GL, Teixeira M, Marques-Toledo C. Enhancement of epidemiological models for dengue fever based on twiter data. 2017. pp. 109–118.

128. Zhu G, Hunter J, Jiang Y. Improved Prediction of Dengue Outbreak Using the Delay Permutation Entropy. 2017. pp. 828–832.

129. Zainudin Z, Shamsuddin SM. Predictive analytics in Malaysian dengue data from 2010 until 2015 using BigML. Int J Adv Soft Comput Its Appl. 2016; 8: 18–30.

130. Milinovich GJ, Avril SMR, Clements ACA, Brownstein JS, Tong S, Hu W. Using internet search queries for infectious disease surveillance: Screening diseases for suitability. BMC Infect Dis. 2014; 14. https://doi.org/10.1186/s12879-014-0690-1 PMID: 25551277

131. Ongruk P, Siriyasatien P, Kesorn K. New key factors discovery to enhance dengue fever forecasting model. Adv Mater Res. 2014; 931–932: 1457–1461. https://doi.org/10.4028/www.scientific.net/AMR.931-932.1457

132. Balasundaram A, Bhuvaneswari PTV. Comparative study on decision tree based data mining algorithm to assess risk of epidemic. 2013. pp. 390–396.

133. Wu Y, Lee G, Fu X, Soh H, Hung T. Mining weather information in dengue outbreak: Predicting future cases based on wavelet, SVM and GA. Lect Notes Electr Eng. 2009;39 LNEE: 483–494.

134. Zhang Y, Ibaraki M, Schwartz FW. Disease surveillance using online news: Dengue and zika in tropical countries. J Biomed Inform. 2020; 102. https://doi.org/10.1016/j.jbi.2020.103374 PMID: 31911171

135. Souza RCSNP. Detecting spatial clusters of infection risk with geo-located social media data. 2018.

136. Coberly JS, Fink CR, Elbert Y, Yoon I-K, Velasco JM, Tomayao AD, et al. Tweeting Fever: Can Twitter Be Used to Monitor the Incidence of Dengue-Like Illness in the Philippines? JOHNS HOPKINS APL Tech Dig. 2014; 32: 12.

137. Gomide J, Veloso A, Meira Jr. W, Almeida V, Benevenuto F, Ferraz F, et al. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. 2011.

138. Fang Z-H, Tzeng J-S, Chen CC, Chou T-C. A study of machine learning models in epidemic surveillance: Using the query logs of search engines. 2010. pp. 1438–1449.

139. Souza J, Leung CK, Cuzzocrea A. An Innovative Big Data Predictive Analytics Framework over Hybrid Big Data Sources with an Application for Disease Analytics. Adv Intell Syst Comput. 2020;1151 AISC: 669–680.

140. Yogapriya P, Geetha P. Dengue disease detection using K-means, hierarchical, kohonen-SOM clustering. Int J Innov Technol Explor Eng. 2019; 8: 904–907. https://doi.org/10.35940/ijitee.J9066.0881019

141. Adias Sabara M, Somantri O, Nurcahyo H, Kurnia Achmadi N, Latifah U, Harsono. Diagnosis classification of dengue fever based on Neural Networks and Genetic algorithms. 2019.

142. Jongmuenwai B, Lowanichchai S, Jabjone S. Comparision using data mining algorithm techniques for predicting of dengue fever data in northeastern of Thailand. 2019. pp. 532–535.

143. Balasaravanan K, Prakash M. Detection of dengue disease using artificial neural network based classification techniquetion. Int J Eng Technol. 2018; 7: 13–15. https://doi.org/10.14419/ijet.v7i1.3.8978

144. Acosta Torres J, Oller Meneses L, Sokol N, Balado Sardiñas R, Montero Díaz D, Balado Sansón R, et al. Decision tree technique applied to the clinical method in the dengue diagnosis. Rev Cuba Pediatr. 2016; 88: 441–453.

145. Soonthornphisaj N, Thitiprayoonwongse D. Knowledge discovery on dengue patients using data mining techniques. 2016. pp. 371–375.

146. Fathima SA, Hundewale N. Comparitive analysis of machine learning techniques for classification of arbovirus. 2012. pp. 376–379.

147. Fathima S, Hundewale N. Comparison of Classification Techniques-SVM and Naives Bayes to predict the Arboviral Disease-Dengue. In: Chen B, Chen J, Chen X, Chen Y, Cho YR, Cui J, et al., editors. 2011 Ieee International Conference on Bioinformatics and Biomedicine Workshops. Los Alamitos: Ieee Computer Soc; 2011. pp. 538–539.

148. Long ZA, Abu Bakar A, Razak Hamdan A, Sahani M. Multiple attribute frequent mining-based for dengue outbreak. Lect Notes Comput Sci Subser Lect Notes Artif Intell Lect Notes Bioinforma. 2010;6440 LNAI: 489–496.

149. Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, et al. The Global Burden of Dengue: an analysis from the Global Burden of Disease Study 2013. Lancet Infect Dis. 2016; 16: 712–723. https://doi.org/10.1016/S1473-3099(16)00026-8 PMID: 26874619

150. 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

151. World Health Organization. A Global Brief on Vector-Borne Diseases. Geneva: World Health Organization; 2014. https://apps.who.int/iris/bitstream/handle/10665/111008/WHO_DCO_WHD_2014.1_eng.pdf

152. Brownstein JS, Freifeld CC, Madoff LC. Digital Disease Detection—Harnessing the Web for Public Health Surveillance. N Engl J Med. 2009; 360: 2153–2157. https://doi.org/10.1056/NEJMp0900702 PMID: 19423867

153. Choi J, Cho Y, Shim E, Woo H. Web-based infectious disease surveillance systems and public health perspectives: a systematic review. BMC Public Health. 2016; 16: 1238. https://doi.org/10.1186/s12889-016-3893-0 PMID: 27931204

154. Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and internet-based data in global systems for public health surveillance: a systematic review. Milbank Q. 2014; 92: 7–33. https://doi.org/10.1111/1468-0009.12038 PMID: 24597553

155. Leta S, Beyene TJ, De Clercq EM, Amenu K, Kraemer MUG, Revie CW. Global risk mapping for major diseases transmitted by Aedes aegypti and Aedes albopictus. Int J Infect Dis. 2018; 67: 25–35. https://doi.org/10.1016/j.ijid.2017.11.026 PMID: 29196275

156. Monaghan AJ, Sampson KM, Steinhoff DF, Ernst KC, Ebi KL, Jones B, et al. The potential impacts of 21st century climatic and population changes on human exposure to the virus vector mosquito Aedes aegypti. Clim Change. 2018; 146: 487–500. https://doi.org/10.1007/s10584-016-1679-0 PMID: 29610543

157. Reis BY, Kohane IS, Mandl KD. An epidemiological network model for disease outbreak detection. PLoS Med. 2007; 4: e210. https://doi.org/10.1371/journal.pmed.0040210 PMID: 17593895

158. Thacker SB, Qualters JR, Lee LM, Centers for Disease Control and Prevention. Public health surveillance in the United States: evolution and challenges. MMWR Suppl. 2012; 61: 3–9.

159. Romette JL, Prat CM, Gould EA, de Lamballerie X, Charrel R, Coutard B, et al. The European Virus Archive goes global: A growing resource for research. Antiviral Res. 2018; 158: 127–134. https://doi.org/10.1016/j.antiviral.2018.07.017 PMID: 30059721

160. Dos Reis IC, Gibson G, Ayllón T, de Medeiros Tavares A, de Araújo JMG, da Silva Monteiro E, et al. Entomo-virological surveillance strategy for dengue, Zika and chikungunya arboviruses in field-caught Aedes mosquitoes in an endemic urban area of the Northeast of Brazil. Acta Trop. 2019; 197: 105061. https://doi.org/10.1016/j.actatropica.2019.105061 PMID: 31194961

161.   Jones R, Kulkarni MA, Davidson TMV, Team R-LR, Talbot B. Arbovirus vectors of epidemiological concern in the Americas: A scoping review of entomological studies on Zika, dengue and chikungunya virus vectors. PLOS ONE. 2020; 15: e0220753. https://doi.org/10.1371/journal.pone.0220753 PMID: 32027652

162.   Ajin VW, Kumar LD. Big data and clustering algorithms. 2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS). 2016. pp. 1–5.

163.   Dave M, Gianey H. Different clustering algorithms for Big Data analytics: A review. 2016 International Conference System Modeling Advancement in Research Trends (SMART). 2016. pp. 328–333.

164.   Pearl J. Causality: models, reasoning, and inference. Cambridge University Press; 2009.

165.   Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. 2009; 569.