

# Dengue Prediction Through Machine Learning and Deep Learning: A Scoping Review Protocol.

Ewerthon Dyego de Araujo Batista (✉ [edabew@gmail.com](mailto:edabew@gmail.com))

IFPB: Instituto Federal de Educacao Ciencia e Tecnologia da Paraiba <https://orcid.org/0000-0003-4993-9900>

**Frederico Moreira Bublitz**

UEPB: Universidade Estadual da Paraiba

**Wellington Candeia de Araujo**

UEPB: Universidade Estadual da Paraiba

**Romeryto Vieira Lira**

IFPB: Instituto Federal de Educacao Ciencia e Tecnologia da Paraiba

---

## Protocol

**Keywords:** scoping review, predict, forecast, machine learning, deep learning, dengue

**Posted Date:** October 26th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-95498/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** Dengue is an endemic disease caused by the DENV virus. There are four types of serology for this virus (DENV1, DENV2, DENV3 e DENV4). All of these variations can cause the disease and, once infected with one type, the patient is not immune against other serologies. Due to the particularity of the virus serology, as well as the ease of reproduction of the transmitting mosquito, approximately 4.3 million people suffered from this disease in 2019. Although it is not a new disease, there is still no effective vaccine against the virus. The best form of combat is prevention against mosquito proliferation. In this sense, machine learning and deep learning techniques have been used to predict dengue cases. In this work we show a scope review to clarify how it is possible to predict dengue cases through machine and deep learning.

**Methods:** This scope review will follow the methodology defined in the article "Scoping studies: advancing the methodology". The methodology consists of six phases. We chose to use only the mandatory ones: 1 - Identify the research question, 2 - Identify the relevant studies, 3 - Select the studies, 4 - Map the data and 5 - Compile, summarize and make the report. The main research question is to verify the feasibility of using machine learning and deep learning in the prediction of dengue cases. Derived from this question, the machine learning and deep learning techniques used will be investigated, where the studies are carried out, which data are being used, how the models are validated and which produce better results. The review used electronic databases: Scopus Document Search, IEEE Xplore Digital Library, PubMed, ACM Digital Library, and Web of Science.

**Results:** After completing this study, a technical-scientific opinion was created and the suggested protocol was executed. As a result of the execution, 301 papers were selected and 14 approved.

**Conclusions:** We can prove the effectiveness of using Machine Learning and Deep Learning techniques to predict dengue cases.

**Systematic Review registrations:** Submitted on October 16,2020 Open Science Framework

## Background

Dengue is an endemic disease caused by the DENV virus and transmitted through the *Aedes Aegypti* mosquito (1). There are four variations of the virus (DENV1, DENV2, DENV3 and DENV4) and all can cause the disease (2). The main symptoms of dengue are: high fever, muscle pain, malaise, lack of appetite and headaches. In some more severe cases, dengue can cause bleeding, difficulty breathing and cause the patient death (3).

Although it is not a new disease, there is still no effective vaccine for immunizing the population against all types of viruses. Once infected with any of the four types of the virus, the patient acquires immunity to this variation. However, it remains susceptible to the others. Additionally, in cases of reinfection by another serology, the symptoms are stronger (4). Therefore, efforts to combat dengue are directed against the proliferation of the vector.

In 2019, the World Health Organization (WHO) accounted for approximately 4.2 million cases of dengue worldwide. Previously, this same agency issued an alert classifying dengue as one of the main diseases for the year 2019 (5). In Brazil, due to the circulation of a new type of virus (DENV-2), there was a new outbreak of dengue in 2019 with an increase of 149% of cases in some states (6).

The *Aedes Aegypti* mosquito reproduces through standing water and finds in countries with tropic climate and high level of rain precipitations, that is an ideal place for its reproduction (7). To combat the disease, government systems invest in public awareness campaigns, asking for the correct disposal of tires and containers in the open, since they can accumulate water and become, in the future, a breeding ground for mosquito proliferation. (8).

Given the global health problem caused by dengue and its limitations in combat, machine learning (ML) and deep learning (DL) techniques emerged as allies in its combat. The idea is to create models for the prediction of dengue cases and outbreaks, thus providing Governments with accurate information and helping managers in decision making.

Using machine learning techniques, the Canadian company Bluedot wrote, in January 2019, an article talking about the possible outbreak of transmission of COVID-19 through the air traffic of China (9). If the recommendations were taken seriously and restrictions were imposed, how many of the 996.342 lives lost by COVID-19(on 09/28/20) could have been saved (10)?

Disease prediction is not an easy job. There are several influencing and impacting factors when creating models. Here are some examples: climatic economic and social factors, urban mobility (11). Doni et al. (12) conducted a study in India, using deep learning techniques to analyze climate data, such as temperature, precipitation and humidity data, combining them with historical data on dengue. Its model was able to predict dengue cases with an accuracy of 89%. Another example of the use of prediction through ML and DL occurred in Thailand (13). Here, in addition to climate data, research data obtained from Google Trends was used.

When conducting research in bases such as Scopus, more than 107 articles were returned on the theme of prediction combining machine learning and deep learning during the prediction of dengue cases. This fact demonstrates the attention given by science in the search for a prediction model. Therefore, a scope review is justified to verify which machine learning or deep learning techniques are being used, where studies are being carried out, how and what data are being used for predictions. Finally, we show which techniques are showing the best results and the reasons.

## Methods And Design

Levec at al (14) details a scope review methodology composed of five mandatory phases: 1 - Identify the research question, 2 - Identify the relevant studies, 3 - Select the studies, 4 - Map the data and 5 - Compile, summarize and make the report. Additionally, there is a sixth phase called consultation. Although, this is optional.

This systematic review will be structured based on the model mentioned above. However, only the mandatory stages will be applied. Moreover, researchers will use the software StArt<sup>1</sup> to assist in the definition and execution of the research protocol.

## Identifying the research question

The main objective of this systematic review is to verify the feasibility of using machine learning and deep learning techniques in predicting cases of dengue disease. To achieve the main response and achieve the goal of the review, the team used strategy defined by the acronym PICO: Population(P): patients with dengue, Intervention (I) assist health systems in predicting dengue cases, Comparisons (C) compare which ML and DL techniques are used in the prediction, Outcomes (O) the result of the predictions as well as the evaluation of the performance of these. Furthermore, in order to improve the strategy for responding, the team divided the problem into five parts: techniques, data, approach, validation and outcomes. That done, subdivisions of the main question were created, following the details separated by parts:

1. Techniques: Which machine learning and deep learning techniques are used in predictions?
2. Data: Which country was the study conducted? How was the data collected?
3. Approach: How many years of data were used in the models and which items were considered when creating the models? Example: climatic factors, economics, data from social networks, among others.
4. Validation: How were the models validated? What statistical techniques did the study use to evaluate the performance of these?
5. Outcomes: Which technique or combination of techniques achieved the best results and the reasons?

## Identifying relevant studies

After defining the first research question and its derivations, the next step was to define which databases would be relevant for the review. To research the studies, the main electronic databases were used in the areas of Health and Computer Science: Scopus, IEEE Xplore<sup>2</sup>, PubMed (Medline)<sup>3</sup>, ACM Digital Library<sup>4</sup> e Web of Science<sup>5</sup>. Other databases such as Cochrane were tested, but they did not have good indexing of articles for the subject of this review or access to data was limited. Once the bases were defined, the team started the study on the terms for the formation of the search string.

The terms used in the search string refer to forecasting, machine learning (here, deep learning is understood as a type of ML), and dengue. They are: predict\* (refer to terms like predict, prediction, predicted), forecast\* (including the words forecast, forecasting, forecasted), machine learning (there is no variation for that term), deep learning (there is no variation for that term) and dengue (references for dengue fever, fever hemorrhagic dengue and, in some countries, only dengue).

The search string was tested on the five bases selected in rounds. In each round, small adjustments were made to the terms and logical operators. The evaluation of the quality of the string was made by the team

analyzing the relevance, quality, relationship of the articles returned with the time of this research. Moreover, seventeen articles were chosen by the team and classified as indispensable in the search return. Therefore, if any of the seventeen articles were not returned, the adjustment in the string was discarded. Finally, after four rounds, the team defined the following strings by base:

- Scopus: *TITLE-ABS-KEY ((predict\* or forecast\*) AND ("machine learning" or "deep learning")) AND (dengue);*
- IEEE: *(("Full Text & Metadata":predict\* or forecast\*) AND "Full Text & Metadata":machine learning or deep learning) AND "Full Text & Metadata":dengue);*
- PubMed: *All Fields ((predict\* or forecast\*) AND ("machine learning" or "deep learning")) AND (dengue);*
- ACM: *[All: predict or forecast] AND [All: machine learning or deep learning] AND [All: dengue] AND [All: predict\* or forecast\*] AND [All: machine learning or deep learning] AND [All: dengue];*
- Web of Science: *TOPIC: (predict\* or forecast\*) AND TOPIC: (machine learning or deep learning) AND TOPIC: (dengue)*

## Study selection

The goal of this review is to understand whether it is possible to make predictions of dengue cases based on machine learning and deep learning techniques. In addition, surveys which are the main approaches and which techniques are producing the best results.

Machine learning and deep learning are current themes, constantly evolving and widely researched in medicine and computer science. Primarily, studies will be selected to predict dengue cases performed using ML and DL techniques. In order to have a better direction with the research question, the team defined inclusion (INC) and exclusion (EXC) criteria. Are they:

- INC01 - The study uses machine learning techniques?
- INC02 - The study uses deep learning techniques;
- INC03 - The study was statistically validated?
- INC04 – There is a comparison and use of more than one ML or DL technique or model in the study?
- INC05 – The study contains prediction of Dengue cases?
- INC06 – The paper must be written in English or Portuguese;
- INC07 – The study contains risk of bias (outcome or study level)?
- EXC01 - Duplicate publications. Some databases like Scopus index articles, are shown in other databases. With that, some articles may come duplicated;
- EXC02: Publications that do not meet the inclusion criteria mentioned above.

After defining the criteria, he began screening articles using the concepts and techniques described in PRISMA (15). Initially, this stage was carried out by two researchers. In order to avoid influence on the results, each one made the classification individually and without knowing the result of the other researcher.

In this stage of the process, the following steps were performed: 1 - Eliminate duplicate articles; 2 - Quick reading in the abstract and results of the articles. 3 – Apply the inclusion and exclusion criteria previously defined. In case of exclusions by any criterion, these must have their justification registered. After this process had been done, the results obtained were compared. Articles with conflicting opinions were screened by a third researcher. This, discussed the points raised by previous researchers, and, finally, deferred the final opinion in relation to the articles.

## **Charting the data**

In the Charting the Data stage, the team defines the information to be extracted from the articles. There was a meeting with the team to equalize the information to be extracted. As a result of the meeting, the researchers entered into a concession to extract the data:

- Bibliographic data;
- Which ML or DL techniques were used;
- Which country did the study take;
- How the data was collected; They are official bases?
- How many years of data were used in the samples;
- Which statistical techniques were used to validate predictions;
- Which technique performed best;
- Reviewer's considerations about the study.

The StArt tool allows researchers to include fields to catalog data from articles. As the articles were read, the reviewers, individually, imputed the information in these fields. Finally, all item data, including custom fields, are exported to a file in .xls format. The export was done individually for each reviewer. The separate files needed to be combined into a final file. In this stage conflicts arose. Once again, the existing conflicts were resolved through the analysis of a third evaluator.

## **Collating, summarizing, and reporting the results**

During the collection, summarization and reporting of results Levec at al (14) suggests division into topics: analysis, reporting e implications. We will follow this suggestion and the following details will be inserted:

- Analysis: the analysis stage will be divided into two stages. First, the team will do the quantitative analysis. Here, bibliographic data, information about machine and deep learning will be considered, in addition to the data defined in the topic chart criteria. In the quantitative analysis, the studies will be classified and verified how the articles answer the research questions. Following, a qualitative analysis will be carried out. In this, the data and methods of ML and DL discussed in the articles will be used combined with the considerations of our team of researchers.
- Reporting: aiming at a better structuring, during this phase, a table will be created, separated by columns, listing the strengths and weaknesses of the articles.

- Implications for future research: A technical opinion will be created about the models used in the prediction of dengue. The opinion will follow the model suggested by (16) and will serve as a basis for future research on the topic.

## Patient and Public Involvement

There was no involvement of the public or patients in this study.

<sup>1</sup>[http://lapes.dc.ufscar.br/tools/start\\_tool](http://lapes.dc.ufscar.br/tools/start_tool)

<sup>2</sup><https://ieeexplore.ieee.org/Xplore/home.jsp>

<sup>3</sup><https://www.ncbi.nlm.nih.gov/pubmed/>

<sup>4</sup><https://dl.acm.org/>

<sup>5</sup>[www.webofknowledge.com](http://www.webofknowledge.com)

## Discussion

This review will verify the feasibility of using Machine Learning and Deep Learning techniques to predict cases of dengue. Furthermore, it will be verified what are the main techniques used, which countries where the studies are carried out, how and what data are used in the construction of the forecast models and, finally, how these models are validated.

With this information, future research will be able to verify what are the best techniques and approaches, as well as what gaps to be filled in this area. Due to the lack of an effective vaccine, mosquito prevention is still the best way to fight the disease. Being able to produce a model with assertive prediction will help governments to fight dengue.

## Abbreviations

ML: Machine Learning, DP: Deep Learning, COVID-19: Coronavirus disease 2019, StArt: (State of the Art through Systematic Review, PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis.

## Declarations

### Ethics and Dissemination

As this is a scope review, ethical approval of this study is not necessary, since primary data will not be collected.

### Consent for publication

This review does not contain data from any individual person.

## Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the Scopus repository, <https://www.scopus.com>, IEEEExplore, <https://ieeexplore.ieee.org/Xplore/home.jsp> PubMed(Medline), <https://www.ncbi.nlm.nih.gov/pubmed/> ACM Digital Library, <https://dl.acm.org/> and Web of Science, [www.webofknowledge.com](http://www.webofknowledge.com)

## Competing interests

There are no conflicts of interest to register.

## Funding statement

This research did not receive assistance of any public, private or non-profit institution.

## Authors contributions

During the writing of this article, the team followed the division of activities: conception and writing: Ewerthon D. A. Batista, review: Frederico M. Bublitz, Wellington C. Araujo and Romeryto V. Lira.

## Acknowledgements

Not applicable

## References

1. Appice A, Gel YR, Iliev I, Lyubchich V, Malerba D. A Multi-Stage Machine Learning Approach to Predict Dengue Incidence: A Case Study in Mexico. *IEEE Access*. 2020;8:52713–25.
2. Pham DN, Aziz T, Kohan A, Nellis S, Jamil JBA, Khoo JJ, et al. How to Efficiently Predict Dengue Incidence in Kuala Lumpur. In: *Proceedings - 2018 4th International Conference on Advances in Computing, Communication and Automation, ICACCA 2018*. 2018.
3. Carvalho, T. M., Tenório, G. L., Figueiredo, K., Vellasco, M., Caarls W. Comparison of Machine Learning Models for Total Dengue Cases Prediction. 2019;
4. Swaminathan S, Khanna N. Dengue vaccine development: Global and Indian scenarios. *Int J Infect Dis* [Internet]. 2019 Jul 1 [cited 2020 Sep 21];84:S80–6. Available from: <https://doi.org/10.1016/j.ijid.2019.01.029>
5. Kerdprasop K, Kerdprasop N, Chuaybamroong P. Forecasting Dengue Incidence with the Chi-squared Automatic Interaction Detection Technique. In: *ACM International Conference Proceeding Series*. 2019. p. 37–42.
6. MinistérioDaSaúde. Ministério da Saúde alerta para aumento de 149% dos casos de dengue no país. Ministério da Saúde, Bras [Internet]. 2019 [cited 2020 Sep 21];2020. Available from: <http://www.saude.gov.br/noticias/agencia-saude/45257-ministerio-da-saude-alerta-para-aumento-de-149-dos-casos-de-dengue-no-pais>

7. 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(10).
8. Norrby R. Outlook for a dengue vaccine [Internet]. Vol. 20, *Clinical Microbiology and Infection*. 2014 [cited 2020 Sep 21]. p. 92–4. Available from: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>
9. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *J Travel Med* [Internet]. 2020;2020:1–3. Available from: [www.who.int/csr/don/](http://www.who.int/csr/don/)
10. WHO. WHO Coronavirus Disease (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard [Internet]. 2020 [cited 2020 Sep 21]. Available from: <https://covid19.who.int/>
11. Mussumeci E, Codeço Coelho F. Large-scale multivariate forecasting models for Dengue - LSTM versus random forest regression. *Spat Spatiotemporal Epidemiol*. 2020;35.
12. Doni AR, Sasipraba T. Lstm-Rnn Based Approach for Prediction of Dengue Cases in India. *Ing des Syst d'Information*. 2020;25(3):327–3355.
13. Puengpreeda A, Yhusumrarn S, Sirikulvadhana S. Weekly Forecasting Model for Dengue Hemorrhagic Fever Outbreak in Thailand. 24(3).
14. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci*. 2010 Dec;5(1):69.
15. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Available from: <http://www.prisma-statement>.
16. Brasil. Ministério da Saúde. Secretaria da Ciência TEIEDDCET, Ministério da Saúde (Brasil). Secretária de Ciência- Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. Diretrizes metodológicas: elaboração de pareceres técnico-científico. Ministério da Saúde, Secr Ciência, Tecnol e Insumos Estratégicos, Dep Ciência e Tecnol ed, revisada e atualizada-Brasília Ministério da Saúde, 2011 [Internet]. 2014;(1):80. Available from: [http://bvsmms.saude.gov.br/bvs/publicacoes/diretrizes\\_metodologicas\\_elaboracao\\_parecer\\_tecnico.pdf](http://bvsmms.saude.gov.br/bvs/publicacoes/diretrizes_metodologicas_elaboracao_parecer_tecnico.pdf)

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PRISMAPchecklist.docx](#)