Dengue prediction model: A systematic review using social network analysis

S N Wahyuni¹, E Sediyono², and I Sembiring³

¹ Management Informatics, Universitas Amikom Yogyakarta, Jl. Ringroad Utara, Condongcatur, Depok, Yogyakarta, 55283, Indonesia.
² Information Technology, Universitas Kristen Satya Wacana, Jl. Diponegoro 52-60, Salatiga, Central Java 50 711, Indonesia

Email: yuni@amikom.ac.id¹

Abstract. Since 1968, Dengue Harmonic Fever's incidence in Indonesia has continued to rise and has become a public health issue. Indonesia has the largest number of Dengue Harmonic Fever cases than 30 other epidemic countries worldwide. It is very important to carry out research related to dengue cases' prediction to prevent the spread of Dengue. This literature review is intended to determine the extent of the dengue prediction approach carried out by previous researchers, and a research gap will be obtained. The algorithm used to cluster articles is a modularity algorithm, using several open-source tools to process data. The online databases used are Google Scholar and Crossref by using keywords: journal, algorithm, prediction, and Dengue. The data are taken from the expansion of 1928-2020. This study's results are 200 articles that are suitable and divided into four clusters of important articles. Also, several important parameters were obtained in the prediction study of dengue fever, namely humidity, temperature, rainfall, and population density.

1. Introduction

The Covid-19 pandemic has become an intensive concern nationally and globally. However, Dengue Harmonic Fever (DHF) in Indonesia can't ignore as it becomes a public health problem and caused death every year. The DHF disease because of the Aedes aegypti mosquito and Aedes albopictus mosquito. Both of them are female and are capable of infected children and adults [1][2]. In Indonesia, DHF is a severe case since 1968, and every year it always increases. It makes Indonesia becomes the country with the highest DHF death cases in Southeast Asia [3]. In recent years, Dengue Harmonic Fever has infected all tropical and non-tropical countries worldwide and has become one of the leading public health problems in many countries [4].

In 2014, DHF broke out in Japan with 160 confirmed cases. In European countries such as Italy, France, Tanzania, and Medina in the same year, there was also an increase in dengue cases [5]. In 2017, Sri Lanka had 186,000 DHF issues with 440 death of problems [6]. According to World Health Organization, Indonesia is the highest country for Dengue Harmonic Fever cases than 30 epidemic countries worldwide. In 2020, there were 71,633 cases reported in Indonesia. Thus, the DFH cases prediction model is vital to help the government take precautionary measures before real cases occur. The accuracy of the DHF cases prediction model will help (1) Carrying out early alertness for the explosion of DHF cases at a certain period, and they can make preparations for dealing with DHF cases (2) To control and prevent DHF cases during a Covid-19 pandemic, (3) Assisting medical officers and government or non-government agencies to take preventive measures before the occurs outbreak. Based on the description above, the research purpose in this article is to get the DHF prediction model citation review using a social network to map the DHF prediction model research area and to get the DHF research gap in the prediction model.

2. Related Works

The systematic review is one way of mapping the research to be carried out. Many researchers conduct periodic reviews to mapping research gaps. Some studies that do systematic reviews are [7] use the
Institute of Electrical and Electronics Engineers (IEEE) explore for mapping Dengue prediction literature review. The data article was taken from August 2014 - January 2015, and the result is 96 articles were clustered into five areas. Dengue Fever (DF) epidemiology, DF forecasting model, DF classification, and spatial model visualization. Paper [8] uses scientific publication data from Web of Science Core Collection - articles indexed in Science Citation Index Expanded (SCI-EXPANDED). The data are taken from 1945 – 2014. The study is searching by a combination of title, author, abstract, and author affiliation, including city and country of origin. The article results are normalized using fuzzy logic. Paper [9] uses Pubmed, Cochrane Library, ScienceDirect, and Herdin. The data was taken from 1 January 1958 – 31 December 2017. The articles checklist uses PRISMA and the result analysis based on objectives, methods, results, descriptive epidemiologic, and case reports. In this study, two article databases were compared, namely Crossref and google scholar. And the comparison results are clustered by a modular algorithm. The articles are taken from 1928 until 2020, the proposed workflow show in Figure 1.

![Proposed workflow](image)

**Figure 1. Proposed workflow**

### 3. Material and Methodology

#### 3.1. Data Collection

The data articles were taken from Crossref and google scholar. All items through several online databases, The Association for Computing Machinery (ACM) digital library, Elsevier, Institute of Electrical and Electronics Engineers (IEEE) explore, and ScienceDirect.

#### 3.2. Methodology

The study does in four methodologies. The first step is comparing all article data sources. The author compares four data sources that are SCOPUS, Google Scholar, Web of Science, and Crossref. Based on the comparison, Crossref was considered more effective because all articles on Crossref are included on several online databases. The papers were collected using the free open sources software. The second step, taking all articles relevant and related to DHF model predictions, uses some keywords: approach, parameters, and prediction model. The third step is database analysis using social network analysis with modularity algorithm proposed and using open-source network analysis software for taken the graph. The networking analysis graph is developed by connecting all keywords included in some articles. And the last step is clustering important articles into some clusters.

#### 3.3. Main path article

In this part, the principle of path analysis is identifying the process using a citation network model. The first step is to take the metadata from Crossref with open-source software. The Crossref is a choice because it is free to access, and some articles publish in some journals, including in it. Each item will extract into a network note where connect in the next step. The Network analysis develops with the following steps: (1) Article retrieval based on entering the paper title, keywords, publication journal...
name, and year of publication, (2) Storing article metadata in the form of a reference manager file and, (3) Calculating the amount of traversal for each quote on the network. The number of traversals is the frequency with which a particular quote appears on different paths in several articles, both the source article and the one cited by others. The traversal count value significantly indicates how many links with other papers [10].

3.4. Clustering of the article analysis
The article clustering analysis is doing by identify groups of similar articles and specific research domains. Some of the items have the same research topic. The article clustering develops by the modularity algorithm proposed. This algorithm is carried out in two steps: iterating the notes on the graph and assigning each letter to it by increasing modularity to lead to its group. The second step is to create the supernotes from the Cluster in step one, and this process always repeats so that this algorithm relies on efficient and effective heuristics. VOSviewer uses a modularity algorithm to build article clustering, which shows on Equation 1 [10] [11] [12]:

\[
Q = \frac{1}{2m} \sum_{ij} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)
\]

Where \( Q \) is the fraction of edges that fall within group 1 or 2, \( A_{ij} \) is the weight of the boundary between nodes \( i \) and \( j \), \( k_i \) and \( k_j \) the sum of the consequences of the advantages attached to nodes \( i \) and \( j \), respectively; \( 2m \) indicates the total sum of all the edge weights in the graph; \( c_i \) and \( c_j \) are the communities of the nodes \( i \) and \( j \), and \( \delta \) is a simple delta function.

4. Result And Observation
4.1. Analysis of Main Path
The google scholar database articles are The global main of the papers with citation networks that help define the research area's diffusion pathways. Research of Dengue prediction is conduct from 1928-2020. Globally, an article search with 32 keywords resulted in 1353 articles. The articles are searching by open-source application and produced more than 200 articles related to the keywords. The data of the paper on DHF prediction research trend is shown in Figure 2.

![Figure 2. Dengue Prediction Research Trends in 1928-2020](image)

According to Figure 2, all articles are connected in the same keyword and break down into four clusters and indicated with different colors. The first Cluster shows in red color where almost all articles talking about the start of DHF spread in the world study, DHF characteristics, symptoms, and the DHF spread
pattern. The critical papers in the first Cluster shown in Table 1, and the citation network analysis is shown in Figure 3.

The Table 1 article focuses on DHF disease, trigger factors, spread rate of DHF cases, area of spread, the death rate caused by DHF, and dengue-endemic countries [15]. One of the causes of Dengue is animal, weather, and mosquito [15-16]. The second Cluster describes all the articles that focus on developing a prediction model with several approaches and various data variables (such as the number of cases, spatial, climate, etc.). This Cluster also focuses on mathematical models, statistical models, and Machine learning models. The machine learning approach includes Linear Regression (LR) Logistics Regression, Neural Network (NN), and Random Forest. The statistical process is a time-series technique where has Moving Average (MA), Simple Regression (SR) and, Exponential Smoothing (ES). Kalman proposed the Moving Average (MA) approach in 1960. Currently, the linear function consists of past values and random noise in the Auto-Regressive (AR) model. In contrast, the time series's present value is a linear function consisting of values residues in the MA model. The Auto-Regressive Moving Average (ARMA) was developed into ARIMA to produce accurate short-term forecasting with the keys in the dependent variable's past and present values [18]. While the data has a seasonal pattern, the appropriate model is the Simple Regression Moving Average (SARIMA). The DHF prediction model algorithm is Exponential Smoothing (ES) developed by Brown in 1950 and modified by Gilchrist, generalization by Gilchrist in 1976, Roberts in 1982 and Gardner in 1985, and McKenzie in 1989 [19-22]. Several ES method types are Single Exponential Smoothing (SES). A characteristic is random data, with a small weighting model for old data and large weighted values for new data, no trend, and no seasonal patterns [23-25]. The critical article in Cluster two shows in Table 2.


Following the description in Table 2, the several parameters use for DHF prediction models are weather factors, temperature, humidity, rainfall, and spatial factors include population density and water deposition [20][26][27][28]. The third cluster focus on DHF epidemiology and Dengue endemic countries as a center of Dengue. The critical articles show in Table 3.

Table 2. Essential Articles on the Second Cluster

| Year | Author         | Journal                     | Model                                      | Parameters                                                                 |
|------|----------------|-----------------------------|--------------------------------------------|---------------------------------------------------------------------------|
| 2017 | W. Anggraeni et al. [3] | Procedia Comput. Sci        | Regression, Least Square and Natural Logarithmic | Number of dengue cases, monthly mean temperature, number of rainy days per month, monthly mean humidity. |
| 2018 | A. L. Buczak et al. [30] | PLOS One                    | Holt-Winters seasonal exponential smoothing method | The minimum and maximum temperature, rainfall, and number of Dengue cases |
| 2018 | F. Cortes et al. [31] | Acta Tro                    | ARIMA Model                                | The population density, temperature                                        |
| 2019 | V. K. Mishra [32] | International Conference on Advanced Computing | Multiple Regression.                       | Humidity, rainfall, and temperature                                        |
| 2019 | P. Somboonsak [34] | Association for Computing Machinery | An autoregressive integrated moving average (ARIMA) | The number Dengue of cases, an average temperature, and population density |
| 2019 | T. Chakraborty, S. Chattopadhyay, and I. Ghosh [2] | Phys. A Stat. Mech. its Appl | ARIMA model Neural network autoregressive model | The number of Dengue cases, temperature, humidity, and rainfall |
| 2019 | P. Somboonsak [35] | ACM Int. Conf. Proceeding Ser | SARIMA                                    | Number of dengue cases, temperature average, number of rainy, humidity.   |
| 2019 | V. Jayaraj [36] | Acta Tro                    | a multivariate Poisson regression model, SARIMA model and a SARIMA with external regressors for selection | Dengue incidence data per month, average of temperature, average of humidity and average rainfall. |

Table 3. Essential Articles on the Third Cluster

| Year | Author         | Title                                                    | Journal                                      |
|------|----------------|----------------------------------------------------------|----------------------------------------------|
| 2004 | H. T. Pohan et al., | Interleukin-18 levels in adult dengue fever and Dengue hemorrhagic fever[38] | Medical Journal of Indonesia. |
| 2017 | F. Debnath, M. Ponnaiah, and P. Acharya, | Dengue fever in a municipality of West Bengal, India, 2015: An outbreak investigation[39] | Indian Journal of Public Health, |
| 2018 | A. Chakravarty and A. Krishnan | Seasonal Outbreak of Dengue Fever in Northern India[40] | International Journal of TROPICAL DISEASE & Health, |
| 2019 | S. Lamtha and K. Bhutia, | The first outbreak of dengue fever in East Sikkim in the Northeastern part of India[41] | Journal of Family Medicine and Primary Care |
Table 3 explains DHF endemic countries such as Indonesia, India, and others. Both countries have the same climate and temperature[38][39]. Meanwhile, in cluster 4, not many researchers have conducted research related to other causes of dengue fever outbreaks. The research on this Cluster focuses on the DHF of countries endemic, such as Indonesia, India, etc., with the same climate and temperature [40]. Meanwhile, in cluster 4, few researchers have researched other causes of the DHF outbreak in India, Indonesia, and several other countries, which have tropical and sub-tropical.

5. Conclusion
The results of the comparison and grouping of articles using this network analysis technique obtained several things, among others, (1) several parameters in the prediction study of DHF were temperature, rainfall, number of cases in recent years, population density, and land use, (2) There are four crucial clusters of articles in research and, (3) The research area that is still very interesting to discuss is prediction using a machine learning approach. The development of a dengue prediction model using machine learning in research is very recommend. Machine learning approach is considered simpler and more practical. So it is expected to be able to reduce the number of dengue fever spread in Indonesia. Future research will compare several prediction models using a machine learning approach to evaluate model predictions' accuracy.

6. References
[1] G. J. Ebrahim, "Dengue and dengue hemorrhagic fever," J. Trop. Pediatr., vol. 39, no. 5, pp. 262–263, 1993, doi: 10.1093/trope/39.5.262.
[2] T. Chakraborty, S. Chattopadhyay, and I. Ghosh, "Forecasting dengue epidemics using a hybrid methodology," Phys. A Stat. Mech. its Appl., vol. 527, p. 121266, 2019, doi: 10.1016/j.physa.2019.121266.
[3] W. Angraeeni et al., "Modified Regression Approach for Predicting Number of Dengue Fever Incidents in Malang Indonesia," Procedia Comput. Sci., vol. 124, pp. 142–150, 2017, doi: 10.1016/j.procs.2017.12.140.
[4] "Treatment, prevention, and global control strategy for dengue prevention and control 2."
[5] S. Polwiang, "Estimation of dengue infection for travelers in Thailand," Travel Med. Infect. Dis., vol. 14, no. 4, pp. 398–406, 2016, doi: 10.1016/j.tmvid.2016.06.002.
[6] V. S. Aryaprema and R. De Xue, "Breteau index as a promising early warning signal for dengue fever outbreaks in the Colombo District, Sri Lanka," Acta Trop., vol. 199, no. August, p. 105155, 2019, doi: 10.1016/j.actatropica.2019.105155.
[7] A. Munir and A. Sari, "Sistematic Review: Model peramalan wabah penyakit demam berdarah," Semin. Nas. Apl. Teknol. Inf. 2015, pp. 117–124, 2015.
[8] F. H. Nieto, "Ex post and ex-ante prediction of unobserved multivariate time series: a structural-model based approach," Journal of Forecasting, vol. 26, no. 1. Wiley, pp. 53–76, 2007, doi: 10.1002/for.1017.
[9] K. A. Agrupis, M. Slade, J. Aldaba, A. L. Lopez, and J. Deen, "Trends in dengue research in the Philippines: A systematic review," PLoS Negl. Trop. Dis., vol. 13, no. 4, pp. 1–18, 2019, doi: 10.1371/journal.pntd.0007280.
[10] A. Sivaprasad, N. S. Beevi, and T. K. Manojkumar, "Dengue and Early Warning Systems: A review based on Social Network Analysis," Procedia Comput. Sci., vol. 171, no. 2019, pp. 253–262, 2020, doi: 10.1016/j.procs.2020.04.027.
[11] C. Prem Sankar, K. Asokan, and K. Satheesh Kumar, "Exploratory social network analysis of affiliation networks of Indian listed companies," Soc. Networks, vol. 43, pp. 113–120, 2015, doi: 10.1016/j.socnet.2015.03.008.
[12] S. Emmons, S. Kobourov, M. Gallant, and K. Börner, "Analysis of network clustering algorithms and cluster quality metrics at scale," PLoS One, vol. 11, no. 7, pp. 1–18, 2016, doi: 10.1371/journal.pone.0159161.
[13] "DENGUE FEVER," Journal of the American Medical Association, vol. 91, no. 18. American Medical Association (AMA), p. 1377, 1928, doi: 10.1001/JAMA.1928.02700180047015.
[14] N. ISHII, "STUDIES ON DENGUE FEVER I. STUDIES ON DENGUE VIRUS AND IMMUNITY. (2)," The Japanese Medical Journal, vol. 1, no. 3. Editorial Committee of Japanese Journal of Infectious Diseases, National Institute of Infectious Dis, pp. 228–238, 1948, doi: 10.7883/yoken1948.1.228.

[15] K. C. Shekhar and O. L. Huat, "Epidemiology of Dengue/Dengue Hemorrhagic Fever in Malaysia-A Retrospective Epidemiological Study. 1973-1987. Part II: Dengue Fever (DF)," Asia Pacific Journal of Public Health, vol. 6, no. 3. SAGE Publications, pp. 126–133, 1992, doi: 10.1177/101053959200600302.

[16] G. J. EBRAHIM, "Dengue and Dengue Haemorrhagic Fever," Journal of Tropical Pediatrics, vol. 39, no. 5. Oxford University Press (OUP), pp. 262–263, 1993, doi: 10.1093/tropej/39.5.262.

[17] "Dengue and Dengue Hemorrhagic Fever in the Americas, 1986," JAMA: The Journal of the American Medical Association, vol. 259, no. 12. American Medical Association (AMA), p. 1781, 1988, doi: 10.1001/jama.1988.03720120009008.

[18] H. B. Zaman et al., "SnoezelenCAVE: Virtual Reality CAVE Snoezelen Framework for Autism Spectrum Disorders," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9429, pp. 443–453, 2015, doi: 10.1007/978-3-319-25939-0.

[19] L.-M. Liu and G. Hudak, "Forecasting and Time series analysis using the SCA statical system," Sci. Comput. Assoc., vol. 142, no. 3, p. 614, 1992, doi: 10.1007/SpringerReference_6246.

[20] A. L. Rosas and V. M. Guerrero, "Restricted forecasts using exponential smoothing techniques," Int. J. Forecast., vol. 10, no. 4, pp. 515–527, 1994, doi: 10.1016/0169-2070(94)90020-5.

[21] E. S. Gardner, "Exponential Smoothing: The State of the Art," vol. 4, no. August 1984, pp. 1–28, 1985.

[22] E. S. Gardner, "Exponential smoothing: The state of the art-Part II," Int. J. Forecast., vol. 22, no. 4, pp. 637–666, 2006, doi: 10.1016/j.ijforecast.2006.03.005.

[23] K. K. Lai, L. Yu, S. Wang, and W. Huang, "Hybridizing Exponential Smoothing and Neural Networks," Int. Conf. Comput. Sci., pp. 493–500, 2006.

[24] D. E. Gardner, "Weight factor selection in double exponential smoothing enrollment forecasts," Res. High. Educ., vol. 14, no. 1, pp. 49–56, 1981, doi: 10.1007/BF00995369.

[25] P. H.M.NishanthiHerath, A. A. I. Perera, and H. P. Wijekoon, "Prediction of Dengue Outbreaks in Sri Lanka using Artificial Neural Networks," Int. J. Comput. Appl., vol. 101, no. 15, pp. 1–5, 2014, doi: 10.5120/17760-8862.

[26] D. Mahdiana, A. Ashari, E. Winarko, and H. Kusnanto, "A Model for Forecasting the Number of Cases and Distribution Pattern of Dengue Hemorrhagic Fever in Indonesia," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 11, 2017, doi: 10.1109/KCIC.2018.8628533.

[27] M. R. Mufid, N. R. Kusuma Saginta Putri, A. Fariza, and Mu’Arifin, "Fuzzy logic and exponential smoothing for mapping implementation of dengue hemorrhagic fever in Surabaya," Int. Electron. Symp. Knowl. Creat. Intell. Comput. IES-KCIC 2018 - Proc., pp. 372–377, 2019, doi: 10.1109/KCIC.2018.8628533.

[28] J. Albinati, W. Meira, G. L. Pappa, and A. G. Wilson, "Efficient Gaussian process-based inference for modeling Spatio-temporal dengue fever," Proceedings - 2017 Brazilian Conference on Intelligent Systems, BRACIS 2017, vol. 2018-Janua. pp. 61–66, 2017, doi: 10.1109/BRACIS.2017.13.

[29] A. L. Bucazk, B. Baugher, L. J. Moniz, T. Bagley, S. M. Babin, and E. Guven, "Ensemble method for dengue prediction," PLoS One, vol. 13, no. 1, pp. 1–23, 2018, doi: 10.1371/journal.pone.0189988.

[30] F. Cortes et al., "Time series analysis of dengue surveillance data in two Brazilian cities," Acta Trop., vol. 182, pp. 190–197, 2018, doi: 10.1016/j.actatropica.2018.03.006.

[31] V. K. Mishra, "Dengue Disease Spread Prediction Using Twofold Linear Regression," Proceedings of the 2019 IEEE 9th International Conference on Advanced Computing, IACC 2019. pp. 182–187, 2019, doi: 10.1109/IACC48062.2019.8971567.
[33] P. Rangarajan, S. K. Mody, and M. Marathe, "Forecasting dengue and influenza incidences using a sparse representation of Google trends, electronic health records, and time-series data," *PLoS Comput. Biol.*, vol. 15, no. 11, pp. 1–24, 2019, doi: 10.1371/journal.pcbi.1007518.

[34] P. Somboonsak, "Forecasting Dengue Fever Epidemics using ARIMA Model," pp. 144–150, 2019, doi: 10.1145/3375959.3375970.

[35] P. Somboonsak, "Time series analysis of dengue fever cases in Thailand utilizing the SARIMA model," *ACM Int. Conf. Proceeding Ser.*, pp. 439–444, 2019, doi: 10.1145/3377170.3377215.

[36] V. Jayaraj, "Developing a dengue prediction model based on the climate in Tawau, Malaysia," *Acta Trop.*, vol. 197, 2019, doi: 10.1016/j.actatropica.2019.105055.

[37] Y. C. Cheng et al., "Real-time dengue forecast for outbreak alerts in southern Taiwan," *PLoS Negl. Trop. Dis.*, vol. 14, no. 7, pp. 1–18, 2020, doi: 10.1371/journal.pntd.0008434.

[38] H. T. Pohan et al., "Interleukin-18 levels in adult dengue fever and dengue hemorrhagic fever," *Medical Journal of Indonesia*. Faculty of Medicine, Universitas Indonesia, p. 86, 2004, doi: 10.13181/mji.v13i2.136.

[39] F. Debnath, M. Ponnaiah, and P. Acharya, "Dengue fever in a municipality of West Bengal, India, 2015: An outbreak investigation," *Indian Journal of Public Health*, vol. 61, no. 4. Medknow, p. 239, 2017, doi: 10.4103/ijph.ijph_309_16.

[40] A. Chakravarty and A. Krishnan, "Seasonal Outbreak of Dengue Fever in Northern India - A Clinical Perspective and Predicting Length of Hospital Stay," *International Journal of TROPICAL DISEASE & Health*, vol. 32, no. 3. Science domain International, pp. 1–8, 2018, doi: 10.9734/ijtdh/2018/44585.

[41] S. Lamtha and K. Bhutia, "First outbreak of dengue fever in East Sikkim in Northeastern part of India," *Journal of Family Medicine and Primary Care*, vol. 8, no. 3. Medknow, p. 1007, 2019, doi: 10.4103/jfpmc.jfmpc_397_18.