Literature Review

Sudipta Kumar Das ID: 20-43658-2

November 2022

1 Background

Dengue is a disease, which occurs when a female mosquito called "Aedes aegypti" bites a person who has dengue. The vaccination of Dengue is still under development. One of Philippines's endemic illnesses is dengue. There are several research on dengue to understand how the disease spread and the nature. For that reason, many data has been collected from the patients from various hospitals and medical centers. Dengue depends much on the rain fall. Because, these mosquitos born in a still water. Therefore, the primary cause of the propagation of dengue is rainfall. Because of these rain fall, the rain waters stays in drains and sometimes in nursery areas. As the that Aedes aegypti mosquito born in still water, so it becomes a suitable place for them and a hotspot for us to get infected. In addition, this mosquito bites on the day time.

There are lots of predictions on Dengue outbreak and the deaths. There has a huge data on the dengue outbreak and the deaths of Philippines from 2016 to 2022. There are also statical data available for countries. On those datasets several researches has been conducted. In those researches several methods like linear kernel, radial kernel any several more has been used. They are accurate indeed, but there is no research how the dengue situation and deaths would be in upcoming years for Philippines. Here, in this research we are going to predict the dengue outbreak and deaths in Philippines with the help of machine learning and regression methods which fits best on the data. This research will help the government of Philippines to take preparation to deal with the upcoming situation of dengue outbreak and deaths.

2 Article searching procedure

To search and select the articles for the literature review, we followed systematic process. For this, first of all, we searched as advanced method in 2 different online databases for the research papers and articles. The databases are Science Direct and PubMed. At first, we opened the website of the Science Direct and the PubMed. After that, we searched 2 different keywords "Dengue Case" and "Machine Learning" which are related to this particular research.

Then we got a number of articles and research paper as well. Then we applied some conditions and refining which are "From year 2023", "Review articles", "Research articles", "Open access". After including these conditions, we got a few numbers of articles and research paper. Then, we chose the journal articles and research papers that were pertinent to this particular study. Then, in PDF format, we downloaded the papers and the study paper.

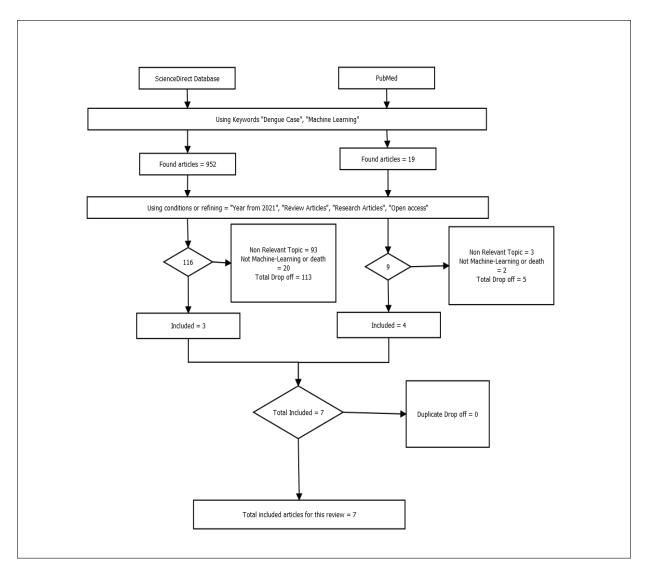


Figure 1: Article Search Result

3 Article inclusion and exclusion criteria

After refining we got a significant amount of papers, each of them are related to the research topic directly and indirectly and also free to access. Then we go through each of the papers. We go through the introduction first then the conclusion, then the methodology. By doing this, we got a clear idea about that research papers about what they have done and how then have done. Then those papers which methods is not related to regression techniques or forecast about the dengue outbreak, we drop those off.

4 Data extraction Procedure

As soon as we go through those research paper one by one by going through the introduction first then the conclusion, then the methodology, we got a clear idea about that research papers about what they have done and how then have done. Then we extract out the similar topics but differences and different techniques they have applied, different goal they have achieved. To store those extracted data, we use a standard format. This is a tabular format, which is very easy to understand and also very easy to read through the data.

Reference	Data Collections	Number of Factors	Model Used	Results
[1]	Data was gathered from hospital patients and the Bangladesh Bureau of Statistics	3	SVR and MLR model	This study forecasted dengue cases from August 2021 to May 2022 for both regression models (10 months). SVR surpassed MLR with a prediction accuracy of 75%, whereas MLR only managed a prediction accuracy of 67% (Mean Absolute Error (MAE): 4.57). (MAE:4.95).
[2]	Field data collecting over 81 weeks in 2 significant regions. Aedes data from 50 and 55 ovitraps in Selayang and Bandar Baru Bangi, respectively, were gathered	3	Autoregressive Distributed Lag (ADL) model	The notified cases related to this week's and last week's onset cases, the last three weeks' worth of larvae, the last three weeks' worth of positive PCR results, the last three weeks' worth of maximum humidity, the last four weeks' worth of rainfall, and the last four weeks' worth of air pollution index (API), all of which were significant relationships in the pooled data between notified dengue cases and all the predictors
[r3]	Data were gathered from the Sistema de Informaço de Agravo de Notifcaço (SINAN) Brazilian Information System for Notifiable Diseases.	Not Mentioned	Unclassified Machine and Deep learning model	16 show a data collection containing records relating to the transmission of Dengue, Chikungunya, and Zika, as well as analyze the introduction of Dengue and associated arboviruses (Zika and Chikungunya) in Córdoba, Argentina
$[\mathbf{r4}]$	Patient data on dengue	Not Mentioned	ANN and Random Forest Classifiers	Had the best accuracy of 58% for predicting the clinical severity of dengue during the crucial period
[r5]	Epidemiological information from the New Caledonia general population census and meteorological information from daily rainfall and the highest temperature recorded at weather stations run by Météo-France. Data on climate change from several global warming scenarios	Not Mentioned	Support Vector Machines (SVM)	The number of days where the maximum temperature surpassed 30.8°C and the mean of the daily precipitation over 80 and 60 days before to the projected week, respectively, were the strongest predictors of the weekly risk of a dengue epidemic.
[r6]	Information about dengue cases from the Guangdong Center for Disease Control and Prevention (Guangdong CDC) in Guangzhou. Data about the weather collected in Guangzhou by almost 300 weather stations	4	PSPNet, Linear Kernel and SVM	About 50–60% of the dengue cases in the city were found in the top 30% of the anticipated high-risk townships
[r7]	Weather data from the Meteorology, Climatology, and Geophysical Agency (BMKG	3	Support Vector Regression, RMSE and MAE	The latency chosen for the weather variables is longer than the lag used for earlier occurrence

Table 1: Data Extraction

5 Research questions

After going through sufficient number of review papers and research papers, we found some research questions. Those are,

- How the dengue outbreak and deaths would be in upcoming years for Philippines?
- What would be the predicted dengue cases in 2023?
- What would be the predicted dengue deaths in 2023?
- What is our current model accuracy?

6 Research Objectives

- To predict the dengue outbreak and deaths in Philippines with the help of machine learning and regression methods which fits best on the data.
- To use machine learning and regression techniques that best suit the data to forecast the dengue spread and fatalities in the Philippines.
- To get the most prediction accuracy in the model.

7 Significance of the output

Our study's findings might be used in the fields of medicine, health, disaster relief, and other things. So that the government can create public awareness of dengue and can take necessary steps, if the dengue situation get worse, then they can handle it.

References

- [1] Samrat Kumar Dey et al. "Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach". In: *PloS one* 17.7 (2022), e0270933.
- [2] Suzilah Ismail et al. "The practicality of Malaysia dengue outbreak forecasting model as an early warning system". In: *Infectious Disease Modelling* 7.3 (2022), pp. 510–525.