Early Identification Model for Dengue Haemorrhagic Fever (DHF) Outbreak Areas Using Rule-Based Stratification Approach

Agus Qomaruddin Munir[1,2]   Sri Hartati[3]*   Aina Musdholifah[3]

[1]Doctoral Program Department of Computer Science and Electronics,
Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta, Indonesia
[2]Department of Management Informatics, Faculty of Sciences Technology,
Universitas Respati Yogyakarta, Yogyakarta, Indonesia
[3]Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science,
Universitas Gadjah Mada, Yogyakarta, Indonesia
* Corresponding author's Email: shartati@ugm.ac.id

Abstract: The spread of dengue hemorrhagic fever (DHF) globally with a frequency level that tends to be high in the past 50 years raises a systematic idea of prevention. One of the efforts to prevent DHF is the need for early identification of areas that are potentially epidemic. Early identification is carried out by getting an overview of the incident one step ahead by data forecasting. The focus of the study was the development of area stratification algorithms as an early identification of DHF outbreak areas by using data forecasting methods with surveillance data variables. Surveillance data which became the references for system modeling were DHF case data, rainfall, humidity, air temperature, wind speed and Larva-free Number (ABJ) for the span of 2010-2016 in 17 districts in Sleman Regency, Special Region of Yogyakarta. There were four steps during the study, i.e., 1) Forecasting of DHF case for the period of 12 months, 2) Forecasting of Larva-free Number (ABJ), 3) Determination of DHF case pattern for the last three years and the average of ABJ in the third year and 4) Area classification into stratification class. A method used for data forecasting of DHF case was seasonal autoregressive moving average (SARIMA), and the determination of area class pattern was conducted by using a neural network, meanwhile to obtain area stratification class used rule-based approach referring to guidelines controlling DHF outbreaks of the Ministry of Health of the Republic of Indonesia. Early identification was carried out by dividing into 4 area classes. Area class target included endemic (K1), sporadic (K2), a potential (K3) and free (K4). The testing of accuracy forecasting used relative mean absolute error (RMAE) for 12 months period. The results of the forecasting accuracy test on 17 districts in Sleman Regency showed RMAE average of 1.46 was considered low for it was still below 10%. Furthermore, the results of the early identification of area stratification classes in 2014 and 2015 from 17 districts showed that 3 of the four regions were endemic areas while in 2016 almost all districts were endemic areas and only one area was classified as sporadic.

Keywords: DHF, Neural network, Rule-based stratification, Seasonal ARIMA, Area stratification.

1. Introduction

Dengue haemorrhagic fever (DHF) is a disease caused by dengue virus infection transmitted from one human to another through the bite of the *Aedes* mosquito from the subgenus *Stegomyia* [1], found in almost all parts of the world, especially in tropical and subtropical countries, both as endemic and epidemic diseases [2, 3]. The distribution of DHF was influenced by many factors (multi-factors) [4-7] among others climate & weather, population characteristics and environment [6, 8]. Based on data from the Ministry of Health in Indonesia, in 2015 there were 90,425 cases of DHF, and 816 patients died. In 2016, there were 112,511 cases, and 871 patients died. In 2017 to mid-December, it was recorded that DHF patients in 34 provinces in Indonesia were 71,668 people, and 641 of them died based on data in Ministry of Health. It happens due to

many factors which are known as the health epidemiology in public health science.

The health epidemiology triangle consists of host, agent, and an environment in which each has a role in DHF transmission. The agent factor in the form of the dengue virus is transmitted through *Aedes aegypti*. Another factor had a big part in the environment, for example, geographical location, climatic and weather conditions in an area, which affects the increase of the spread pattern of DHF cases [9-11]. The trend of DHF cases generally occurs in the tropics and appears in the rainy season then decreases after the rainy season ends. This virus arises due to the influence of seasons and nature as well as human behavior. Next is the host factor, i.e., from the human side itself has the same role as the two previous factors, for example, the population, the number of dengue cases, the number of DHF patients died, and so on. The increase in the prevalence of infectious diseases can be influenced by regional spatial phenomena which have similar characteristics [12, 13]. DHF transmission pattern is formed from each health epidemiological factors, the interrelationship between factors, daily local conditions affecting the source of the disease, and environmental and population risk factors [14].

Development of the early identification of maps of DHF outbreaks had been carried out by previous researchers using several approaches. Research results showed that the risk map for the spread of DHF outbreaks is essential as an area early identification [15, 16] to reduce the tendency for a full spread. The same approach was conducted in Selangor Malaysia to map the distribution of DHF outbreaks. The purpose was to find out which areas were having the potential risk of DHF prevalence [12, 17]. Several regions in Indonesia had already conducted researches on DHF risks mapping such as Kendari City – South East Sulawesi [18], South Denpasar, Magelang, Special Region of Yogyakarta, Banggai Regency and Palopo City, in which the purpose was to figure out which areas are most predominantly affected by DHF outbreaks. The problem of undetected virus spread supports the occurrence of dengue outbreaks. DHF usually occurs in parallel from one region to another due to a uniform spread of the virus [19]. Some previous studies performed the mapping after the DHF case occurred while this study developed an early identification model of areas having the potential to epidemic before dengue cases arise, by forecasting the number of dengue cases data involving health epidemiology variables including data on the number of dengue cases, larva-free number (ABJ), rainfall, air humidity, temperature, and wind speed. Based on

this reason, an early identification of area stratification was needed to provide an overview of the severity of an area against the high prevalence of infectious diseases [19]. The stratification was carried out by referring to the control regulation of directorate general on the spread of the illness of the Ministry of Health of the Republic of Indonesia.

The model developed after this referred to as the DHF epidemic area stratification system (DSSys), is one model that supports surveillance activities that shows that an area has the potential for epidemic so that a control tool is needed in the form of actual early identification as a form of prevention. It is one of the main advantages for the Ministry of Health to carry out prevention in the period before the outbreak.

In [20] to get the stratification class of the dengue outbreak area, it must wait for DHF cases to occur while the proposed model does not have to wait for dengue cases to happen.

Differences from other studies are 1) Some studies do not involve the larva-free number variable (ABJ) for DHF surveillance activities for the concept of prediction. 2) The combination of the seasonal ARIMA method for forecasting the number of cases and ABJ which then utilizes the stratification algorithm to obtain trends in regional patterns that have the potential to occur in dengue cases.

2. Related work

The purpose of early identification area is to prevent the occurrence of outbreaks before the outbreak. The early identification area to predict the occurrence of an epidemic can be done by involving several variables in the study, for example in the case of dengue outbreak using climate and weather variables [6, 7, 21], vector variables mosquitoes and economic indicators [22, 23] and so on.

SARIMA method in forecasting the incidence of DHF has been done by [24, 25], the advantage of using the SARIMA method is that the data has an internal structure, such as autocorrelation, seasonal trends or variations so affect climate and weather conditions in tropical regions such as Indonesia. SARIMA is popular concerning seasonal conditions in the data, and also in classifying the value of forecasting [26].

The application of forecasting of event data in Indonesia has been carried out in several studies, for example forecasting the incidence of dengue in the city of Palembang by determining the potential of the occurrence of the month but only involves one variable namely the incident case. Furthermore, the prevention model for DHF outbreaks is a prediction

Table 1. Presentation table of DHF patients number per year at Area-X, the year of 2014-2016

No	Area	2014	2015	2016	Larva Index	Stratification
1	Mekar	6	5	8	--	Endemic (K1)
2	Jaya	5	0	3	--	Sporadic (K2)
3	Megah	0	0	0	< 95%	Potential (K3)
4	Sukasari	0	0	0	≥ 95%	Free (K4)

model [27, 28] which only has the objective of short-term forecasting of the ongoing outbreak or model early warning system [16, 29, 30] which aims to be a descriptive analysis of the region that has the potential to endemic but no regional cluster patterns in each area are predicted.

The research that has been carried out is still limited to descriptive analysis in the form of a numerical table of each variable characteristic. Also, the results of the forecasting are not discussed about the threshold of the accuracy of the results which are then made into early identification model for the area of dengue occurrence. The main problem in this study is that to determine the class of stratification of the region must wait for dengue cases to occur. Therefore we need a prediction model of regional stratification patterns before the DHF cases occurs.

The early identification of the DHF event area referred to in this study is the information presented on the regional map is the result of the forecasting process using the Seasonal Autoregressive Moving Average (SARIMA) method to obtain the value of the number of cases and the potential occurrence of the month, while the Neural Network is used to form a cluster pattern region based on data input on dengue cases and larva-free numbers. Determination of regional classification patterns is carried out by referring to the rules for determining regional stratification from the Ministry of Health of the Republic of Indonesia regarding the status of the region against DHF.

3. Area stratification by the rule of the ministry of health in Indonesia

Area stratification needs investigation following procedures and stages of the steps. The first step that must be carried out is the epidemiological surveillance of dengue hemorrhagic fever (DHF) in community health centers which includes the collection and recording of data on suspected DHF in the context of Epidemiology Investigation (PE) [20]. Data processing and presenting is according to the following table example:

From table 1 above, it is found out that the prevalence of DHF cases was calculated based on the year and location of the incidence in the last three years. Then, larva-free number (ABJ) was estimated on average each year in 12 months in one area. Stratification status consists of 4 classes namely endemic, sporadic, potential and free. This study developed a class model of this area stratification using rules-based made into algorithm form of DHF outbreak area stratification.

4. Research methodology

4.1 Study area

Sleman Regency is a part of Special Region of Yogyakarta province covering an area of 574.82 km^2 and divided into 86 villages. It has 25 community health centers covering 17 districts based on data in Statistic Centers Bureau (BPS District Sleman). The estimated population in 2015 was 1,148,479 people with a population growth of 1.97% per year.

4.2 Data and materials

This study used data on cases of DHF prevalences in 17 districts of Sleman Regency, Special Region of Yogyakarta province. DHF is closely related to climate and weather [31, 32], thus in addition to data on the number of DHF prevalences collected from the Health Service of Sleman Regency, the other data used were the climate & weather data obtained from the Meteorology, Climatology and Geophysics Agency and Adi Sucipto Airport Yogyakarta including variables (in the form of rainfall, humidity, air temperature and wind speed).

Another data input was Larva-free Number (ABJ) obtained from Health Service of Sleman Regency, Yogyakarta. The larva-free number in the epidemiological triangle hints that the virus is one of the major contributors to the occurrence of DHF outbreaks [16, 33]. The period used is monthly starting from 2010 to 2016. Next, from the three variables, the forecasting process was conducted using the seasonal ARIMA method to obtain the output in the form of the number of DHF cases. Table 2 shows data characteristics used in this study. Whereas for data example of District_1 area is available in table 3 below:

Table 2. Variables and data characteristics of early identification model of DHF outbreaks Area Stratification

No	Variable's Name	Unit	Symbol	Function
1	Number of DHF Prevalence	Case	X_1	A parameter for the number of DHF Prevalence
2	Larva-free number (ABJ)	%	X_2	A parameter of density level of mosquitoes
3	Rainfall	mm	X_3	Average of rainfall in each area
4	Humidity	%	X_4	A parameter of humidity in each area
5	Temperature	ºC	X_5	A parameter of temperature in each area
6	Wind Speed	Knot	X_6	A parameter of wind speed in each area

Table 3 Data sample of district_1 area with variables including number of cases, larva-free number, rainfall, air humidity, air temperature, and wind speed

Month	DBD	ABJ	CH	KL	SH	KA
1	3	87.93	400.7	84	26.3	6
2	9	94.33	322.7	84	26.4	5
3	11	88.66	334.7	82	26.3	5
4	5	86.44	264.6	84	26.8	4
5	11	86.65	177.3	81	27.0	4
6	2	85.61	0.000	75	26.3	4
7	3	85.63	0.000	74	26.3	4
8	1	86.31	0.000	70	26.0	4
9	1	86.57	0.500	71	26.5	6
10	1	88.84	19.00	71	27.9	5
11	3	88.34	371.3	83	27.2	4
12	1	88.35	388.5	83	27.2	5

Where:

DBD = Number Cases of DHF
ABJ = Larva-free Number (%)
CH = Rainfall (mm)
KL = Air Humidity (%)
SH = Air Temperature (ºC)
KA = Wind Speed (Knot)

4.3 Pre-processing data

Pre-processing data was performed before forecasting using Seasonal ARIMA, and to obtain

Table 4. The result of test of data stationary using ADF test

No	Name Of District	ADF Value of DHF Case Numbers	Lag Order	P-Value
1	District_1	-2.1046	4	0.5327
2	District_2	-2.7882	4	0.2544
3	District_3	-3.4123	4	0.0607
4	District_4	-3.0695	4	0.1398
5	District_5	-5.2621	4	0.0100
6	District_6	-3.1851	4	0.0971
7	District_7	-3.2941	4	0.0796
8	District_8	-3.7845	4	0.0243
9	District_9	-3.4535	4	0.0541
10	District_10	-3.8666	4	0.0206
11	District_11	-3.7113	4	0.0300
12	District_12	-3.9918	4	0.0150
13	District_13	-3.8351	4	0.0220
14	District_14	-3.5061	4	0.0476
15	District_15	-3.2506	4	0.0866
16	District_16	-2.8656	4	0.2229
17	District_17	-2.9222	4	0.1998

good forecasting value in statistics, it is necessary to test the stationary of time series data. Test of data stationery was conducted to series data of DHF number cases in 17 districts and series data of climate such as rainfall, air humidity, air temperature, and wind speed for an area of Sleman Regency. The testing was executed by using Augmented Dicky Fuller test (ADF Test) technique with the results of a trial that the time series data is stationarily shown in Table 4.

Testing using the ADF Test method in Table 3 shows that all data series used in this research (number of dengue cases, number of larvae, humidity, air temperature, rainfall and wind speed) are stationary because they have negative ADF values [34], so the method Seasonal ARIMA can be used in forecasting. In 17 sub-districts in the Sleman district, for example, the Gamping sub-district, the ADF Test value -2.1046, on the lag order 4 and the p-value 0.5327 shows the stationary data series which is ready to be implemented in the data forecasting process.

Furthermore, after the data have been tested for its stationary, analysis is needed for determining the amount of lag that will be used in forecasting. In the trials of this study, the best lag number was 72 lags.

4.4 Forecasting step

The next step of forecasting the number of cases was to detect the sequence of events in the last three years, the larva-free number per month on average in the past year and then categorized into area classes.

From the data on the number of DHF cases in the last three years, ABJ and area classes; the distribution of training data for the 2010-2015 period and data testing for the 2016 data period were carried out and then they were used as input for neural network data to obtain area class patterns. Determination of area classes was made using a rule-based approach with the basic rules of stratification determination of the DHF outbreak area of the Ministry of Health of Indonesia which was divided into four namely endemic, sporadic, potential and free. Forecasting results, area class patterns, and determination of area stratification was tested respectively. Forecasting of DHF, ABJ and area class patterns used RMSE and MAE, while a determination of area stratification was performed by comparing real data in the field for the previous year data period. The technique of determining the stratification of this area is then used as a basis for reference in prevention, area control and supervision efforts before the outbreak takes place.

4.5 Validation model

Validation of data forecasting models for DHF and ABJ cases was carried out to determine the accuracy of the model in predicting the two variables in the next 12 months. The results of the model validation used data on monthly DHF prevalences in 2010-2015 as data training to predict the incidence and average of ABJ in 2016 which were then compared with the empirical data of 2016 DHF incidence and average of ABJ.

5. Early identification model of area stratification using forecasting technique

This study focused on the development of early identification stratification models for DHF outbreak areas using forecasting techniques involving predictor variables such as the number of dengue cases, climate & weather (rainfall, air humidity, air temperature and wind speed) and larva-free number. The model was developed by predicting the number of cases using the seasonal ARIMA and the neural network method. A neural network was used to overcome data with linear and nonlinear patterns in time series problems [35]. Fig. 1 shows the development flow of the early identification stratification model for outbreak areas from the results of forecasting the number of DHF cases. Data input used was the number of DHF cases, weather & climate, and larva-free number (ABJ). The training data used was monthly data for the period of 2010 to 2015, while the 2016 data was used as testing data. The value results of forecasting the number of DHF cases using the seasonal ARIMA method were then used as neural network input to obtain the area classes classified as K1, K2, K3 or K4 classes. The next step was to detect the prevalence of DHF cases in the last three years and to identify larva-free number (ABJ) in which the results were then classified using the rule-based approach using area stratification of the Ministry of Health of the Republic of Indonesia to obtain area stratification class. The primary output in the development of this model was the area stratification class based on the input variable of the number of cases, climate & weather, and larva-free number (ABJ).
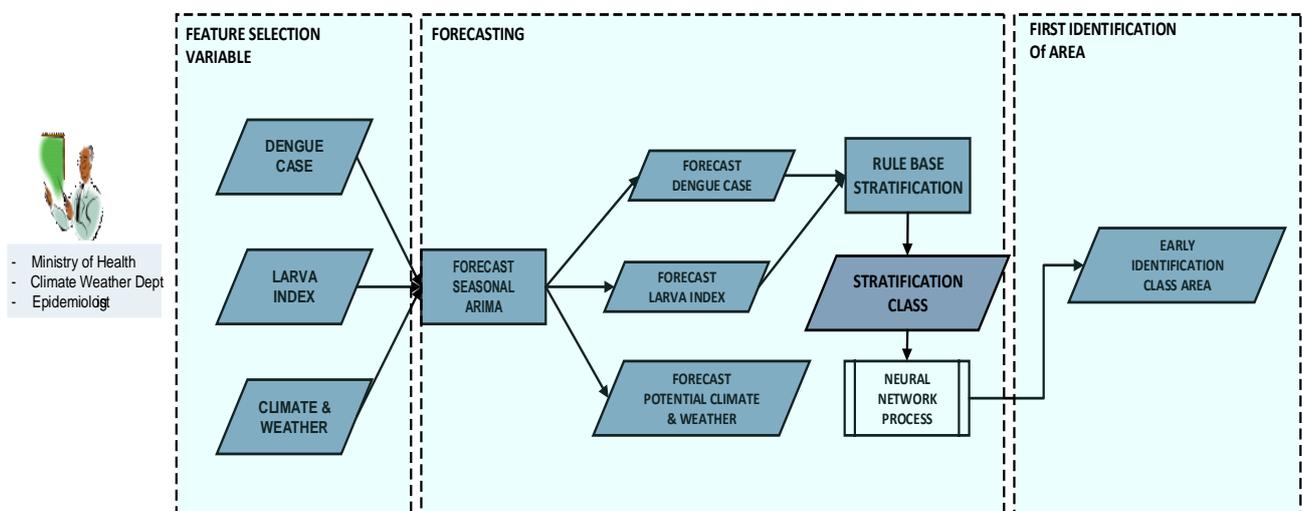

Figure. 1 Early identification model for area stratification class of DHF outbreaks using forecasting technique
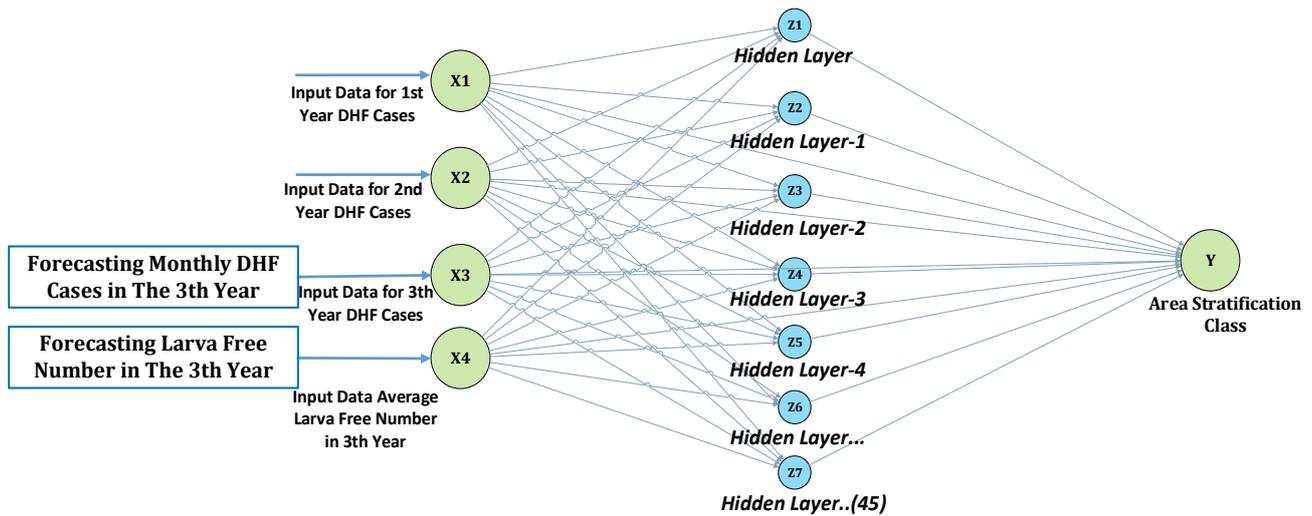
Figure. 2 Architecture of the neural network sub process early identification model

Table 5. Sample classification test data stratification of outbreak areas

NO	Name Of District	DHF Year In			ABJ	Class
		2012	2013	2014		
1	Kec Gamping	60	128	106	87.1	K1

Table 6. Calculation results of hidden layer

Kec	X1	X2	X3	X4	Y	Pred. Out 1
1	0.415	0.255	0.3	0.7347	0.4	0.3559036

In the architectural model of Fig. 1, the rule base stratification is used to obtain the class stratification area which is then used as an initial identification of the area of DHF outbreak. Neural network architecture from the early identification model for area stratification sub-process consists of 3 layers, namely the input layer, hidden layer, and output layer. Furthermore, the sub-process to determine the best architecture of the neural network in the stratification class area can be seen in Fig. 2.

The best architecture is obtained by dividing into 10 training test each hidden neuron 5, 10, 20, 30 and 45. Next from each trial sum square error (SSE) is calculated to get the error value. The input layer is an input that contains a vector from the test data that the class will find. The input layer aims to distribute the input data in the hidden layer. The following is presented one sample vector test data that is used can be seen in the following Table 5.

Hidden layers contain vector data training that has been known for its class. Hidden layers are used to calculate the distance between the test data vector and the training data vector.

The output layer is the process of finding the maximum value of the vector output, then generating the value of the stratification class decision. Based on table 1, the test data in table 2 belongs to class K1.

A rule-based approach was used as rule base in determining whether the district in mention was endemic (K1), sporadic (K2), a potential (K3) or free (K4) stratification class. To conduct area stratification class, the following three steps were required:

1. Determination of DHF case numbers in Certain Area

Data of the number of DHF cases were obtained from the Health Service of Sleman regency with a division of 17 districts in the form of 1800 Excel MS Excel records. To get data on the number of cases per month for each district was conducted by following the formula below:

$$JK = \sum_{i=1}^{n}(C) \qquad (1)$$

Where;

JK = *Number of DHF cases per month for each district.*

C = *The number of cases DHD that occur in the district.*

n = *The number of districts.*

For example Gamping District in 2016 is calculated using Eq. (1) above, by making the following Table 7:

Table 7. Sample of calculating the number of DHF cases in each district per year

Name of District (Wi)	Year	Month (b)					Case (JKi)
		Jan (b1)	Feb (b2)	Mar (b3)	to ...	Dec (b12)	
Distcrict_1	2015	4	16	7	...	0	57
......	...	...	...	...	...	...	...
District_ 17	2015	1	1	0	...	0	2

Table 8. The example lag event in the past 3 years

Period (P)	Year (T)	Month (B)				
		Jan (B1)	Feb (B2)	March (B3)	Up to	Dec (B12)
P_1	2014	Weeks $_{1-4}$	Weeks $_{1-4}$	Weeks $_{1-4}$	....	Weeks $_{1-4}$
P_2	2015	Weeks $_{1-4}$	Weeks $_{1-4}$	Weeks $_{1-4}$	....	Weeks $_{1-4}$
P_3	2016	Weeks $_{1-4}$	Weeks $_{1-4}$	Weeks $_{1-4}$	....	Weeks $_{1-4}$

The calculation is done by summing up the incidence of DHF cases every month for one year for each sub-district area, while for this study it takes a period of three years for each region with a total of 17 sub-districts in the Sleman district.

2. Detection of DHF Case Prevalence in the Last 3 Years

The prevalence of DHF that occurred in each area in the last three years was detected whether there was an event for three consecutive years, or there were cases but not consecutively or there were no cases during the previous three years but had a larva-free number of less than 95%. The flow arrangement for event detection in the last three (3) years used the following table example.

The sequence of DHF cases was detected by calculating the number of cases each month then summed up within 12 months (1 year) and sorted by year of occurrence following the Table 4 format for the last three years. After the identification of the number of cases based on the following year, the area stratification is processed using the rule-based approach.

Where:
- T = Year of DHF prevalence in an area in the last three years, T={1..3}.
- B = A month of prevalence in a year of DHF incidence, B={1..12}.
- K = The number of cases of DHF prevalence in month series in 3 years of prevalence, K={1..∞}.
- P = Period occurring from the lag of year of DHF prevalence, P={1..12}
- A = The average percentage of a larva-free number in a previous year in the area of DHF cases, A={1..100}

- S = Stratification class in the area of DHF cases, S={1..4} → $S_{1 = endemic}$, $S_{2 = sporadic}$, $S_{3 =potential}$ and $S_{4 = free}$.
- W = District area in which DHF cases occur, W={1..17}

3. Rule-Based for the Determination of Area Stratification Class

The following is the rule for determination of area stratification class following DHF control guidelines of the Minister of Health of the Republic of Indonesia.

1. **IF** Case **in** Year=[1..3] Consecutively **THEN** Area =$S_{1=endemic}$
2. **IF** Case in Year=[1..3] Not Consecutively **THEN** Area=$S_{2=sporadic}$
3. **IF** Case **in** Year =[1..3] NONE **AND** (A_{BJ} <= 95%) **THEN** Area=$S_{3=potential}$
4. **IF** Case **in** Year =[1..3] NONE **AND** (A_{BJ} > 95%) **THEN** Area=$S_{4=free}$

Then, the algorithm flow was derived into pseudo-code. The following is a pseudo-code to detect the sequence of DHF events within the last three years classified into endemic class, while for sporadic, potential and free class follow the rule-based determination of outbreak areas stratification.

Algorithm: Stratification Class Area
Input: JK: The number of DHF cases
       W: District that DHF Cases Occur
       A: The average percentage of Larva Free Number
         In previous year in the area of DHF case
Output: Class of Stratification Area

Start:
1. *for* first_year=1 *to* third_year *do*
2. *begin*
3. *for* month ← [1..12] *to* month ← [25..36] *do*
4. *if* (JK_first_year ≥ 1) *and* (JK_second_year ≥ 1) *and*

---

5.   (JK_third_year ≥ 1) *then*   W ← Endemic;
6.   *elseif* (JK_first_year ≥ 1) *and* (JK_second_year = 0)
     and (JK_third_year ≥ 1) *then* W ← Sporadic;
7.   *elseif* (JK_first_year = 0) *and* (JK_second_year ≥ 1)
     and (JK_third_year ≥ 1) *then* W ← Sporadic;
8.   *elseif* (JK_first_year ≥ 1) *and* (JK_second_year ≥ 1)
     and (JK_third_year = 0) *then* W ← Sporadic;
9.   *elseif* (JK_first_year = 0) *and* (JK_second_year = 0)
     and (JK_third_year = 0) *and* (A ≤ 95%)
       then W ← Potential;
10. *elseif* (JK_first_year = 0) *and* (JK_second_year = 0)
     and (JK_third_year = 0) *and* (A ≥ 95%)
       then W ← Free;
11.   *endif;*
12. *endfor;*
13. *end.*

4.   Determination of Area Classes

The next process was to identify whether the area was classified into endemic (K1), sporadic (K2), potential (K3) and free (K4) area.

6.   Result and discussion

6.1 Forecasting of DHF prevalence and larva-Free number (ABJ)

This study generated the prediction of the number of DHF cases using 6-years data series (between 2010 and 2015) on DHF cases in Sleman Regency consisting of 17 districts. The forecasting technique used was the seasonal ARIMA method, then the results of the forecasting were grouped into four classes called stratification classes, namely; 1) endemic (K1), 2) sporadic (K2), 3) potential (K3) and 4) free (K4).

To determine the area stratification class in 2014, the data on the number of dengue cases in the previous two years, 2012 and 2013, were needed. Sleman regency consisted of 17 sub-districts which 16 districts belong to K1 stratification class, 1 district belong to K2 with ratio of 94.2 % endemic and 5.68 % sporadic as shown in Table 9.

In 2015, with the same provisions as in Table 9, Sleman regency had a stratification class which 14 districts belong to K1 stratification class, 3 districts belong to K2 with ratio of 82.36 % endemic and 17.64 % sporadic as shown in Table 10.

In 2016, Sleman regency had a stratification class which 16 districts belong to K1 stratification class, 1 district belong to K2 with ratio of 94.2 % endemic and 5.68 % sporadic as shown in Table 11.

Forecasting result of the DHF and ABJ cases number in 2016, an area stratification rule table was made in the order of the table column, namely the number of DHF cases in the 1st, 2nd, 3rd year and the average percentage of ABJ in the 3rd year.

6.2 Evaluation of forecasting models

Evaluation of the results of DBD forecasting with SARIMA method in 17 sub-districts in Sleman-Yogyakarta-Indonesia District using relative mean absolute error (rMAE). The resulting error value has a relatively small value with an average of less than 5% for each region.

Table 9. Area stratification class based on forecasting value in Sleman regency in the year of 2014

No	District	DHF In Year Of			ABJ	Class
		2012	2013	2014		
1	Gamping	60	128	106	87.1	K1
2	Godean	49	114	85	86.8	K1
3	Moyudan	3	1	4	75.5	K1
4	Minggir	4	8	1	80.5	K1
5	Seyegan	1	9	12	94.4	K1
6	Mlati	18	130	72	86.0	K1
7	Depok	10	82	80	89.5	K1
8	Prambanan	9	32	29	79.7	K1
9	Berbah	7	40	46	90.2	K1
10	Kalasan	27	72	64	90.2	K1
11	Ngemplak	6	19	13	95.1	K1
12	Ngaglik	17	54	51	85.5	K1
13	Sleman	15	31	32	90.7	K1
14	Tempel	1	6	6	95.0	K1
15	Turi	1	3	0	91.4	K2
16	Pakem	2	1	4	95.8	K1
17	Cangkringan	6	6	10	85.6	K1

Table 10. Area stratification class based on forecasting value in Sleman regency in the year of 2015

No	District	DHF In Year Of			ABJ	Class
		2013	2014	2015		
1	Gamping	128	106	57	87.6	K1
2	Godean	114	85	77	87.1	K1
3	Moyudan	1	4	10	72.2	K1
4	Minggir	8	1	4	83.2	K1
5	Seyegan	9	12	23	83.8	K1
6	Mlati	130	72	39	83.2	K1
7	Depok	82	80	88	86.4	K1
8	Prambanan	32	29	20	95.2	K1
9	Berbah	40	46	51	90.7	K1
10	Kalasan	72	64	63	91.3	K1
11	Ngemplak	19	13	13	95.3	K1
12	Ngaglik	54	51	47	83.8	K1
13	Sleman	31	32	25	92.3	K1
14	Tempel	6	6	3	95.3	K1
15	Turi	3	0	0	92.5	K2
16	Pakem	1	4	0	96.6	K2
17	Cangkringan	6	10	0	84.9	K2

Table 11. Area stratification class based on forecasting value in Sleman regency in the year of 2016

No	District	DHF In Year Of			ABJ	Class
		2014	2015	2016		
1	Gamping	60	128	106	87.1	K1
2	Godean	49	114	85	86.8	K1
3	Moyudan	3	1	4	75.5	K1
4	Minggir	4	8	1	80.5	K1
5	Seyegan	1	9	12	94.4	K1
6	Mlati	18	130	72	86.0	K1
7	Depok	10	82	80	89.5	K1
8	Prambanan	9	32	29	79.7	K1
9	Berbah	7	40	46	90.2	K1
10	Kalasan	27	72	64	90.2	K1
11	Ngemplak	6	19	13	95.1	K1
12	Ngaglik	17	54	51	85.5	K1
13	Sleman	15	31	32	90.7	K1
14	Tempel	1	6	6	95.0	K1
15	Turi	1	3	0	91.4	K2
16	Pakem	2	1	4	95.8	K1
17	Cangkringan	6	6	10	85.6	K1

The result of forecasting value of the number of DHF cases using SARIMA and Vector Autoregressive method for the rMAE value is shown in Table 12. The comparison of an error value in Table 12 the proposed model has a lower error value compared with Vector Autoregressive method.

Evaluation fitting results the number of DHF cases in 17 districts using the SARIMA has lower error compare with VAR method, for the average relative mean absolute error (rMAE) value of 0.1460 and 0.2508.

6.3 Prediction of area class pattern using neural networks

The neural network model prediction was used to obtain area classification pattern learning in each area. Class prediction input was the number of DHF cases in 1st, 2nd, and 3rd year, and larva-free number (ABJ) in the third year and area classes, while the target output was the area stratification class. Tables 13, 14 and 15 show the prediction result using the neural network.

Table 12. Evaluation forecasting result relative mean absolute error value SARIMA and VAR method

No	District	*SARIMA*	*VAR*
1	Gamping	0.0181	0.0663
2	Godean	0.3700	0.2028
3	Moyudan	0.2350	0.8860
4	Minggir	0.0970	0.1944
5	Seyegan	0.3472	0.3333
6	Mlati	0.0232	0.5775
7	Depok	0.0544	0.0951
8	Prambanan	0.1236	0.1389
9	Berbah	0.0893	0.3849
10	Kalasan	0.1274	0.0703
11	Ngemplak	0.4229	0.5194
12	Ngaglik	0.0317	0.3095
13	Sleman	0.0071	0.0477
14	Tempel	0.2569	0.2569
15	Turi	0.0001	0.0000
16	Pakem	0.0417	0.0417
17	Cangkringan	0.2639	0.1389
	Average	0.1460	0.2508

Table 13. Area classes of DHF cases data forecasting using neural network in year of 2014

District	X1	X2	X3	X4	Y	Pred_Out_1
1	0.415	0.255	0.300	0.7347	0.4	0.3559036
2	0.320	0.125	0.245	0.7077	0.4	0.4055714
3	0.035	0.000	0.015	0.3979	0.8	0.6421382
4	0.025	0.005	0.020	0.7033	0.4	0.5744711
5	0.105	0.025	0.005	0.7394	0.4	0.5418918
6	0.255	0.120	0.090	0.7323	0.4	0.4630306
7	0.440	0.105	0.050	0.7566	0.4	0.4124174
8	0.040	0.005	0.045	0.6304	0.4	0.5776527
9	0.155	0.005	0.035	0.7327	0.4	0.5176745
10	0.690	0.080	0.135	0.7609	0.4	0.3276753
11	0.125	0.015	0.030	0.7871	0.4	0.5176513
12	0.195	0.035	0.085	0.7515	0.4	0.4840204
13	0.170	0.055	0.075	0.7257	0.4	0.4999551
14	0.010	0.000	0.005	0.7967	0.8	0.5644953
15	0.010	0.000	0.005	0.7660	0.8	0.5710398
16	0.005	0.000	0.010	0.7794	0.8	0.5684305
17	0.020	0.000	0.030	0.7257	0.8	0.5685284

Table 13 is the result of normalization of data the number of DHF case in 2014 such as shown in Table 9. The data are inputted to neural network process and the output of neural network shown as predicted output (Pred_Out_1).

Table 14 is the result of normalization of data the number of DHF case in 2015 such as shown in Table 10. The data are inputted to neural network process and the output of neural network shown as predicted output (Pred_Out_1).

Table 15 is the result of normalization of data the number of DHF case in 2015 such as shown in Table 10. The data are inputted to neural network process and the output of neural network shown as predicted output (Pred_Out_1). Process training neural network is carried out ten times with the number of hidden units of 5, 10, 20, 30 and 45, respectively. In the information, the table above column Y is the target class that is used as a goal of the class of stratification (0.4 = endemic, 0.8 = sporadic), while the column Pred_Out_1 is a column for class prediction.

Table 13, 14 and 15 is test-1 training for example in the search for regional stratification class area. Whereas for testing the evaluation results from test-1 to test-10 are presented in Table 16, namely the presentation of the MSE value of the early identification model for stratification class area. A description of the column is in the following information:

Table 14. Area classes of DHF cases data forecasting using neural network in year of 2015

District	X1	X2	X3	X4	Y	Pred_Out_1
1	0.255	0.300	0.640	0.869	0.4	0.3032914
2	0.125	0.245	0.570	0.851	0.4	0.3546066
3	0.000	0.015	0.005	0.469	0.8	0.6403346
4	0.005	0.020	0.040	0.811	0.4	0.5516995
5	0.025	0.005	0.045	0.922	0.4	0.5210359
6	0.120	0.090	0.650	0.862	0.4	0.3430622
7	0.105	0.050	0.410	0.890	0.4	0.3979291
8	0.005	0.045	0.160	0.789	0.4	0.5190650
9	0.005	0.035	0.200	0.889	0.4	0.4874211
10	0.080	0.135	0.360	0.906	0.4	0.4118421
11	0.015	0.030	0.095	0.943	0.4	0.5043691
12	0.035	0.085	0.270	0.877	0.4	0.4576891
13	0.055	0.075	0.155	0.873	0.4	0.4848140
14	0.000	0.005	0.030	0.945	0.8	0.5298144
15	0.000	0.005	0.015	0.910	0.8	0.5413521
16	0.000	0.010	0.005	0.929	0.8	0.5403610
17	0.000	0.030	0.030	0.852	0.8	0.5475355

Table 15. Area classes of DHF cases data forecasting using neural network in year of 2016

District	X1	X2	X3	X4	Y	Pred. Out 1
1	0.300	0.640	0.530	0.871	0.4	0.3025537
2	0.245	0.570	0.425	0.868	0.4	0.3399673
3	0.015	0.005	0.020	0.755	0.4	0.5668112
4	0.020	0.040	0.005	0.805	0.4	0.5574203
5	0.005	0.045	0.060	0.944	0.4	0.5174600
6	0.090	0.650	0.360	0.860	0.4	0.3978530
7	0.050	0.410	0.400	0.895	0.4	0.4023173
8	0.045	0.160	0.145	0.797	0.4	0.5029460
9	0.035	0.200	0.230	0.902	0.4	0.4592537
10	0.135	0.360	0.320	0.902	0.4	0.3971803
11	0.030	0.095	0.065	0.951	0.4	0.5036870
12	0.085	0.270	0.255	0.855	0.4	0.4420305
13	0.075	0.155	0.160	0.907	0.4	0.4665863
14	0.005	0.030	0.030	0.950	0.4	0.5259665
15	0.005	0.015	0.000	0.914	0.8	0.5428952
16	0.010	0.005	0.020	0.958	0.4	0.5267437
17	0.030	0.030	0.050	0.856	0.4	0.5300086

Note:

Kec   = The name of District in Sleman Regency
X1   = DHF Prevalence in 1st year Period
X2   = DHF Prevalence in 2nd year Period
X3   = DHF Prevalence in 3rd year Period
X4   = Average of ABJ in 3rd year
Y   = The value of target that represent stratification
            Class
Pred Out 1 = Area Class Prediction of Neural Network

The area stratification class patterns indicated that all 17 districts had endemic tendencies in 2014, 2015 and 2016 periods such as shown in Table 16. The model was able to provide an overview of the potential risk of outbreaks using forecast data input in the last 3 years and the average larva-free number (ABJ) in the 3rd year so that the model was capable of providing information one step ahead in a 12-month period and automatically give recommendations for policy holder about the area stratification of the outbreak prevention.

Results of error measurement training data neural network for early identification area stratification class pattern are shown in Table 16. To predict of class area neural network were trained in different number nodes, and resulted in different mean squared error (MSE) such as shown in Table 16. The smallest MSE was obtained in the architecture of neural network with 20 hidden nodes (test-10) in Table 16.

257

Table 16. Test results of mean squared error early identification area stratification cluster pattern

Training	Hidden Neuron				
	5	10	20	30	45
Test-1	0.685	0.732	0.702	0.742	0.705
Test-2	0.603	0.524	0.562	0.665	0.592
Test-3	0.542	0.884	0.601	0.606	0.526
Test-4	0.495	0.875	0.558	0.517	0.495
Test-5	0.467	0.865	0.518	0.480	0.472
Test-6	0.431	0.856	0.497	0.462	0.457
Test-7	0.447	0.843	0.459	0.497	0.441
Test-8	0.419	0.832	0.430	0.423	0.428
Test-9	0.408	0.820	0.397	0.414	0.413
Test-10	0.398	0.818	0.394	0.404	0.403

Table 17. Calculation of suitability for class stratification

No	Y	Pred. Out 1	MAD	Actual	Predict	Suitable
1	0.4	0.5530	0.1530	Endemic	Endemic	Y
2	0.4	0.5291	0.1291	Endemic	Endemic	Y
..	…..	………	…..	……….	……….	…
..	…..	………	…..	……….	……….	…
49	0.8	0.7475	0.0524	Sporadic	Sporadic	Y
50	0.4	0.7029	0.3029	Endemic	Sporadic	N
51	0.4	0.6033	0.2033	Endemic	Sporadic	N

Table 18. Confusion matrix table

Prediction	Actual	
	Positive	Negative
True	400	27
False	27	56

Evaluation of the stratification class model is done by using the k-fold cross validation method by dividing the dataset into two parts, namely 80% training data, and 20% testing data. Class prediction is made by neural network techniques with the number of neurons hidden 5, 10, 20, 30 and 45, learning rates 0.1, 0.05, and 0.01 for 51 regions in Sleman Regency (17 areas in first year, second year and third year). Datasets used to carry out the test is 51 regions by looking at the suitability of the prediction class with the actual class by calculating the calculation are shown in Table 17.

Afterwards, testing evaluation was done using a confusion matrix to calculate accuracy, precision, and recall. Results of the calculation are presented in Table 18. Calculation of the confusion matrix is described as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \qquad (2)$$

$$= \frac{400+23}{400+23+23+64} \times 100\% = 89,41\%$$

$$Precision = \frac{TP+TN}{TP+FP} \times 100\% \qquad (3)$$
$$= \frac{400+23}{400+63} \times 100\% = 93,6\%$$

Table 19. The highest potential month for prevalence in 2010-2015 and the highest prevalence month based on forecasting result

Year	1st Potention	2nd Potention	3rd Potention
2010	February	January	April
2011	March	May	February
2012	December	June	July
2013	July	January	February
2014	March	February	June
2015	February	May	March
2016	March	February	November

Results of the test calculation using a confusion matrix generated scores for accuracy and precision by 89,41%, and 93,6%, respectively. The result is ideal for a large number of regions to know the pattern of stratification class.

6.4 Prediction of the highest prevalence potential month

The highest prevalence month can be predicted in each district. Data on the number of DHF cases from forecasting form a pattern about which months are having the potential of prevalence with a high number of casualties, and for that, the period of transmission identification, the period of the first transmission and the period of an outbreak will be figured out first. The analysis of the highest potential month for prevalence from 2010-2016 is available in Table 19.

The potential for the occurrence of the month each year has a different pattern from 2010 to 2016 as shown in Table 19, but it can be seen that the

monthly patterns of dengue cases often occur with a high number of cases. A month with the highest potential for dengue cases in February (the first highest month of dengue cases happening six times in the last seven years) and March (the second highest number of months in the past seven years). While other months to watch out for are January, May, and June (the third highest month twice in the last seven years), so that these months need to watch out for indications of an increase in dengue cases.

6.5 Area identification for outbreak areas

Visualization of the area early identification was described in the form of area mapping. The map provided information that the percentage of the endemic area had a more significant portion in each year. For example, in 2014, the comparison was four sporadic areas against 13 endemic regions, in 2015, five sporadic regions against 12 endemic regions, and in 2016, one sporadic area against 16 endemic regions thus it requires further investigation about a potential risk of outbreak transmission to the adjacent regions.

In Sleman Regency, stratification class in 2014 consisted of 76.4% endemic areas while the remaining 23.6% was sporadic areas. In 2015, it included 70.6% endemic areas and 29.4% sporadic areas. In 2016, it consisted of 94.1% endemic areas and 5.9% sporadic areas. In 2014 and 2015, areas in Highlands had three sporadic class regions. It was due to the low air temperature in high altitudes areas resulting in the susceptibility of Ae.agepty mosquito life hence those areas had a high percentage of ABJ, above 95%.
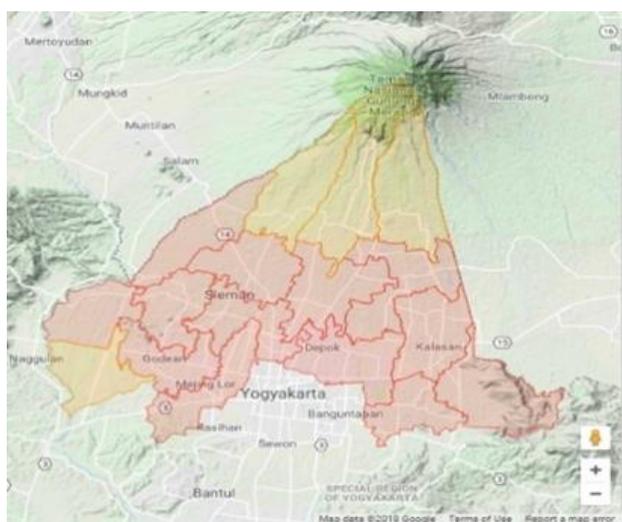

Figure. 3 The map of early identification of area stratification of DHF outbreaks in 2014
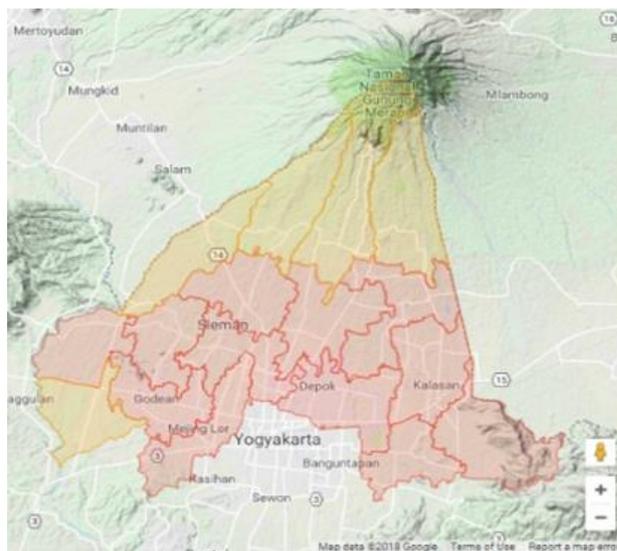

Figure. 4 The map of early identification of area stratification of DHF outbreaks in 2015
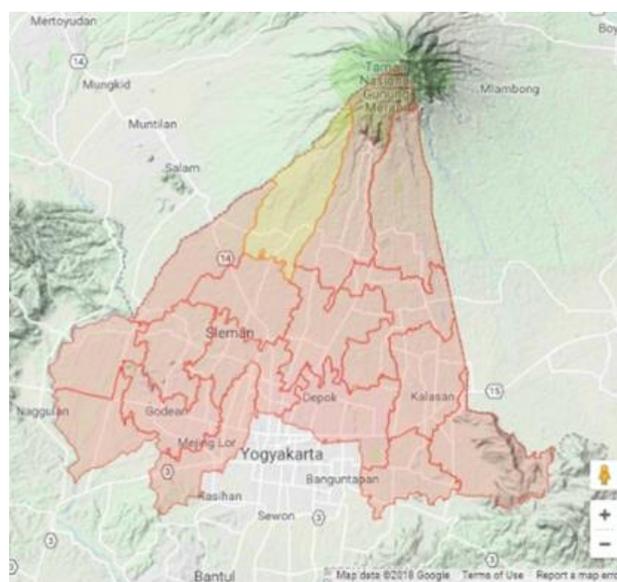

Figure. 5 The map of early identification of area stratification of DHF outbreaks in 2016

7. Conclusion and future work

The early identification model of Dengue Hemorrhagic Fever (DHF) has been made to solve the problem by using several methods. Technically, the solving problem methods in this research are (1) Forecasting the DHF cases, (2) Seeking the class pattern of regional stratification, and (3) Mapping from the forecasting and prediction of area stratification class pattern results.

In this research, seasonal ARIMA appropriate for short term forecasting such as 1, 3 & 6 months. The neural network can be used to find stratification area patterns (endemic, sporadic, potential and free) so that the model can provide temporal and spatial information. Future works of this research is to get an

accurate of map prediction, by reducing the area scope such as villages.

The most significant contribution of the study is development of identification model of dengue outbreak based on rule-based stratification. This model can be predict the dengue outbreak in one step ahead before occurrence.

For the next research, area stratification classes of K1, K2, K3 and K4 are divided into the levels of area risk. The area of potential risk is calculated using spatial autoregressive (SAR), and then categorized as levels high, medium and low in each class, which are visualized in potential risk mapping.

Acknowledgments

References

[1] J. Vontas, H. Ranson, and L. Alphey, "Transcriptomics and disease vector control", *BMC Biol.*, Vol. 8, p. 52, 2010.

[2] S. Naish, P. Dale, J. S. Mackenzie, J. Mcbride, K. Mengersen, and S. Tong, "Spatial and temporal patterns of locally-acquired dengue transmission in Northern Queensland, Australia, 1993-2012", *PLoS One*, Vol. 9, No. 4, pp. 1981–1991, 2014.

[3] D. N. Pham, S. Nellis, A. A. Sadanand, A. Jamil, J. J. Khoo, T. Aziz, 1 D. Lukose, S. bin A. Bakar, and A. Sattar, "A Literature Review of Methods for Dengue Outbreak Prediction", In: *Proc. of eKNOW 2016 Eighth Int. Conf. Information, Process. Knowl. Manag.*, No. c, pp. 7–13, 2016.

[4] M. Andraud, N. Hens, C. Marais, and P. Beutels, "Dynamic epidemiological models for dengue transmission: a systematic review of structural approaches", *PLoS One*, Vol. 7, No. 11, p. e49085, Jan 2012.

[5] A. M. Stewart-ibarra, Á. G. Muñoz, S. J. Ryan, E. B. E. B. Ayala, M. J. Borbor-cordova, M. A. Stewart-Ibarra, G. ï¿½ngel Muï¿½oz, J. S. Ryan, E. B. E. B. Ayala, J. M. Borbor-Cordova, L. J. Finkelstein, R. Mejï¿½a, T. Ordoï¿½ez, G. C. Recalde-Coronel, and K. Rivero, "Spatiotemporal clustering, climate periodicity, and social-ecological risk factors for dengue during an outbreak in Machala, Ecuador, in 2010", *BMC Infect. Dis.*, Vol. 14, No. 1, pp. 1–16, 2014.

[6] S. Sang, W. Yin, P. Bi, H. Zhang, C. Wang, X. Liu, B. Chen, W. Yang, and Q. Liu, "Predicting local dengue transmission in Guangzhou, China, through the influence of imported cases, mosquito density and climate variability", *PLoS One*, Vol. 9, No. 7, p. e102755, 2014.

[7] J. Fan, H. Lin, C. Wang, L. Bai, S. Yang, C. Chu, W. Yang, and Q. Liu, "Identifying the high-risk areas and associated meteorological factors of dengue transmission in Guangdong Province, China from 2005 to 2011", *Epidemiol. Infect.*, Vol. 142, No. 3, pp. 634–43, 2014.

[8] P. Sirisena, F. Noordeen, H. Kurukulasuriya, T. A. Romesh, and L. K. Fernando, "Effect of climatic factors and population density on the distribution of dengue in Sri Lanka: A GIS based evaluation for prediction of outbreaks", *PLoS One*, Vol. 12, No. 1, 2017.

[9] C. T. Tsai, F. C. Sung, P. S. Chen, and S. C. Lin, "Exploring the spatial and temporal relationships between mosquito population dynamics and dengue outbreaks based on climatic factors", *Stoch. Environ. Res. Risk Assess.*, Vol. 26, No. 5, pp. 671–680, 2012.

[10] H. M. Khormi and L. Kumar, "The importance of appropriate temporal and spatial scales for dengue fever control and management", *Sci. Total Environ.*, Vol. 430, pp. 144–9, 2012.

[11] Z. Zhang, *Multivariate Time Series Analysis in Climate and Environmental Research*, 2018.

[12] N. C. Dom and A. H. Ahmad, "Spatial Mapping of Temporal Risk Characteristic of Dengue Cases in Subang Jaya," No. Cssr, pp. 361–366, 2010.

[13] C.-H. Chiu, T.-H. Wen, L.-C. Chien, and H.-L. Yu, "A probabilistic spatial dengue fever risk assessment by a threshold-based-quantile regression method", *PLoS One*, Vol. 9, No. 10, pp. e106334, 2014.

[14] A. Redaksi, "Buletin Jendela Epidemiologi, Volume 2 , Agustus 2010", *Achmadi Umar Fahmi*, Vol. 2, p. 48, 2010.

[15] V. R. Louis, R. Phalkey, O. Horstick, P. Ratanawong, A. Wilder-Smith, Y. Tozan, and P. Dambach, "Modeling tools for dengue risk mapping-a systematic review", *Int. J. Health Geogr.*, Vol. 13, No. 1, p. 50, 2014.

[16] V. Racloz, R. Ramsey, S. Tong, and W. Hu, "Surveillance of dengue fever virus: a review of epidemiological models and early warning systems", *PLoS Negl. Trop. Dis.*, Vol. 6, No. 5, p. e1648, 2012.

[17] Y. L. Cheong, P. J. Leitão, and T. Lakes, "Assessment of land use factors associated with dengue cases in Malaysia using Boosted

Regression Trees", *Spat. Spatiotemporal. Epidemiol.*, Vol. 10, pp. 75–84, 2014.

[18] I. K. Mulyawan, "Pola sebaran dan faktor risiko kejadian dbd di kota kendari tahun 2010", pp. 1–211, 2011.

[19] X. Porcasi, C. H. Rotela, M. V. Introini, N. Frutos, S. Lanfri, G. Peralta, E. A. De Elia, M. A. Lanfri, and C. M. Scavuzzo, "An operative dengue risk stratification system in Argentina based on geospatial technology", *Geospat. Health*, Vol. 6, No. 3 SUPPL., 2012.

[20] D. A. N. P. Lingkungan, Modul Pengendalian Demam Berdarah Dengu*e*. 2011.

[21] Y. L. Hii, H. Zhu, N. Ng, L. C. Ng, and J. Rocklöv, "Forecast of Dengue Incidence Using Temperature and Rainfall", *PLoS Negl. Trop. Dis.*, Vol. 6, No. 11, 2012.

[22] D. J. Rogers, J. E. Suk, and J. C. Semenza, "Using global maps to predict the risk of dengue in Europe", *Acta Trop.*, Vol. 129, pp. 1–14, 2014.

[23] B. Maximiliano and G. Gonz, "Predictive model of dengue focus applied to Geographic Information Systems", 2015.

[24] Y. Song, F. Wang, B. Wang, S. Tao, H. Zhang, S. Liu, O. Ramirez, and Q. Zeng, "Time series analyses of hand, foot and mouth disease integrating weather variables", *PLoS One*, Vol. 10, No. 3, pp. 1–18, 2015.

[25] Ö. Ö. Bozkurt, G. Biricik, and Z. C. Tayşi, "Artificial neural network and SARIMA based models for power load forecasting in Turkish electricity market", *PLoS One*, Vol. 12, No. 4, p. e0175915, 2017.

[26] I. N. Soyiri and D. D. Reidpath, "Evolving forecasting classifications and applications in health forecasting", *Int. J. Gen. Med.*, Vol. 5, pp. 381–389, 2012.

[27] H. Halide, P. Ridd, N. Sciences, P. Sciences, and J. C. Universty, "Early Warning System for Dengue Hemorrhagic Fever (DHF) Epidemics in Makassar", pp. 26–34, 2011.

[28] Y. Yusof and Z. Mustaffa, "Dengue Outbreak Prediction : A Least Squares Support Vector Machines Approach", Vol. 3, No. 4, pp. 489–493, 2011.

[29] N. Degallier, C. Favier, C. Menkes, M. Lengaigne, W. M. Ramalho, R. Souza, and J. S. J. Boulanger, "Toward an early warning system for dengue prevention : modeling climate impact on dengue transmission", pp. 581–592, 2010.

[30] R. Lowe, C. Barcellos, C. S. Coelho, T. C. Bailey, G. E. Coelho, R. Graham, T. Jupp, W. M. Ramalho, M. S. Carvalho, D. B. Stephenson, and X. Rodó, "Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts", *Lancet. Infect. Dis.*, Vol. 14, No. 7, pp. 619–26, 2014.

[31] R. A. Morello-frosch and S. L. Syme, "Advancing Health Equity and Climate Change Solutions in California Through Integration of Public Health in Regional Planning By Solange M . Gould A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Public Healt", 2015.

[32] P. Guo, T. Liu, Q. Zhang, L. Wang, J. Xiao, Q. Zhang, G. Luo, Z. Li, J. He, Y. Zhang, and W. Ma, "Developing a dengue forecast model using machine learning: A case study in China", *PLoS Negl. Trop. Dis.*, Vol. 11, No. 10, pp. 1–22, 2017.

[33] F. Schaffner and A. Mathis, "Dengue and dengue vectors in the WHO European region: past, present, and scenarios for the future", *Lancet. Infect. Dis.*, Vol. 3099, No. 14, pp. 1–10, 2014.

[34] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*, 2011.

[35] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomputing*, Vol. 50, pp. 159–175, 2003.