ASSESSMENT AND COMPARISON OF MACHINE LEARNING ALGORITHM CAPABILITY IN SPATIAL MODELING OF DENGUE FEVER VULNERABILITY BASED ON LANDSAT IMAGE 8 OLI/TIRS

Rahmat Azul Mizan[1], Prima Widayani[2], Nur Mohammad Farda[3]
Program Studi Magister Penginderaan Jauh, Departemen Sains Informasi Geografi,
Fakultas Geografi, Universitas Gadjah Mada
Jl. Kaliurang, Sekip Utara, Bulaksumur, Sleman, Yogyakarta, 55281, Indonesia
e-mail: primawidayani@ugm.ac.id

Abstrak

Penyebarluasan penyakit demam berdarah di Indonesia telah menjadi masalah utama bidang kesehatan. Pemodelan spasial distribusi kerentanan penyakit demam berdarah merupakan langkah yang penting untuk mendukung penyusunan perencanaan dan penanggulangan penyakit demam berdarah di Indonesia. Penelitian ini bertujuan menilai dan membandingkan kemampuan dua algoritma machine learning dalam membuat model spasial kerentanan demam berdarah. Penelitian dilakukan di Kota Baubau Provinsi Sulawesi Tenggara dengan mengambil 129 kasus kejadian selama periode 2015 hingga Februari 2016. Dalam penelitian ini, model dibuat menggunakan software R dan algoritma machine learning mencakup support vector machine (SVM) dan random forest (RF). Enam variabel pemodelan terlibat mencakup penggunaan/tutupan lahan, BLFEI, NDVI, LST, curah hujan dan kelembapan udara yang diekstrak dari citra Landsat 8 OLI/TIRS serta data iklim BMKG dan BWS. Kemampuan model dinilai menggunakan kurva Area Under Curve-Receiver Operating Characteristic (AUC-ROC). Hasil penilitian menunjukan informasi bahwa kedua algoritma memberikan akurasi model yang sangat baik dengan nilai AUC 1 untuk SVM dan 0,997 untuk RF dengan SVM sebagai algoritma yang dipandang paling baik memodelkan demam berdarah di Kota Baubau.

Kata Kunci: *Machine Learning, Kerentanan, Demam Berdarah, Citra Landsat 8*

Abstract

The spread of dengue fever in Indonesia has become a major health problem. Spatial modeling for the distribution of dengue fever vulnerability is an important step to support the planning and mitigation of dengue fever in Indonesia. This study aims to assess and compare the capability of two machine learning algorithms to create a spatial model of dengue fever vulnerability. The research was conducted in Baubau City, Southeast Sulawesi Province by taking 129 cases that occurred from 2015 to February 2016. In this study, the model was created using R software and machine learning algorithms including support vector machine (SVM) and random forest (RF). The six modeling variables involved include land use/cover, BLFEI, NDVI, LST, rainfall and humidity extracted from Landsat 8 OLI/TIRS imagery as well as BMKG (Meteorological, Climatological, and Geophysical Agency of Indonesia) and BWS climate data. The model's capability was assessed using the Area Under Curve-Receiver Operating Characteristic (AUC-ROC) curve. The results of the research show that both algorithms provide excellent model accuracy with AUC values of 1 for SVM and

0.997 for RF with SVM as the best algorithm for modeling dengue fever in Baubau City.

INTRODUCTION

Various serious problems are being faced by the world today, where climate change ranks first of the ten problems (World Economic Forum, 2017). Furthermore, climate change plays a role in the transmission and expansion of various infectious diseases spread by mosquito species such as Aedes aegypti and Aedes albopictus which are known as dengue hemorrhagic fever (Lee et al., 2018).

DHF has become a serious problem because it causes more than half of the world's population to be hospitalized and is responsible for death (Gubler, 2012). (WHO, 2019) stated that DHF has become the main arbovirus in the world with an increase in the number of sufferers reaching three times in the last five decades. Another worrying fact is that the worst dengue cases occurred in Southeast Asia and the Western Pacific with an increase in cases reaching 2.2 million from 2008 to 2010 (WHO, 2019).

Indonesia was one of the endemic areas with the third largest incidence rate in Southeast Asia and the Western Pacific (Ministry of Health, 2017). Based on the report (Ministry of Health of the Republic of Indonesia, 2019) as of February 2016, dengue fever caused 169 deaths from 16,692 new cases.

The implementation of handling efforts that have not been maximized (WHO, 2007) and the absence of vaccines or appropriate control efforts (Wilder-Smith et al., 2019) have made the spread and outbreak of DHF continue to increase, especially in tropical areas including Indonesia (WHO, 2019). Therefore, prevention efforts are considered as the main step in reducing the incidence.

One way of preventions carried out in the spatial field is spatial modeling of regional vulnerabilities, both based on physical environmental conditions (Ding et al., 2018) and socio-economic (Yue et al.,

2018) or a combination of both (Adzan & Danoedoro, 2012); Widayani, 2010).

Spatial modeling is an activity to simplify complex spatial phenomena in the real world with the aim of understanding how spatial phenomena are formed. Furthermore, the best accuracy of DHF modeling can vary from one region to another, depending on several factors, one of which is the complexity of the modeling method (Pineda-Cortel et al., 2019).

The presence of remote sensing technology is very helpful in spatial modeling because it can overcome the problem of availability of physical environmental data which is generally difficult to obtain. Landsat 8 OLI/TIRS imagery is one of the remote sensing products that is widely used in natural resource management to the field of public health (NASA, 2019).

Along with the increase in spatial data collection tools and sensors, the term geospatial big data has emerged which is the forerunner of the field of artificial intelligent geospatial science (geoAI). GeoAI itself is a combination of spatial science, machine learning, data mining, and advanced computational methods to reveal more meaningful information from geospatial big data.

In line with that (Dempsey, 2012) states that the increasing number and variety of data formats from geospatial big data collected today presents its own challenges in terms of storing, managing, processing, analyzing, visualizing, and checking data quality. Thus, the presence of machine learning methods in spatial data analysis is a challenge as well as a new opportunity in spatial modeling of dengue fever

As it is known that there is no single modeling method that is suitable for all regions, it is necessary to deepen the method that is seen as emerging as a novelty in spatial modeling. In terms of the methods used, various spatial modeling methods have now been applied and one of them is machine learning. Therefore, the presence of machine learning, especially the support vector machine (SVM) and random forest (RF) algorithms which are present as a novelty of analytical methods in spatial modeling, especially in geospatial big data, need to be assessed for their capability to model dengue fever.

Various previous machine learning studies using SVM, and RF algorithms are generally found in disaster modeling such as landslides (Motevalli et al., 2018), forest fires (Ngoc Thach et al., 2018) and have not been widely carried out in the field of disease epidemiology, especially dengue fever. (Ding et al., 2018). Apart from not being explored much, previous studies on spatial modeling of DHF also only compared the capabilities between algorithms on a global scale. Therefore, we need a study of spatial modeling of DHF on a local scale that focuses on the assessment and comparison of model capabilities.

This study aims to assess and compare the capability of two machine learning algorithms to produce a spatial model of regional vulnerability to dengue fever in Baubau City, Southeast Sulawesi Province (Figure 1). The selection of research locations is based on the incident rate (IR) value where this number is used to assess the high and low of an area affected by DHF. Based on the Indonesian Ministry of Health (2018), Southeast Sulawesi was the province with the third highest IR rate nationally. Moreover, according to data from the Southeast Sulawesi Provincial Health Office (2018), Baubau City was designated as one of the areas of extraordinary occurrence (KLB) of DHF with the third largest incidence of 17 regencies/cities in Southeast Sulawesi. Administratively, the research area covers the Districts of Sorawolio, Bungi, Lea-Lea, Wolio, Betoambari, Kokalukuna, Murhum and Batupuaro Districts (Statistics Indonesia of Baubau City, 2018).
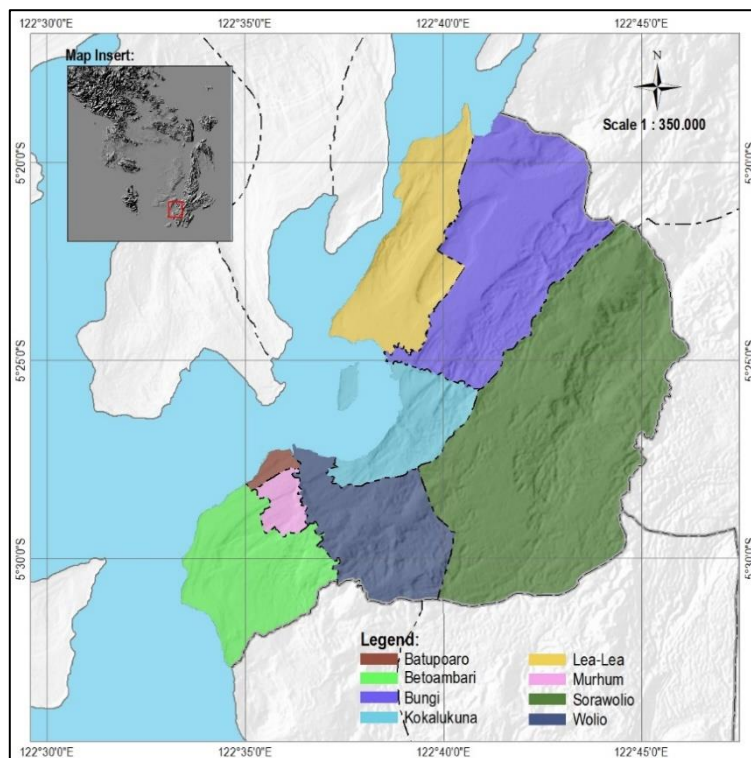

Figure 1. Research Location

RESEARCH METHODS

1. Tools and Materials

The tools and materials used to carry out this research are described in Tables 1 and 2 below:

Table 1. Research Tools

Tools	Use
GPS	Marking the location of DHF patients.
Software ArcGIS/QGIS	A spatial analysis tool to generate research variables and visualize models in the form of digital maps.
Software ENVI	A digital image processing tool to generate research variables.
Software R	A tool to carry out the classification process using machine learning.

Table 2. Research Materials

Data and Material	Source	Function
Rainfall and humidity	Meteorological, Climatological, and Geophysical Agency and BWS	to create a rainfall and humidity map
DHF sufferers	Baubau City Health Office	to create training data and model validation
Climate observation station location	Meteorological, Climatological, and Geophysical Agency and BWS	to make reference point data for interpolation of rainfall data
Landsat 8 OLI/TIRS Citra image	*USGS Earthexplorer*	for NDVI, BLFEI, ESG and land cover index maps

2. Extraction of Modeling Variables

The involvement of spatial variables was taken through an in-depth literature search to explore how closely the relationship between the incidence of dengue fever was and the variables to be used. The spatial variables in question include land cover/use (Tiong et al., 2015; Vanwambeke et al., 2007), surface temperature (Kamimurai et al., 2002), humidity (Costa et al., 2010), rainfall (Iriani, 2012; Valdez et al., 2018), built-up land index (BLFEI) and vegetation density index (NDVI) (Palaniyandi, 2014; Widayani et al., 2018).

2.1. Classification of Land Cover and Use

Classification of land cover/use referred to Danoedoro (2012) with regard to three main factors that affect the quality of the classification including classification schemes, sample criteria and algorithms. The classification scheme used was SNI-7645-1: 2014 regarding land cover classification-part 1: small and medium scale (BSN, 2014) with a maximum likelihood classification algorithm. This algorithm requires that the sample distribution is evenly distributed over the image coverage with an area of at least 10-40 pixels.

2.2. Vegetation Index (NDVI)

The calculation of the vegetation density index refers to the equation used by (Ganie & Nusrath, 2016) in their research.

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \qquad (1)$$

In which NIR and RED were near infrared channels (band 5) and red channels (band 4) Landsat 8 OLI/TIRS images with wavelengths of 0.851 – 0.879μm and 0.636 – 0.673μm, respectively.

2.3 Built-up Land Index (BLFEI)

Meanwhile, the appearance of built-up land and open land and water bodies was calculated using the building

JURNAL ge grafi

density index equation proposed by Bouhennache, et al. (2018).

$$BLFEI = \frac{\frac{OLI3 + OLI4 + OLI7}{3} - OLI6}{\frac{OLI3 + OLI4 + OLI7}{3} + OLI7} \quad (2)$$

With $OLI3$, $OLI4$, $OLI6$, and $OLI7$ were green channel (band 3= 0.533 – 0.590μm), red channel (band 4 = 0.636 – 0.673μm), middle infrared channel I (band 6 = 1.566 – 1.651μm), and middle infrared channel II (band 7= 2.107 – 2.294μm) Landsat 8 OLI/TIRS image.

2.4. Surface Temperature (LST)

The surface temperature was calculated referring to the equation issued in the Landsat 8 handbook (USGS, 2019).

$$T_{rad} = \frac{K_2}{ln\left(\frac{K_1}{L_\lambda} + 1\right)} - 273,15 \quad (3)$$

In which K_1 and K_2 are spectral radian calibration constants and band 10 absolute temperature calibration constants, L_λ is TOA spectral radians or corrected spectral radian value in units (Watts/(m2*srad*μm)). This corrected spectral radian was obtained from the calibration of the radian band 10 value using the object emissivity value (ε_λ) which was calculated using the equation (Coll et al., 2010) in (Fawzi, 2014) obtained through the http://atmcorr.gsfc.nasa.gov/.

2.5. Interpolation of Rainfall and Humidity Data

Rainfall and humidity are variables obtained through field measurements by Meteorological, Climatological, and Geophysical Agency and BWS climate observation stations. Data from these two agencies was in the form of daily climate data which was tabulated into annual averages and then transformed into spatial data by interpolation method using ArcGIS 10.6 software. The interpolation method chosen was the inverse distance weight (IDW) in the study (Kurniadi et al., 2018).

$$Z_0 = \frac{\sum_{i=1}^{s} Z_i \frac{1}{d_i^k}}{\sum_{i=1}^{s} \frac{1}{d_i^k}} \quad (4)$$

In which Z_0 is the approximate value at point 0; Z_i is the value of z at control point i; d_i is the distance between point i and point 0 and k is the influence of neighboring points and s is the number of station points used.

3. Data Analysis Techniques

Vulnerability classification based on machine learning algorithms was carried out using software R. R is a statistical and graphic computing program which in its development has been known as one of the most reliable software in data analysis and data science (R Foundation, 2020). Its reliability is due to its open-source nature so that it allows the source code to be checked, modified, added and shared by anyone so that the distribution of the package continues to grow following various data science analysis needs.

Package is a collection of scripts which are generally in the form of functions or data that can be used for certain needs (Hidayatuloh, 2020). Various packages were involved in this research including *the raster, rgdal, rgeos, ggplot2, RSNNS, e1071 package* and *the kernlab package* (Hijmans & Elith, 2019).

The dengue fever modeling in this study is aimed at assessing and comparing the capabilities of two machine learning algorithms that are known to have been widely used and relatively strong in various spatial models including support vector machine (SVM) and random forest (RF) algorithms.

3.1. SVM Algorithm

SVM was first introduced by Vapnik (1995) in (Motevalli et al., 2018) as a statistical learning theory. In its development, SVM has been widely used in various environmental fields and natural disaster studies such as (Motevalli et al., 2018; Nguyen et al., 2018) and even health studies, especially those related to

the environment (Ding et al., 2018). The advantage of the SVM method is its good ability to classify even with limited training data (Mountrakis et al., 2011).

Taking into account the training dataset, $TD = (x_i, y_i)_i = 1^N$ in which $x_i \in R^n$ is the input variable including land cover/use, NDVI, BLFEI, LST, annual rainfall and humidity. N is the number of samples from the training data, n is the dimension of the training data, $y_i \in \{-1, +1\}$ describes the label class. Then the best decision function can be formed using equation (5).

$$f(x) = sign \left[\sum y_i \alpha_i K(x_i, x_j) + b \right] \quad (5)$$

In which b is the offset value; α_i is a lagrange; and $K(x_i, x_j)$ are kernel functions. The ability of the SVM model depends on the kernel function, in this study was the radial basis function (RBF).

Since RBF was chosen as a kernel function, the ability of SVM is controlled by two main parameters, namely kernel regularization (C) and kernel width (γ) which must be chosen correctly in order to provide maximum SVM algorithm capabilities. The selection method was done by trial and error based on the range of values that had been used by previous researchers such as and that was done using the grid search method.

In terms of trial and error, the value range proposed by (Ding et al., 2018) was six kernel regularization values (C), namely 2, 4, 8, 16, 24 and 32, while the value of the kernel width (γ) being tested consists of from 0.01 to 0.08. There were 48 combinations of values that would be tested based on the two sets of values proposed. The values proposed by (Nguyen et al., 2018) were 44 for regularization and 0.075 for kernel width.

3.2. RF Algorithm

Random forest (RF) is one of the most popular techniques in machine learning and is widely used for classification and regression of factor inputs (Breiman, 2001). The advantage of this method is that it is able to provide very high classification accuracy and a faster computational process, thus making RF receive more attention in various fields, one of which is remote sensing (Yue et al., 2018).

Conceptually, RF is a learning approach that is built to make predictions based on a set of decision tree classifiers. Based on this concept, several parts of the data set are generated randomly through alternations derived from the training data set. Furthermore, each part of the data set is used to build a decision tree using CART (classification and regression tree) (Breiman, 2001). Based on the above concept, the RF capability is influenced by two parameters, namely the number of decision trees (*N-tree*) and the number of modeling variables (*N-fact*) that must be selected and determined correctly. The method of determining the value of the two parameters was done by trial and error from the values proposed by previous researchers.

In terms of the number of variables involved (*N-fact*), this study used six input variables which were also used when modeling dengue fever using the SVM algorithm. The six variables consisted of land use/cover, NDVI, BLFEI, LST, rainfall and humidity. Meanwhile for the number of decision trees (*N-tree*), several values had been proposed by previous researchers such as (Lawrence et al., 2006; Stevens et al., 2015) who proposed 500 *N-tree*, other researchers such as (Motevalli et al., 2018) found that the value of 1,000 for the *N-tree* and 3 for the m_{try} parameter gave maximum results for landslide modeling, also (Ding et al., 2018) stated that 500 was the best value for the *N-tree* for modeling the distribution of *Ae. Aegypti* dan *Ae. Albopictus* mosquitos. The m_{try} parameter is used to control the number of variables that are randomly sampled as candidates for each variable separation in forming a decision tree model (Liaw & Wiener, 2002). The range of mtry parameter values to be tested in this study was 1 to 8.

4. Assessment and Comparison of Model Capability

The ability of the dengue fever vulnerability model generated from the two algorithms was assessed using the ROC-AUC (area under curve-receiver operating characteristic) curve. Where the qualitative relationship between ROC-AUC and model capability is 0.9-1 (perfect), 0.8-0.9 (very good), 0.7-0.8 (good), 0.6-0.7 (moderate), and 0.5–0.6 (poor) (Swets, 1988).

5. Research Flowchart

Figure 2 would explain step by step how this research was carried out. Broadly speaking, the research phase was grouped into three main parts, namely (1) satellite image processing and spatial data analysis to produce research variables, (2) the process of training algorithms based on historical data and research variables, and (3) the process of testing the model that had been obtained based on DHF history data.
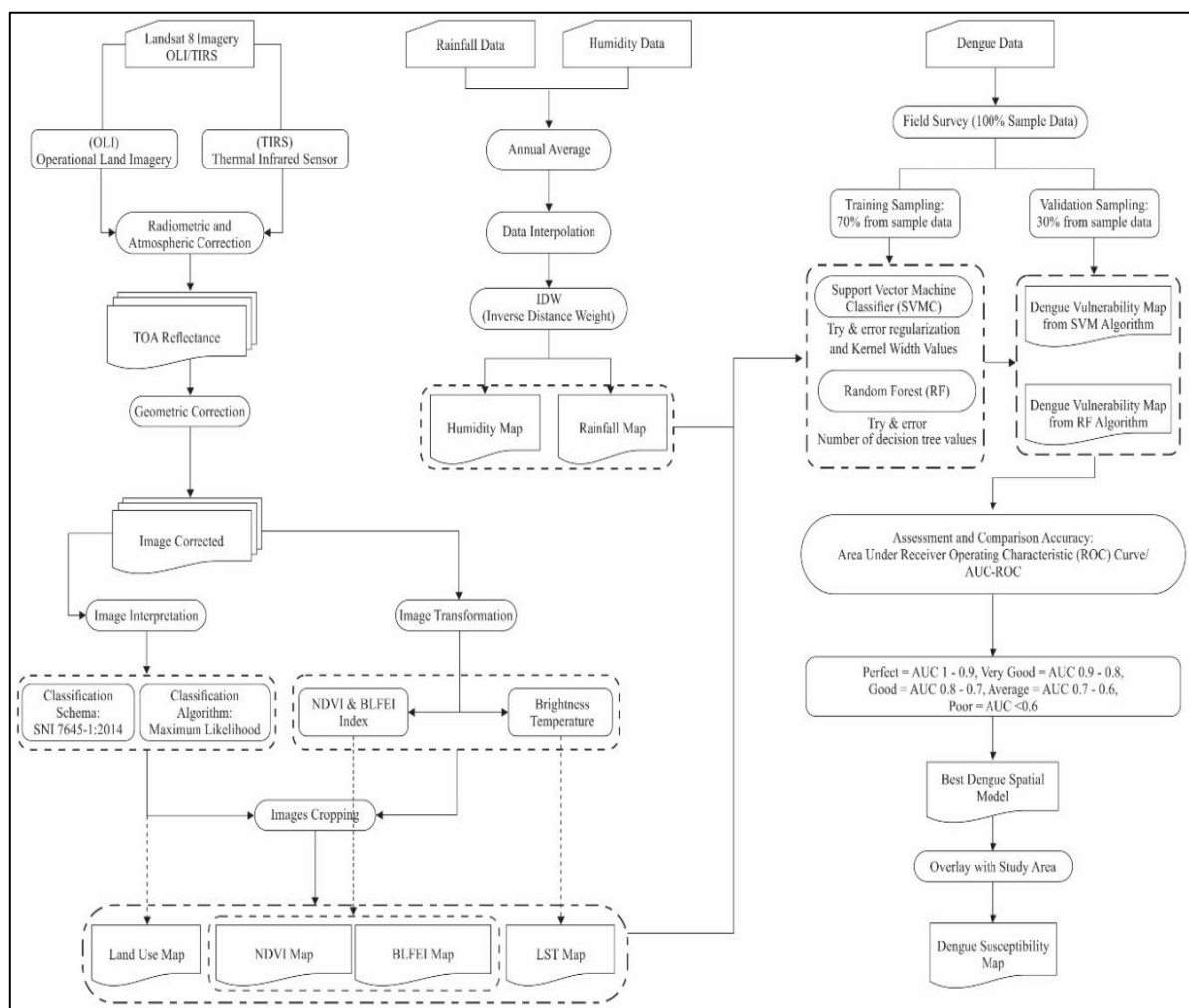

Figure 2. Research Flowchart

RESULTS AND DISCUSSION

The classification results show as many as 13 types of land cover/use at the study site (Figure 4a). Based on the classification scheme used, the 13 types of land cover/use can be grouped into built - up area, cultivated open land and hardened surface, forest and natural/semi-natural vegetation, permanent cultivated vegetation, and bare land (3).

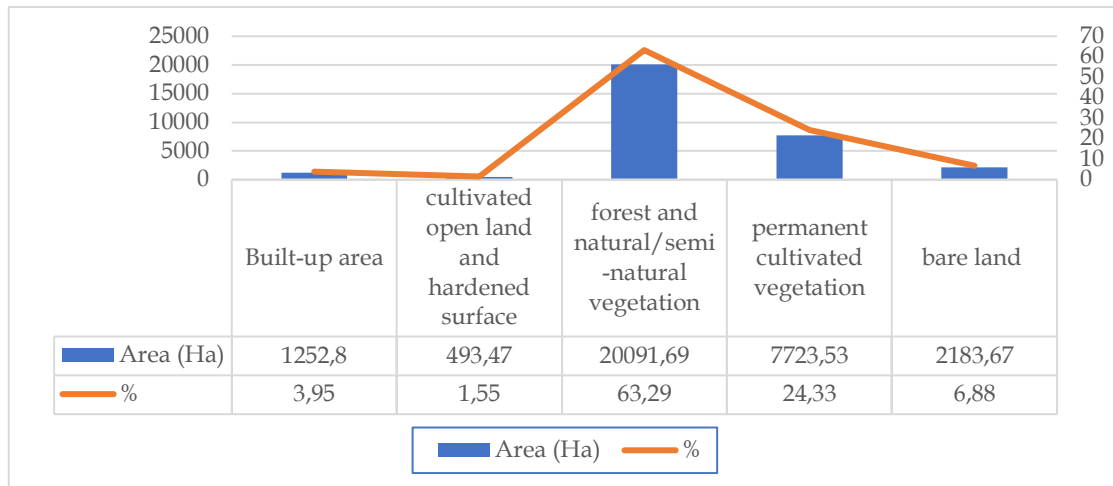|	Built-up area	cultivated open land and hardened surface	forest and natural/semi-natural vegetation	permanent cultivated vegetation	bare land
Area (Ha)	1252,8	493,47	20091,69	7723,53	2183,67
%	3,95	1,55	63,29	24,33	6,88

Figure 3. Area of Land Cover/Use.

The research location is still dominated by forest land cover and natural vegetation of 63.29% while the use of building land is only 3.95% of the total area.

The NDVI value of the research location ranges from 0.66 to 0.86 with a value of 0.75-0.81 being the value with the highest number of pixels (Figures 3a and 4b). The NDVI value of 0.7-0.8 is the value with the most pixels due to the land cover of the research location being dominated by forest land cover and natural/semi-natural vegetation and the permanent cultivation vegetated land cover of 87.62% of the total area (3).

Based on the processing results (Figure 4b), it is known that the BLFEI pixel value of the research location ranged from -1.05 to 0.84 as shown in Figure 5b. From the histogram, it is also known that the highest number of pixel frequencies ranged from -0.9 to -0.6. Referring to the histogram generated by (Bouhennache et al., 2018), an initial conclusion can be drawn that the land cover of the study site was dominated by vegetation. The number of pixels in the built-up histogram (Figure 4c) began to increase in the value range of -0.6 to -0.2 with a pixel frequency of about 200 pixels.
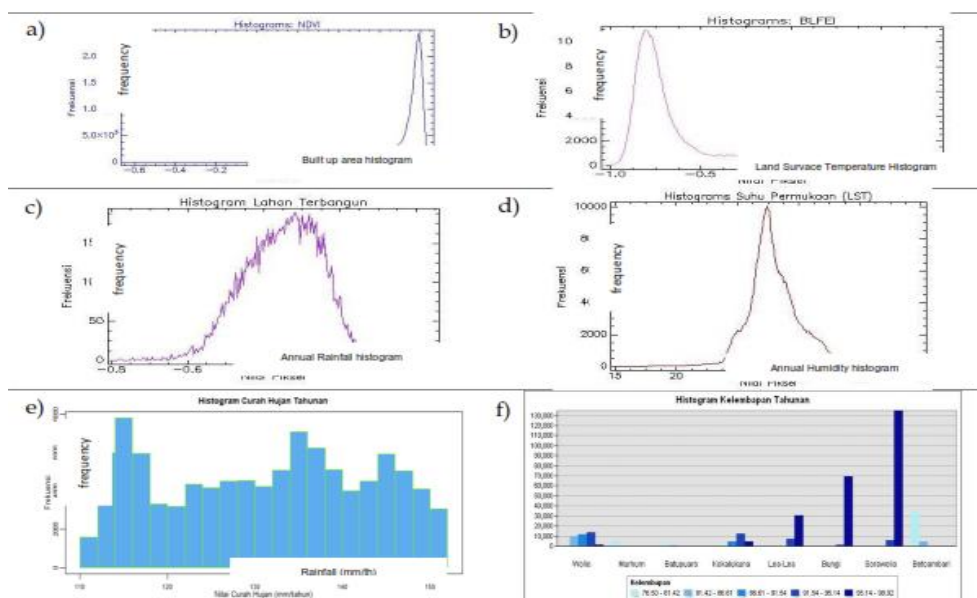

Figure 4. Histogram of Research Variables. (a). NDVI. (b). BLFEI.
(c). Histogram of Built-up Area. (d). LST. (e). Rainfall. (f). Air Humidity.

The surface temperature value is obtained by correcting the radian band 10 value using atmospheric parameters including the downwelling and upwelling values of atmospheric radiation and the object's emissivity value (Fawzi, 2014). The emissivity value of the object at the study site ranged from 0.96 to 0.97 (Figure 5d).

The surface temperature of the object at the research location based on calculations ranged from $14,64^0C$ to $40,43^0C$ with the most common temperature values found in the temperature range of 24^0C to 32^0C as seen in Figure 4d above.

The average annual rainfall at Betoambari, Kaisabu, Ngkari-ngkari and Post Rain Pada stations were 140.66, 151.84, 111.39 and 122.02 mm/year, respectively. This value was then interpolated using the IDW method and shows information that the rainfall in the study area ranged from 111.39 to 151.84 mm/year.

Figure 5e shows the range of rainfall between 111.39 to 151.84 mm/year with variations in pixel values in which the pixel groups of 112, 136 and 138 mm/year were the pixel values with the highest frequency, respectively. Meanwhile, the group of pixel values with the lowest frequency was in the group of values of 110 and 152 mm/year (Figure 4e).

In line with rainfall, humidity was also obtained by interpolating the annual average of Betoambari, Kaisabu and Ngkari-ngkari climate stations with values of 76.50, 96.75 and 98.93 mm/year, respectively. The interpolation results show a vulnerable value of 76.50-98.93 mm/year (Figure 5f) with the group value of 95.14–98.92 mm/year being the pixel value with the highest number of frequencies and located in the Sorawolio District (Figure 4f). On the other hand, Batupuaro District is the area with the smallest number of frequencies with pixel values in the group of 76.50 to 81.42 mm/year.
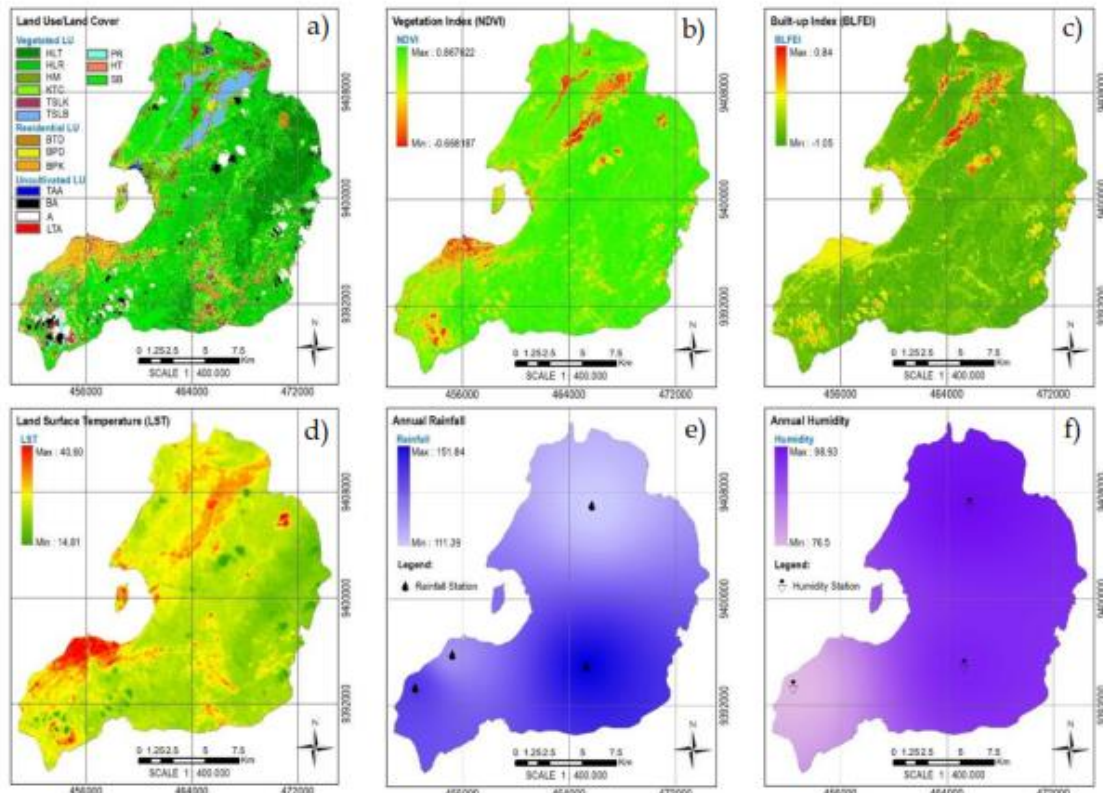

Figure 5. Dengue Fever Modeling Variables. (a) Land Cover/Use. (b) Vegetation Density Index (NDVI). (c) Building Density Index (BLFEI). (d) Surface Temperature (LST). (e) Rainfall. (f) Air Humidity.

A few programming languages (scripts) are run to classify vulnerabilities using the SVM and RF algorithms in the R software. This script was created first in the preparation stage. For SVM, a total of 48 combinations of combined values of kernel regularization (C) and kernel width (γ) were tested using the trial-and-error method, which is presented in Table 3 below:

Table 3. Results of Trial-and-Error Parameter Values for Radial Basis Function Kernel Width and Regularization. (a) Width 0.01. (b) Width 0.02. (c) Width 0.03. (d) Width 0.04. (e) Width 0.05. (f) Width 0.06. (g) Width 0.07. (h) Width 0.08.

a)

γ	C	AUC	Correlation
0,01	2	0,99	0,95
0,01	4	1	0,96
0,01	8	1	0.96
0,01	16	1	0,94
0,01	24	0,99	0,96
0,01	32	0,99	0,96

b)

γ	C	AUC	Correlation
0,02	2	0,99	0,95
0,02	4	1	0,96
0,02	8	1	0.96
0,02	16	1	0,95
0,02	24	0,99	0,96
0,02	32	0,99	0,95

c)

γ	C	AUC	Correlation
0,03	2	1	0,95
0,03	4	1	0,95
0,03	8	1	0,96
0,03	16	0,99	0,96
0,03	24	0,99	0,96
0,03	32	0,99	0,94

d)

γ	C	AUC	Correlation
0,04	2	0,99	0,95
0,04	4	1	0,96
0,04	8	1	0,95
0,04	16	0,99	0,96
0,04	24	0,99	0,95
0,04	32	0,99	0,95

e)

γ	C	AUC	Correlation
0,05	2	1	0,95
0,05	4	1	0,96
0,05	8	1	0,94
0,05	16	1	0,95
0,05	24	1	0,95
0,05	32	0,99	0,95

f)

γ	C	AUC	Correlation
0,06	2	0,99	0,95
0,06	4	1	0,96
0,06	8	1	0,96
0,06	16	1	0.96
0,06	24	0,99	0,96
0,06	32	0,99	0,96

g)

γ	C	AUC	Correlation
0,07	2	0,99	0,95
0,07	4	1	0,96
0,07	8	1	0,96
0,07	16	0,99	0,96
0,07	24	1	0,95
0,07	32	0,99	0,96

h)

γ	C	AUC	Correlation
0,08	2	1	0,95
0,08	4	1	0,95
0,08	8	1	0,96
0,08	16	1	0,96
0,08	24	1	0,95
0,08	32	0,99	0,95

Based on Table 3f above, when the kernel width value was set at 0.06 and the regularization parameter at 16, the best AUC value was 1 with a correlation of 0.96. The combination of kernel width values of 0.05 and 0.08 with regularization parameter values generally gives a maximum AUC value of 1 but still with a correlation value below 0.96 as shown in Tables 3e and 3h. Meanwhile, the value test proposed by (Ding et al., 2018) gives an AUC value of 0.99 with a correlation of 094.

As with SVM, the trial-and-error method was also used for the RF

algorithm. There were 16 values obtained from the combination of *N-tree* values of 500 and 1,000 and m_{try} values 1 to; based on Tabel 3a it is known that the highest AUC value was obtained when the *N-tree* parameter values and m_{try} parameters were set at 500 and 1. The highest AUC value obtained was 0.99 with a correlation of 0.94. The smallest AUC value obtained was 0.98 when the *N-tree* and m_{try} parameter values were set to 1000 and 8 respectively. This smallest AUC value was also found in several combinations of *N-tree* values and m_{try} values above 3, namely 5, 7, and 8.

Table 4. Results of Trial-and-Error Parameter Value Number of Decision Trees (*N-Tree*) and m_{try} (a) *N-Tree* 500. (b) *N-Tree* 1000.

N-tree	m_{try}	AUC	Correlation	*N-tree*	m_{try}	AUC	Correlation
500	1	0,99	0,94	1.000	1	0,99	0,93
500	2	0,99	0,92	1.000	2	0,99	0,92
500	3	0,99	0,92	1.000	3	0,99	0,91
500	4	0,99	0,91	1.000	4	0,98	0,91
500	5	0,98	0,91	1.000	5	0,98	0,91
500	6	0,98	0,91	1.000	6	0,98	0,92
500	7	0,98	0,91	1.000	7	0,99	0,91
500	8	0,98	0,91	1.000	8	0,98	0,91

In terms of capability comparison, the model generated by the SVM algorithm was considered to be better than the model generated by the RF algorithm with values of 1 (Figure 6a) and 0.99 (Figure 6b) respectively. Although the AUC RF value was not better than SVM, this value was still classified as an "excellent" class; meanwhile value between 0.8-0.9 is very good, 0.7-0.8 good, 0.6-07 moderate and 0, 5-0.6 bad.
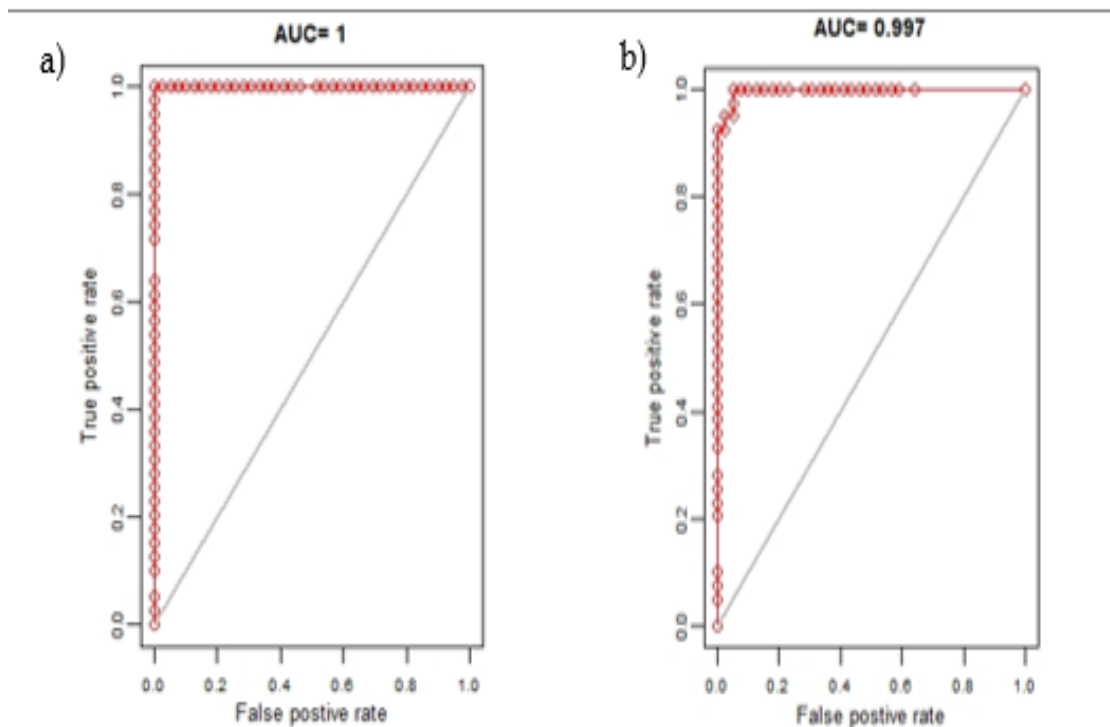

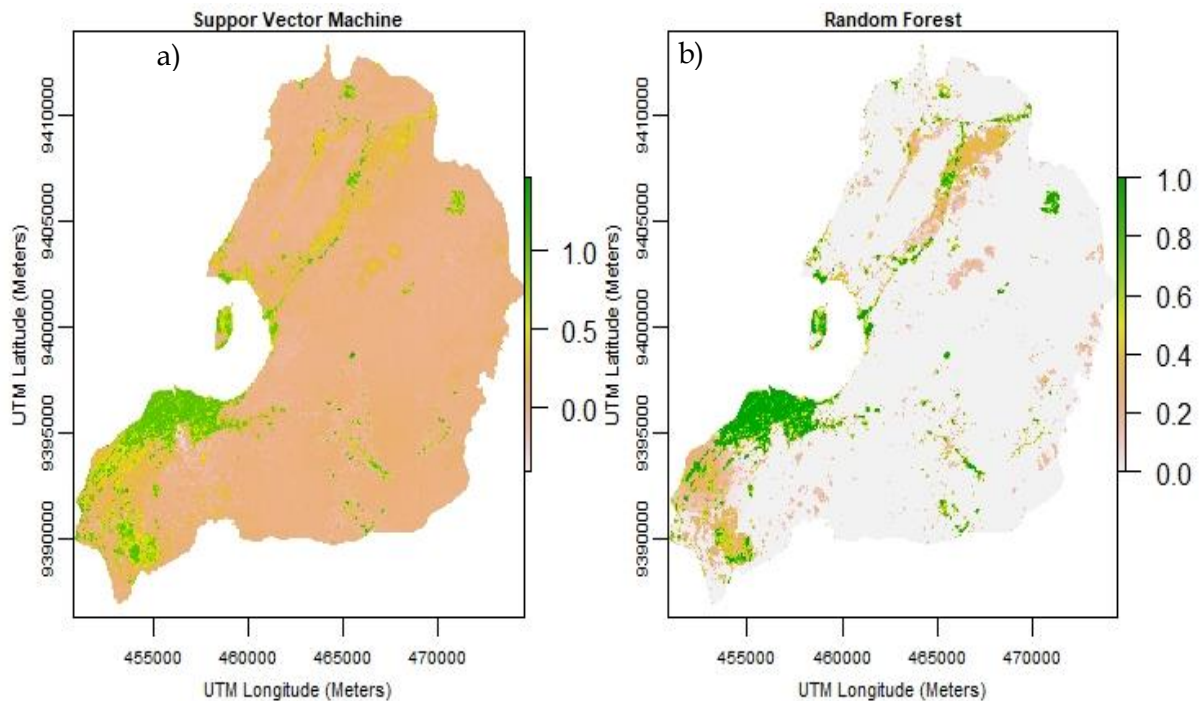Figure 6. Comparison of AUC Curve. (a) *Support Vector Machine*. (b) *Random Forest*.

Figure 7. Comparison in the Map of Classification Results.
(a) SVM Algorithm. (b) RF Algorithm.

CONCLUSION

This study provides two main conclusions including the assessment and comparison of the ability of SVM and RF algorithms in modeling DHF. In terms of assessment, the SVM algorithm provides maximum capability when the kernel width parameter value (γ) is set at 0.06 and 16 for the regularization value (C), while the RF algorithm provides the best performance when the decision tree number (*N-tree*) parameter value is set at 500 and the value of the m_{try} parameter is 1 with the value of the number of variables (*N-fact*) of 6.

For comparison of model capabilities, the SVM algorithm is considered to provide better capabilities than the RF algorithm with an AUC value of 1 and a correlation of 0.96. Although in terms of comparison the capability of the RF algorithm is not better than the SVM algorithm, the AUC value obtained is 0.99 with a correlation of 0.94 has been included in the "excellent" category to express the qualitative relationship between the AUC value and the model's capability.

REFERENCES LIST

Adzan, G., & Danoedoro, P. (2012). *Fuzzy Logic for Spatial Modelling DHF in Yogyakarta.* Jurnal Bumi Indonesia, 1(3).

Bouhennache, R., Bouden, T., Taleb-Ahmed, A., & Cheddad, A. (2018). *A New Spectral Index for the Extraction of Built-Up Land Features From Landsat 8 Satellite Imagery.* Geocarto International, 1–21.

Statistic Indonesia of Baubau City. (2018). *Baubau City in Figures 2018.*

Breiman, L. (2001). *Random Forest* (R. E. Schapire (ed.); pp. 5–32). Kluwer Academic.

National Standardization Agency (BSN). (2014). *SNI 7645-1:2014 Landuse Classification - Part 1: Medium and Small Scale.*

Coll, C., Galve, J. M., Sanchez, J. M., & Caselles, V. (2010). *Validation of Landsat-7/ETM+ Thermal-Band Calibration and Atmospheric Correction With Ground-Based Measurements. IEEE Transactions on Geoscience and Remote Sensing, 48*(1), 547–555.

Costa, E. A. P. de A., Santos, E. M. de M.,

Correia, J. C., & Alburquerque, C. M. R. de. (2010). *Impact of Small Variations in Temperature and Humidity on the Reproductive Activity and Survival of Aedes aegypti (Diptera , Culicidae). Revista Brasileira de Entomologia, 54*(No. 3), 488–493.

Danoedoro, P. (2012). *Digital Remote Sensing* (B. R. W (ed.); 1st ed.). C.V Andi Offset.

Dempsey, C. (2012). *Where is the Phrase "80% of Data is Geographic" From?*

Public Health Office of Southeast Sulawesi Province. (2018). *Southeast Sulawesi Province Health Profile.*

Ding, F., Fu, J., Jiang, D., Hao, M., & Lin, G. (2018). *Mapping the Spatial Distribution of Aedes aegypti and Aedes albopictus.* Acta Tropica, 178(August 2017), 155–162.

Fawzi, N. I. (2014). *Surface Emissivity Mapping Using Vegetation Indices. Majalah Ilmiah Globë, 16*(2), 133–139.

Ganie, M. A., & Nusrath, A. (2016). *Determining the Vegetation Indices (NDVI) from Landsat 8 Satellite Data. International Journal of Advanced Research, 4*(8), 1459–1463.

Gubler, D. J. (2012). *The Economic Burden of Dengue. Am. J. Trop. Med. Hyg, 86*(5), 743–744.

Hidayatuloh, A. (2020). *Introduction of R Programming and RStudio.* https://bookdown.org/aepstk/intror/

Hijmans, R. J., & Elith, J. (2019). *Spatial Distribution Models - R Spatial.*

Iriani, Y. (2012). *Relationship between Rainfall and Increase in Child Dengue Hemorrhagic Fever Cases in Palembang City. Sari Pediatri, 13*(6), 378–383.

Kamimurai, K., Matsusei, I. T., Takahashii, H., Komukaii, J., Fukuda, T., Suzuki, K., Aratani, M., Shirai, Y., & Mogi, M. (2002). *Effect of Temperature on the Development of Aedes aegypti and Aedes albopictus. The Japan Society of Medical Entomology and Zoology, 53*(No. 1), 53–58.

Ministry of Health of the Republic of Indonesia. (2017). *The Situation of Dengue Fever Disease in Indonesia in 2017.*

Ministry of Health of the Republic of Indonesia. (2019, February). *DHF Cases Continue to Increase.* PSN. 1–2.

Kurniadi, H., Aprilia, E., Utomo, J. B., Kurniawan, A., & Safril, A. (2018). *Comparison of IDW and Spline Methods in Interpolation of Rainfall Data (Case Study of Monthly Rainfall in East Java Period 2012-2016). GEOTIK Proceeding* 2018, 215.

Lawrence, R. L., Wood, S. D., & Sheley, R. L. (2006). *Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). Remote Sensing of Environment, 100*(3), 356–362.

Lee, H., Kim, J. E., Lee, S., & Lee, C. H. (2018). *Potential Effects of Climate Change on Dengue Transmission Dynamics in Korea. PLoS ONE, Vol.* 13(No. 28, Juni 2018), 1–23.

Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest. R News,* 2(3), 18–22.

Motevalli, A., Pourghasemi, H. R., & Zabihi, M. (2018). *Assessment of GIS-Based Machine Learning Algorithms for Spatial Modeling of Landslide Susceptibility : Case Study in Iran. Comprehensive Geographic Information Systems,* 266–268.

Mountrakis, G., Im, J., & Ogole, C. (2011). *Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing,* 66(3), 247–259.

NASA. (2019). *Landsat Science Education.* http://landsat.gsfc.nasa.gov/?p=5186

Ngoc Thach, N., Bao-Toan Ngo, D., Xuan-Canh, P., Hong-Thi, N., Hang Thi, B., Nhat-Duc, H., & Dieu, T. B. (2018). *Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study. Ecological Informatics,* 46(May), 74–85.

Nguyen, N. T., Dang, B.-T. N., Pham, X.-C.,

Nguyen, H.-T., Bui, H. T., Hoang, N.-D., & Bui, D. T. (2018). *Spatial Pattern Assessment of Tropical Forest Fire Danger at Thuan Chau Area (Vietnam) using GIS-based Advanced Machine Learning Algorithms : A Comparative Study. Ecological Informatics, 46*(March), 74–85.

Palaniyandi, M. (2014). *The Environmental Aspects of Dengue and Chikungunya Outbreaks in India: GIS for Epidemic Control. International Journal of Mosquito Research, 1*(2), 35–40.

Pineda-Cortel, M. R. B., Clemente, B. M., & Nga, P. T. T. (2019). *Modeling and predicting dengue fever cases in key regions of the Philippines using remote sensing data. Asian Pacific Journal of Tropical Medicine, 12*(2), 60–66.

R Foundation. (2020). *What is R.* https://www.r-project.org/about.html

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). *Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLoS ONE,* 1–22.

Swets, J. A. (1988). *Measuring the Accuracy of Diagnostic Systems Linked references are available on JSTOR for this article : Measuring the Accuracy of Diagnostic Systems. Science,* 240(4857), 1285–1293.

Tiong, V., Abd-jamil, J., Nellis, S., Fauzi, R., & AbuBakar. (2015). *Evaluation of land cover and prevalence of dengue in Malaysia. Tropical Biomedicine,* 32(4), 587–597.

USGS. (2019). *USGS EROS Archive - Landsat Archives - Landsat 8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor).*

Valdez, L. D., Sibona, G. J., & Condat, C. A. (2018). *Impact of Rainfall on Aedes aegypti Populations. Ecological Modelling,* 385(November 2017), 1–27.

Vanwambeke, S. O., Lambin, E. F., Eichhorn, M. P., Flasse, S. P., Harbach, R. E., Oskam, L., Somboon, P., Beers, S. van, Benthem, B. H. B.

van, Walton, C., & Butlin, R. K. (2007). I*mpact of Land-use Change on Dengue and Malaria in Northern Thailand. EcoHealth, May 2014.*

WHO. (2007). *International Health Regulations (2005) Areas of Work for Implementation. In World Health Organization (Issue June).* http://apps.who.int/iris/bitstream/10665/69770/1/WHO_CDS_EPR_IHR_2007.1_eng.pdf?ua=1

WHO. (2019). *Neglected tropical diseases.* https://www.who.int/neglected_diseases/en/

Widayani, P. (2010). *Spatial Modeling of Dengue Hemorrhagic Fever Epidemiology Using Geographic Information System in Depok District, Sleman Regency, Yogyakarta. Jurnal Geografi Gea,* 10(2).

Widayani, P., Yanuar, R., & H, A. Y. (2018). *Relationship Analysis of Environmental Factor Change on the Dengue Fever Diseases Using Image Transformation (Case Study : Surakarta City). IOP Publising,* 169(July), 1–10.

Wilder-Smith, A., Ooi, E. E., Horstick, O., & Wills, B. (2019). *Dengue. The Lancet,* 393(10169), 350–363.

World Economic Forum. (2017). *Global Shapers Survey. In World Economic Forum.* http://www.shaperssurvey2017.org

Yue, Y., Sun, J., Liu, X., Ren, D., Liu, Q., Xiao, X., & Lu, L. (2018). *International Journal of Infectious Diseases Spatial Analysis of Dengue Fever and Exploration of Its Environmental and Socio-Economic Risk Factors using Ordinary Least Squares : A Case Study in Five Districts of Guangzhou City , China, 2014. International Journal of Infectious Diseases*, 75(No. 27 Juli 2018), 39–48.