❑2821

Computer model for tsunami vulnerability using sentinel 2A and SRTM images optimized by machine learning

Sri Yulianto Joko Prasetyo[1], Bistok Hasiholan Simanjuntak[2], Kristoko Dwi Hartomo[3], Wiwin Sulistyo[4]
[1,3,4]Faculty of Information Technology, Satya Wacana Christian University, Indonesia
[2]Faculty of Agriculture and Bussines, Satya Wacana Christian University, Indonesia

Article Info

ABSTRACT

This study aims to develop a software framework for modeling of tsunami vulnerability using DEM and Sentinel 2 images. The stages of study, are: 1) extraction Sentinel 2 images using algorithms NDVI, NDBI, NDWI, MSAVI, and MNDWI; 2) prediction vegetation indices using machine learning algorithms. 3) accuracy testing using the MSE, ME, RMSE, MAE, MPE, and MAPE; 4) spatial prediction using Kriging function and 5) modeling tsunami vulnerability indicators. The results show that in 2021 the area was dominated by vegetation density between (-0.1-0.3) with moderate to high vulnerability and risk of land use tsunami as a result of the decreasing of vegetation. The prediction results for 2021 show a low canopy density of vegetation and a high degree of land surface slope. Based on the prediction results in 2021, the study area mostly shows the existence of built-up lands with a high tsunami vulnerability risk (>0.1). Vegetation population had decreased to 67% from the original areas in 2017 with an area of 135 km^2. Forest vegetation had decreased by 45% from 116 km^2 in 2017. Land use for fisheries had increased to the area of 86 km^2 from 2017 with an area of 24 km^2.

Corresponding Author:

Sri Yulianto Joko Prasetyo
Faculty of Information Technology
Satya Wacana Christian University, Indonesia
Email: sri.yulianto@uksw.edu

1. INTRODUCTION

Throughout history, the two largest tsunami disasters ever recorded were the tsunami on the island of Sumatra, Indonesia in 2004 with casualties reached 220 thousand people, and the tsunami in Tohoku Japan in 2011 with 20 thousand dead and missing victims. In anticipation of similar events in the future, monitoring of various tsunami risk factors is required, such as run-up and inundation forecast data. Researchers from various countries around the world are looking for new data sources to develop tsunami models, one of which is by using remote sensing data, including images of ALOS AVNIR-2, ASTER GDEM, IKONOS, DEM SRTM, Synthetic Aperture Radar, Landsat, and Sentinel [1], [2]. Tsunami risk determination in several countries such as Portugal, Japan, and Indonesia so far is carried out using DEM to obtain an overview of the slope of the coastline [3], [4]. The currently available DEM data is generated from remote sensing image analysis, such as the SRTM interferometric synthetic aperture radar (InSAR) image [5]. The DEM data plays an important role in tsunami modeling because it can provide an overview of the geomorphology, hydrology, and ecology of a potential tsunami area. In addition to the DEM, tsunami modeling requires coastal LULC data for hazard and vulnerability assessments [6]. VI is a new method of LULC analysis based on radiometric measurements of visible light and infrared spectra with indicators of leaf area, percentage of vegetation cover, chlorophyll content, and vegetation biomass [7], [8]. LULC is dynamic and naturally occurring but

sometimes changes rapidly and suddenly due to anthropogenic activity. Analysis of the dynamics of changes in LULC is very important for policy making, planning, and implementation of tsunami risk assessment in coastal areas [9].

The research is focused on providing solutions to current problems, specifically in how to develop algorithms and computer models to classify the height of tsunami's vulnerability using indicators: 1) elevation and contours of coastal areas using the SRTM images; 2) dynamics of LULC using the VI extracted from Sentinel 2; and 3) prediction of LULC spatial pattern using the ML algorithm. The research novelty is development a tsunami high vulnerability area identification software framework using the DEM, LULC, and VI indicators as part of tsunami mitigation[10]. This software framework was developed with the following concepts: 1) the LULC temporal dynamics which are identified using VI indicators consist of normalized difference vegetation index (NDVI), modified soil adjusted vegetation index (MSAVI), normalized difference water index (NDWI), modified normalized difference water index (MNDWI), and normalized difference built-up index (NDBI) [11], [12]; 2) The spatial dynamics of LULC which are predicted using ML consist of Classification and Regression Tree (CART), multivariate adaptive regression spline (MARS), random forest (RF), and k-nearest neighbors (k-nn) [13], [14]; 3) Slopes analysis of DEM SRTM data in areas with a high risk of tsunami waves [15], [16]; 4) Create algortihms for computer modeling framework to classify tsunami high vulnerability areas using the DEM, LULC, and VI indicators.

2. THEORETICAL BACKGROUND

Sentinel 2 is a pair of remote sensing satellites consisting of Sentinel 2A and Sentinel 2B with the ability to provide images at scales 10 m, 20 m, and 60 m using multi-spectral instrument (MSI). Sentinel 2 has a duration of visit for the same point of 5 days. Sentinel 2 is composed of 13 bands (440 nm~2200 nm), which includes visible light, near infrared and infrared. Three of these bands are in the 670 nm~760 nm spectrum, thus provide more effective data for monitoring vegetation growth [17]. Sentinel 2 provides multispectral imagery and radar instruments for monitoring sea, land, and atmosphere particularly ground and water surface conditions to provide imagery of vegetation, land, water cover, and inland waterways. Sentinel 2 using the MSI has a spatial resolution of 10 m in visible light for monitoring, change detection, and coastal zone mapping [18]. The band and spectrum data for the sentinel 2-A image are presented in Table 1.

Table 1. Data band and imagery spectrum of sentinel 2-A [19]

Band	Spatial Resolution (m)	Central wavelength (μm)
B1 –Coastal aerosol	60	0.443
B2 –B	10	0.494
B3 –G	10	0.560
B4 –R	10	0.665
B5 –Red edge 1	20	0.704
B6 –Red edge 2	20	0.740
B7 –Red edge 3	20	0.780
B8 –NIR	10	0.843
B8A –NIR narrow	20	0.864
B9 –Water vapor	60	0.944
B10 –SWIR Cirrus	60	1.375
B11 –SWIR1	20	1.612
B12 –SWIR2	20	2.194

Sentinel 2 MSI multispectral image contains 13 spectral bands in visible light and VNIR, as well as SWIR, with a total of four bands at 10 m, six bands at 20 m, and three bands at atmospheric correction spatial resolution of 60 m. A total of 4 bands of 10 m spatial resolution and 6 bands of 20 m spatial resolution are used for environmental monitoring, whereas 3 bands of 60 m spatial resolution are designed for atmospheric correction and cloud detection [20]. Judging from band composition, Sentinel 2 imagery is suitable for vegetation analysis because it has a finer spatial resolution compared to other satellite images. Besides, Sentinel 2 wavelength is sensitive to chlorophyll contents and its phenological status is higher than other images so that the classification of vegetation indices is more accurate [21]. The SRTM is a radar-based remote sensing image with a spatial resolution of 30-92 meters. SRTM is used as an alternative to replace the need for high-resolution satellite data which is very expensive. The SRTM is chosen usually for its accessibility, feature resolution, vertical accuracy, and lower amount of noise compared to other alternative global DEMs [22]. The SRTM is used to study topography, geomorphology, vegetation cover, tsunami assessment, and urban areas. Until now, SRTM has been available globally and can be accessed easily and

cheaply [23]. The SRTM has a pixel size of 90 m, an accuracy of 16 m in a vertical position, and WGS84 as its reference ellipsoid [16]. The SRTM consistsof single-pass C and X-band interferometric synthetic aperture radars (InSAR). The SRTM can record terabytes of data from topographic maps of the earth with high accuracy from latitudes 60 N to 56 S. More than 95% of the targeted earth's surface areas have been mapped at least twice to reduce interferometric altitude errors. When completed, the SRTM topographic map will provide coverage of approximately 80% of the earth's surface. The C-band SRTM dataset consists of two types of data: 1) principal investigator processor (PI) data; 2) ground data processing system (GDPS) [24].

The VI is a combination of spectral measurements in different wavelengths as recorded by the radiometric sensor and assists in the analysis of multispectral image information by reducing the multidimensional data to a single value. The spectrum and wavelength VI are: ultraviolet (10-380), blue (450-495), green (495-570), red (620-750) anda NIR (850-1700). VI is defined as a dimension-reducing radiometric measurement mechanism involving linear ratios and combinations of the red and near-infrared (NIR) portion of the spectrum [25]. VI is a mathematical combination of various spectrum bands designed to separate the pixel values of different features numerically [26]. The VI equation in the Table 2.

Table 2. The formula for determining the VI value in sentinel 2 imagery

VI	Formula	Description	Reference
NDVI	$NDVI = \dfrac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}}$	The NDVI is calculated using Red and Infrared bands which are reflected by the vegetation canopy, this is because healthy vegetation absorbs visible light and reflects near-infrared band, whereas unhealthy vegetation reflects visible light and reflects near-infrared band in less amounts. The NDVI is in the range of values between (-1) which represents no vegetation and (+1) which represents high vegetation density.	[27], [28]
MNDWI	$MNDWI = \dfrac{\rho_{green} - \rho_{SWIR1}}{\rho_{green} + \rho_{SWIR1}}$	NDWI is designed to maximize the light reflectance of water bodies in the Green band and minimize the reflectance of water bodies in the NIR band [29]. The NDWI values vary from a negative value (-1) which represents soil to a positive value (+1) which represents the surface of open water bodies and thick land cover vegetation.	[30], [31]
NDWI	$NDWI = \dfrac{\rho_{green} - \rho_{NIR}}{\rho_{green} + \rho_{NIR}}$		[32]
NDBI	$NDBI = \dfrac{\rho_{SWIR} - \rho_{NIR}}{\rho_{SWIR} + \rho_{NIR}}$	NDBI is used to assess the reflectance of an established land and has a higher index value compared to other land uses or land covers. The NDBI value varies from a negative value (-1) which represents land with a low proportion of buildings to a positive value (+1) which represents land with a higher proportion of buildings.	[33], [34]
MSAVI	$MSAVI = \dfrac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red+L}} \cdot (1+L)$	The soil surface and vegetation canopy will absorb solar radiation with a wavelength spectrum of 842 nm (NIR) and 705 nm (Red Edge1). The solution to overcome this problem is to calculate the square root of the reflectance absorbed by the canopy of vegetation and soil.	[35], [36]

The data classification method using the ML algorithm has become the focus of many remote sensing researchers because it can model high-dimensional, non-parametric data that is abnormally distributed and produces a high accuracy when compared to conventional classification methods for a large number of predictors [37]. Literature studies show that various ML algorithms have been applied to classify remote sensing data, among which the focus of this study is RF, k-nn, MARS, and CART [38]-[39]. The RF algorithm is a part of the ML method used to classify high complexity and non-parametric data as well as remote sensing image data. The RF algorithm classifies data through a decision tree formation process, where each decision tree is built using bootstrap sample data from two-thirds of the training data taken randomly. The next third of the training data is used to determine the error factor in the data. Each classifying variable is given one label indicating that it is in a certain class. This is done on all sample data and the final result is the classifier that has the most labels [40]. RF uses bagging, which is an ensemble technique to increase the accuracy and stability of a model. RF uses the idea of bagging or bootstrap aggregation, namely selecting the class of classifying variables or predictors to make the final decision. For each the prediction function tree, $f(x)$ is defined by (1) and (2).

$$f(x) = \sum_{m=1}^{M} C_m \Pi(x, R_m) \tag{1}$$

$$\Pi(x, R_m) = \begin{cases} 1, if \; x \in R_m \\ 0, otherwise \end{cases} \tag{2}$$

The notion M is denoted by the number of areas observed, R_m is the area corresponding to the classifying variable or predictor variable with m, and C_m is the constant corresponding to m. The final classification decision is made from the class that has the most labels [41].

MARS algorithm is a nonlinear and nonparametric regression method used to find the relationship between dependent and independent variables. The MARS method predicts using the basis function (BF). The definition of MARS is presented in (3).

$$f(x) = \beta_0 + \sum_{i=1}^{n} \beta_i \lambda_i(X) \tag{3}$$

Where β_0 is the constant, β_i is the BF coefficient, $\lambda_i(X)$ is BF and n is the number of BF in the model. BF is determined using the following function in (4).

$$\max(0, x - \alpha) \; or \; \max(0, \alpha - x) \tag{4}$$

Where x is the independent variable and α is the constant related to the independent variable [42]. CART is a method in ML which is used to classify objects into two or more populations. CART works by identifying and configuring predictor variables to produce more homogeneous data. The determination of the classification variable is carried out using the Gini Index equation as shown in (5) [43].

$$Gini(t) = \sum_{i=1}^{c} P_{\omega i}(1 - P_{\omega i}) \tag{5}$$

Where c is the number of classes and $P_{\omega i}$ is the probability of class ωi at point t. $P_{\omega i}$ is defined as shown in (6).

$$P_{\omega i} = \frac{n_{\omega i}}{N} \tag{6}$$

Where $n_{\omega i}$ is the number of samples in class ωi and N is the number of trained samples [44].

The k-nn algorithm is intrinsically non-parametric and is generally applied for pattern recognition purposes. The k-nn algorithm works by calculating the distance between the unknown sample and the nearest known sample. The unknown sample is assigned a label or a class which is calculated from the average distance of the k-nearest neighbor variables [45]. The main principle of k-nn is that the classified points are determined based on their proximity to k. For example $N = \{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$ is training set and T is the number of observed data entities. The notation $x_i \in R^d$ is a vector and $y_i \in \mathcal{Y} = \{c_1, c_2, ... c_m\}$ as a label for data classification $i = 1,2,...N$. If there is data input x, the entity distance will be determined in the training data as the k-nn value which is denoted as $N_k(x)$. The distance function (d) in k-nn is denoted by the (7) [46].

$$d(x,y) = \sqrt{(x_1 - y_1)^T \sum^{-1}(x_1, y_1)} \tag{7}$$

Ordinary kriging (OK) is an interpolation method that works using the principle of Spatial Autocorrelation, which assumes that a point closer to the prediction point has a greater value than a sample point that is farther from the prediction point [47]. OK is defined using (8).

$$\gamma(d) = \frac{1}{2N(d)} \sum_{i=1}^{N(d)} [Z(x_i) - Z(x_i + d)]^2 \tag{8}$$

Where $\gamma(d)$ is the variogram value at the distance d, $Z(x_i)$ is the value of the observed variable at location (x_i), the notation $N(d)$ is the number of all observation points in the range of distance d. The prediction of the variable distribution is calculated using the $Z*(x)$ linear regression, as shown in (9). Where $\lambda_i(x)$ is weight, $m(x)$ and $m(x_i)$ are mathematical expectations of random variables $Z*(x)$ and $Z*(x_i)$ [48].

$$Z*(x) - m(x) = \sum_{i=1}^{n(x)} \lambda_i(x)[Z(x_i) - m(x_i)] \tag{9}$$

3. RESEARCH METHOD

The research area is Kebumen District, Central Java Province, Indonesia as shown in Figure 1. The research was conducted in five step in Figure 2, are:

a. Data preprocessing. This stage includes collecting Sentinel 2 satellite image data obtained from www.earthexplorer.usgs.gov. The Sentinel 2 satellite images are corrected geometrically, radiometrically, and atmospherically. Geometric correction aims to correct the position of the coordinates of each pixel in the image the same as shown on the earth's surface by taking into account satellite movements, earth rotation and terrain effect [49]. Radiometric correction aims to restore the digital number (DN) value that has been calibrated to reduce the noise component that interferes with the reflectance reception. Atmospheric correction aims to reduce the atmospheric effect that interferes with the reflectance and reduce the DN value [50].

b. Data analysis and classification. This stage includes the extraction of Sentinel 2 image data using the NDVI, NDBI, NDWI, MSAVI, and MNDWI algorithms. The results of data extraction are in the form of numerical values that can be used for classification and prediction using ML. Image data is classified through a supervised classification to form a thematic map of the Land Used land Cover of the study area. The SRTM DEM image data are classified to analyze the contours of the study area.

c. Prediction of NDVI, NDBI, NDWI, MSAVI, and MNDWI data using RF, MARS, and CART.

d. Testing the accuracy of prediction results using ML which is performed statistically using the MSE, ME, RMSE, MAE, MPE, and MAPE equations. If the result of the prediction testing is accurate, the analysis of the level of vulnerability will be continued. If the result of the analysis is not accurate, it will be analyzed again using ML. A high degree of statistical accuracy can be seen from the calculation of MSE, ME, RMSE, MAE, and MPE MAPE which is close to zero.

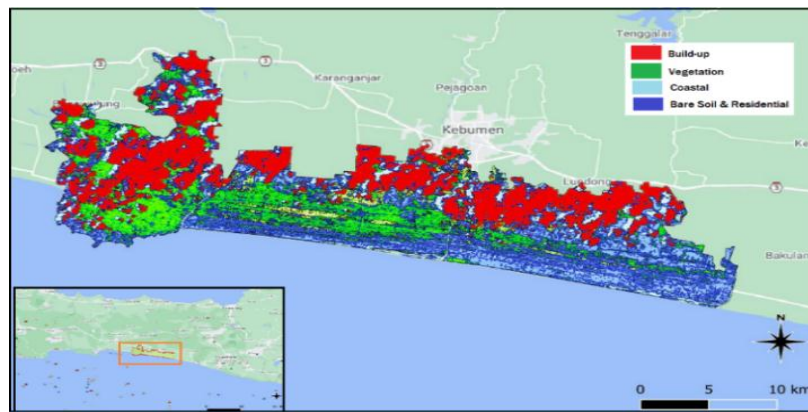e. Data visualization is done by using Geostatistical spatial interpolation, namely Ordinary Kriging.

Figure 1. The research area of kebumen district, central java province, Indonesia
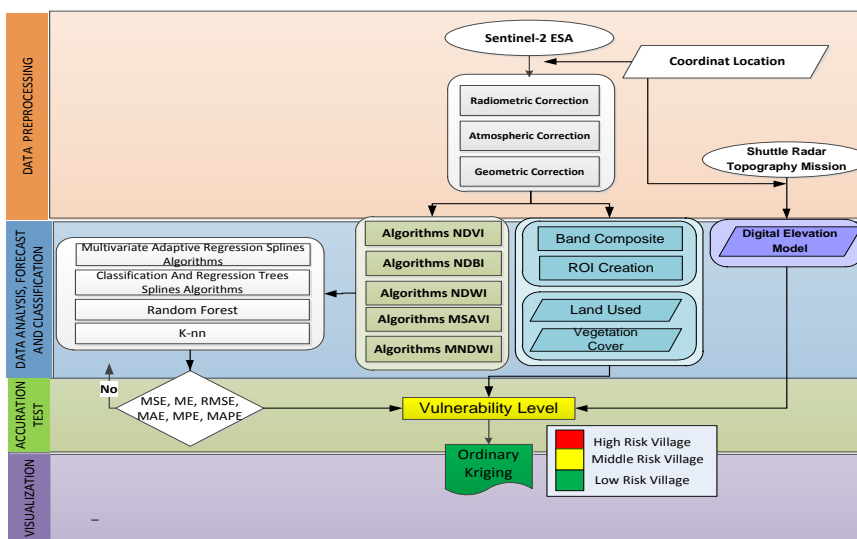

Figure 2. Framework computer model for tsunami vulnerability

4. RESULTS AND DISCUSSION

Vulnerability components are: 1) physical condition; 2) socio-culture, (3) economy and 4) disaster-prone environment. One of the vulnerability components that play an important role in controlling the level of tsunami hazard is LULC. Areas that have LULC with high vegetation density will reduce the speed of free water flow compared to the presence of buildings, water bodies and mixed vegetation [51]. In this study, LC is classified into 5 classes, namely 1) vegetation class; 2) water body class; 3) built-up land class; 4) open land class and 5) land cover class outside the classification. The vegetation class in the study area shows a decline from 2017 that covered an area of 135 km^2, in 2018 that covered an area of 124 km^2, 2019 that covered an area of 96 km^2 and in 2020 that covered an area of 91 km^2. This decrease was caused by changes in land cover from coastal forests to built-up lands for social, cultural and economic functions for the community in Figure 3 (a).

According to USGS (United States Geological survey), things that are included in Build-up or Urban class i.e. residential land; commercial and services land; industrial land, land for transportation, communication, and utilities; industrial and commercial complexes; mixed urban land and built-up land. The unclassified class is a land cover which its condition is between vegetation and bare soil, namely the condition of seasonal plant vegetation such as shrubs, clumps and grasses that are on open land and have not been used optimally. The results show that the land cover for build-up land is very high above the vegetation class. Data show that the class of built-up land in 2017 is 187 km^2, in 2018 is 170 km^2, in 2019 is 170 km^2 and in 2020 is 51 km^2. The class of built-up land is relatively low compared to other classes, namely in 2017 that covered an area of 24 km^2, in 2018 that covered an area of 23 km^2, in 2019 that covered an area of 25 km^2 and in 2020 that covered an area of 86 km^2. The high built-up land class is an indicator of the presence of high social, cultural, physical and economic factors in coastal areas that are at high risk of tsunami waves.

Smaller vegetation area compared to built-up land is an indicator of the low ability of the vegetation to absorb tsunami energy. In this study, land cover is classified into 9 classes, namely 1) rice field; 2) forest 3) river 4) pond 5) shoreline 6) settlement 7) road 8) industry and 9) vegetation. The forest land use is in a separate class from vegetation because forest is defined as an ecosystem in the form of a stretch of land containing biological natural resources dominated by trees with woody, high canopy and decades of age characteristics. Vegetation is defined as the entire plant community at a particular location in the form of forests, gardens, grasslands, agricultural crops and so on. Land use for social, cultural, physical and economic activities in coastal areas was very high from 2017 to 2020. Data show that land use for paddy fields around coasts was relatively widespread and fluctuated due to seasonal patterns. During the planting season in 2018, land use covered 65 km^2. Land use for road access was very high in 2017 that covered an area of 132 km^2, in 2018 covered an area of 130 km^2, in 2019 covered an area of 42 km^2 and in 2020 covered an area of 108 km^2. The land use for the fishing industry has increased from year to year, in 2017 it was 24 km^2, in 2018 it was 23 km^2, in 2019 it was 25 km^2 and in 2020 it was 86 km^2. The use of land for forests from year to year decreased sharply from 2017 that covered an area of 116 km^2, in 2018 that covered an area of 59 km^2, 2019 that covered an area of 73 km^2 and 2020 that covered an area of 53 km^2. This condition indicates that the factor of land use in coastal areas has a high risk of tsunami waves in Figure 3 (b).
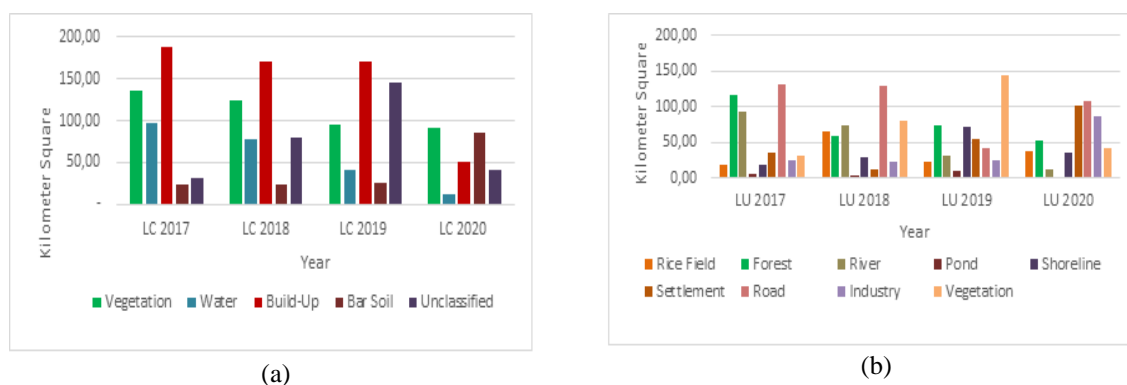

(a)   (b)

Figure 3. These figures are; (a) land cover, and (b) land used, of the study area from 2017 to 2020

Kebumen Regency has 8 sub-districts that are at high risk of tsunami. LULC areas with vulnerability area in Kebumen Regency consisting sub-districts of (a) Ambal, (b) Ayah, (c) Buayan, (d) Buluspesantren, (e) Klirong, (f) Mirit, (g) Petanahan, and (h) Puring. VI is an important instrument for

detecting and monitoring LULC dynamics [52]. In this study, an experiment was performed to detect LULC conditions using VI which includes NDVI as an indicator of vegetation health, NDBI as an indicator of built-up land use, NDWI and MNDWI as indicators of open water, and MSAVI as an indicator of open land. The dynamic changes of the LULC component within a certain period of time can be observed using the VI time series and can be used as an indicator of the level of vulnerability of a region to tsunami waves. VI Prediction aims to predict the changes in LULC in the future caused by various risk factors based on changes in the DN of Sentinel 2 images in different observation periods in the past. VI Prediction is done using ML of MARS, CART, RF, and k-nn.

The data distribution pattern vegetation indices and ML years of 2016-2020 is expressed in the form of Quartile (Q) values which consist of Quartile 1, Quartile 2, and Quartile 3. Quartile 1 is 25% of the data, Quartile 2 is 50% of the data and Quartile 3 is 75% of data using boxplot in Figure 4.
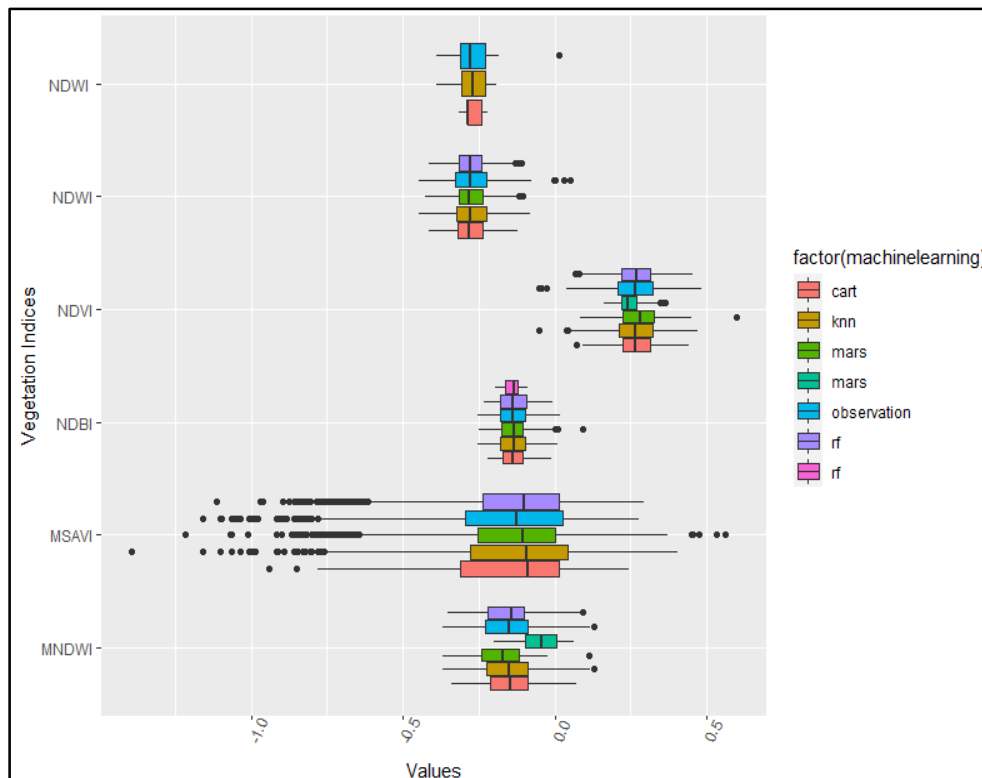

Figure 4. The boxplot of data distribution pattern vegetation indices and ML years of 2016-2020 is expressed in the form of Quartile (Q) values which consist of Quartile 1, Quartile 2, and Quartile 3

Data of MNDWI has a minimum value of -0.3710, Q1 has a value of -0.2250, Q2 has a value of -0.1354, Q3 has a value of -0.930 and the maximum value is 0.1300. The prediction data of NDWI has a value of -0.4470, has a value of -0.3190, Q2 has a value of -0.2721, Q3 has a value of -0.2360 and the maximum is 0.0500. The positive value on the water index indicates that LULC in the coastal area of Kebumen Regency is in the form of water surfaces and water bodies while the negative value represents land surfaces. The testing of the accuracy and validation of NDWI and MNDWI predictions for 182 villages in the study area was carried out using the statistics method as presented in Table 3.

Table 3. The testing of accuracy and validation of MNDWI prediction of 182 villages of study area

	MNDWI				NDWI			
	MARS	CART	RF	k-nn	MARS	CART	RF	k-nn
MSE	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001
ME	0.002	0.003	0.002	0.001	0.002	0.002	0.002	0.001
RMSE	0.031	0.034	0.024	0.025	0.042	0.036	0.031	0.026
MAE	0.024	0.026	0.016	0.018	0.028	0.026	0.018	0.015
MPE	-2.537	-0.800	0.714	-0.952	2.616	1.168	2.026	2.188
MAPE	30.239	33.118	22.657	22.430	13.656	12.108	9.507	7.903

Table 3 show that compared to other ML algorithms, the RF and k-nn algorithms show the smallest value so that it is the most optimal algorithm approaching the observation data. The results of RF and k-nn predictions can be used as an indicator to assess changes in the LULC, especially changes in the water index, which can be seen from changes in MNDWI and NDWI.

The NDVI is the main source of land cover information on LULC, used as an indicator of whether vegetation contains a chlorophyll element. MSAVI is a result of NDVI transformation to reduce the soil effect on the analysis results. A NDVI value closing to zero represents water bodies, water masses, wetlands, while a NDVI value approaching one represents grasses, shrubs, and thick canopy forest vegetation, in Figure 4. The prediction data for NDVI (a) shows a minimum value of -0.0530, Q1 (25% of data) of 0.2180, Q2 (50% of data) of 0.2614, Q3 (75% of data) of 0.3187, and a maximum of 0.5970. The prediction data for MSAVI (b) shows a minimum value of -1.3930, Q1 (25% of data) of -0.2797, Q2 (50% of data) of -0.1653, Q3 (75% of data) of 0.0210, and a maximum value of 0.5590. These results show that most of the value data is close to zero so that LULC in the coastal area of Kebumen Regency is in the form of water bodies, water masses, wetlands, and thick canopy forest vegetation. Testing of the accuracy and validation of NDVI and MSAVI predictions of 182 villages in the study area was carried out using the Statistics method as presented in Table 4.

Table 4. The testing results of the accuracy and validation of NDVI and MSAVI predictions of study area

	NDVI				MSAVI			
	MARS	CART	RF	knn	MARS	CART	RF	knn
MSE	0.002	0.002	0.001	0.001	0.050	0.068	0.031	0.016
ME	-0.003	-0.003	-0.004	0.000	0.007	0.012	0.011	-0.021
RMSE	0.046	0.044	0.035	0.031	0.224	0.260	0.176	0.127
MAE	0.036	0.032	0.023	0.019	0.159	0.161	0.114	0.067
MPE	0.385	1.349	1.096	2.718	-12.273	38.097	-51.864	-14.950
MAPE	14.816	13.811	10.708	8.842	36.444	49.821	22.825	15.452

Table 4 show that compared to other ML algorithms, the RF and k-nn algorithms show the smallest value so that it is the most optimal algorithm approaching the observational data. The results of RF and k-nn predictions can be used as indicators to assess changes in the LULC, especially changes in canopy vegetation seen from changes in NDVI and MSAVI. NDBI is an indicator used to assess built-up lands based on the unique spectral characteristics of an image. NDBI values are between -1 and 1 which are used to differentiate built-up lands from vegetation and water bodies in Figure 4. Negative value represents vegetation and water bodies while positive value represents built-up lands. The data of NDBI prediction shows a minimum value of -0.2560, Q1 (25% of the data) of -0.1777, Q2 (50% of the data) of -0.1355, Q3 (75% of the data) of -0.1030 and the maximum value of 0.0920. Table 5 shows that the smallest values compared to other ML algorithms are RF and k-nn, so that in the prediction of NDBI data, the RF and k-nn are the most optimal algorithms. The results of RF and KNN predictions can be used as an indicator of built-up lands, and the dynamics of the NDBI value shows structural changes in the LULC.

Table 5. The testing results of the accuracy and validation of NDBI predictions of study area

	MARS	CART	RF	k-nn
MSE	0.001	0.001	0.000	0.000
ME	0.000	0.002	0.003	-0.002
RMSE	0.029	0.025	0.020	0.015
MAE	0.023	0.019	0.015	0.012
MPE	-3.927	-7.281	-7.980	0.301
MAPE	22.995	18.014	15.515	10.173

The MARS algorithm is applied to predict VI on LULV in relation to vulnerability because the VI data includes a large number of variables with nonlinear characteristics and contains a high degree of interactions among predictors. The MARS algorithm works by using the cubic spline function, namely the interpolation of predictor data or known as the basic function. This basic function is used as a connectivity representation function between the predictor variables, namely the VI value and the target variable i.e. the value of VI prediction or the value of VI 2021. Testing of the accuracy and validation of the VI prediction of 182 villages in the study area using the CART algorithm is shown in Table 6. The operation of CART Algorithm is based on the Decision Tree concept used for VI prediction purposes using historical data. In this

study, the CART algorithm evaluates the response and independent-dependent correlation between VI attributes which include NDVI, MSAVI, NDWI, MNDWI, and NDBI. The data for each VI is evaluated by partitioning it to achieve the most homogeneous condition according to the characteristics of the training data. The new, more homogeneous data resulting from the partitioning process from data VI is data of the CART RF prediction which this is an algorithm that works by forming a set of decision tree classifiers, which each classifier is built randomly using 70% of the total VI data as bootstrap, and the remaining 30% is used as out-of-bag data for internal error estimation. Each classifier decision tree will provide one vote from data VI, which the votes collection from will form a new class functioning as the prediction result data. VI prediction using RF algorithm is done because this algorithm can handle a large amount of VI time series data and minimize data of outliers and noise. Testing of the accuracy and validation of VI prediction of 182 villages in the study area using the RF algorithm is shown in Table 6. K-nn is an algorithm of classification that works in two stages, namely (1) learning the training data using 70% of the original data to find the characteristics of the data to be labeled in the form of classes and (2) making a prediction by comparing the schemes in the previous stages to process the new data. The unknown scheme of the new data is labeled and entered in each class according to its closest characteristics. Testing the accuracy of the predicted data from the combination of all VIs aims to determine the accuracy of high vulnerability areas. The test is carried out using 2 classification methods of accuracy, namely (1) scale-dependent measure consisting of RMSE, MSE, MAE, and ME (2) percentage error-based measure, namely MPE and MAPE. The test results for each method are shown in Table 6.

Table 6. The testing results of the accuracy of the VI data prediction using ML

Algorithms	MSE	ME	RMSE	MAE	MPE	MAPE
CART	0,000611	-0,00061	0,024708	0,000611	-0,13369	0,133689
Random Forest	0,001832	-0,00183	0,042796	0,001832	-0,09158	0,091575
MARS	0,041514	0,037851	0,20375	0,041514	1,739927	1,984127
k-nn	0,0001	0,0001	0,0004	0,0003	0,0001	0,0002

Table 6 shows that the accuracy test produces values close to 0 so it can be interpreted that the prediction results of VI have a high accuracy. The high accuracy of VI prediction indicates that the future LULC will have characteristics that are not much different from the current or previous LULC characteristics. The characteristics of LULC include the health condition of the vegetation, the proportion of built-up lands, surface water bodies and open lands. Table 6 shows that the CART, k-nn and Random Forest algorithms show the smallest value so that it can be represented as the most accurate VI prediction compared to other algorithms.

The VI spatial distribution as a representation of LULC is predicted using the Ordinary Kriging method. The predicted value of NDVI in 2021 was classified into 3 categories based on the level of vulnerability of LULC as shown in Table 7. and Figures 5 (a). NDVI is a vegetation index which plays as an indicator of vegetation density in LULC. Vegetation density is related to the ability to absorb tsunami energy and withstand various solid materials carried by the tsunami waves. The prediction results using the CART, k-nn and random forest methods for NDVI data in 2021 show that most areas had vegetation densities between (-0.1) to (0.3) with moderate to high LULC tsunami vulnerability, which shown in Figures 5 (b), (c) and (d) in yellow color (moderate) and red color (high). NDVI prediction in 2021 using CART, k-nn, and Random Forest shows a decrease of vegetated land areas compared to VI in 2020 (shown as blue color) which is in the range of 0.1 up to 0.5 see Figure 5 (b), (c) and (d). The area that gets wider in high LULC vulnerability classification illustrates the use of that area for socio-economic, agricultural and urban activities within a distance of < 2404 from the coastline[48].

Table 7. Classification of NDVI prediction of 2021 based on tsunami vulnerability and tsunami risk

Interpretation	Range of NDVI Value	Vulnerability	Color
Low LULC Green Vegetation Density	(-0.1) – 0.1	High	Red
Moderate LULC Green Vegetation Density	0.1 – 0.3	Moderate	Yellow
High LULC Green Vegetation Density	0.3 – 0.5	Low	Blue

The effect of soil background reflectance on NDVI is reduced by adding the factors of canopy density and soil elevation represented in the form of VI with the name of MSAVI. The MSAVI value in 2021 was predicted to be in the range between (-0.2) and - 0.4. This value represents a high vegetation canopy density (maximum 1) and a low vegetation canopy density (minimum -1) dominated by low vegetation

canopy density and soil slope. The classification of the MSAVI 2021 prediction which is based on the Level of Tsunami Risk and Vulnerability can be seen in Table 8.

Table 8. Classification of MSAVI prediction of 2021 based on tsunami vulnerability and risk

Interpretation	Range of MSAVI	Vulnerability	Color
Low LULC Vegetation Density and Land Slope	(-0.8) – (-0.4)	High	Red
Moderate LULC Vegetation Density and Land Slope	(-0.4) – 0.0	Moderate	Yellow
High LULC Vegetation Density and Land Slope	0.0 – 0.4	Low	Blue

Based on this study, it can be seen that the k-nn and Random Forest algorithms show a spatial pattern similar to the LULC condition in 2020. Most of the areas have a range of density from (-0.4) - 0.0 to 0.0 - 0.4 with topographic characteristics show vegetation lands with moderate up to high land sloping and with the prediction of LULC vulnerability towards tsunami waves from low to moderate Table 9. The LULC spatial distribution of vulnerability to tsunami waves has patterns as shown in Figure 5 (g) and (h). The CART algorithm predicts that most areas have moderate to low LULC vegetation density and land sloping. The level of vegetation density and land sloping is in the range of (-0.4) - (0.0) to (-0.8) - (-0.4). The topographical characteristics show that the vegetation land has a moderate to low sloping and the level of vulnerability of LULC to tsunami waves is predicted to be moderate to high (Table 9). The LULC spatial distribution of vulnerability to tsunami waves has a pattern as shown in Figure 5 (b, g, i and q).
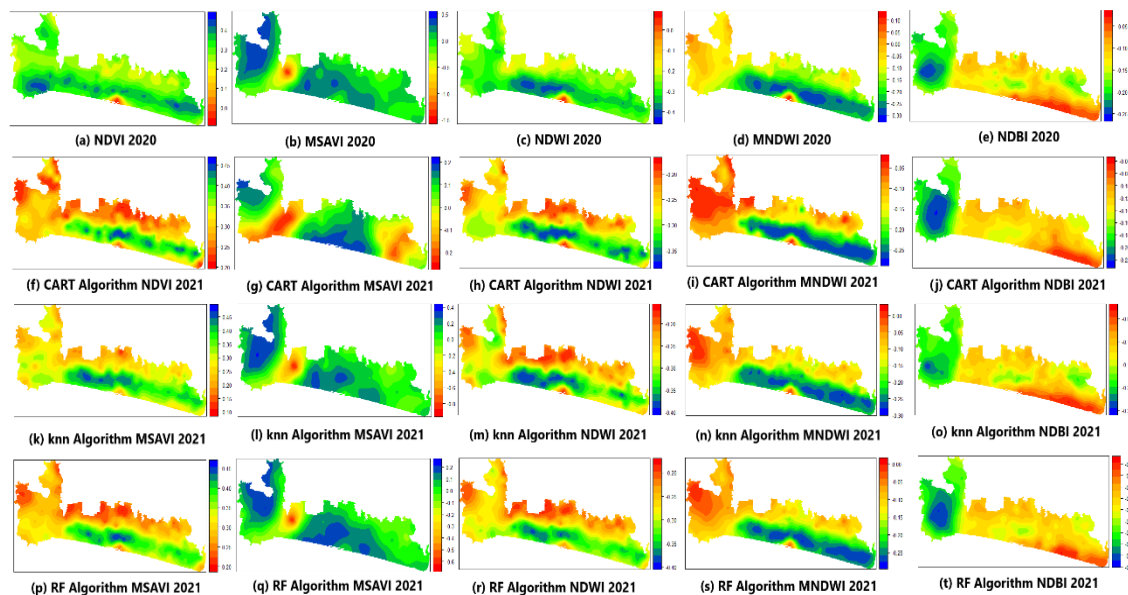

Figure 5. Analysis of ordinary kriging of vegetation indices data on prediction using CART, k-nn, and random forest algorithms

The NDWI and MNDWI are VIs developed for the purpose of measuring moisture content in vegetation, soil moisture, detecting open water, and detecting soil damage caused by liquefaction. Open water detection is carried out using high reflectance water in the green band and low reflectance in the NIR band [53]. The results show that the range of value from (0.1) to (-0.4) with the positive value of open water and the negative value of built-up lands, soil surface, and vegetation (see Table 10 and Figure 5 (a), (f), (k) and (p)).

As seen in Figures 7 (h), (m) and (r) most of the study area in 2021 show built-up lands, land surfaces, and vegetation, which can be concluded in Table 9. The topographic characteristics show that a small part of the areas consists of open water with a high risk of tsunami waves (> 0.1), another part of the areas consists of built-up lands and land surfaces with a range of values from (-0.1) up to (-0.3) with a moderate vulnerability, and some part of the areas consists of vegetation lands with values > (-0.3) and a low vulnerability of tsunami waves.

Table 9. Land slope classes and LULC characteristics

Slope (%)	LULC Characteriatics
< 5.00	Green vegetation, open land, water bodies in the form of beaches and shrimp ponds, and
5.00 – 10.00	built lands.
10.00 – 15.00	Open lands, limestone hills, low plant vegetation, built-up lands.
15.00 – 20.00	Low plant vegetation to forest, limestone hills, and built-up lands.
20.00 – 25.00	
> 25.00	

Table 10. Classification of NDWI prediction of 2021 based on tsunami vulnerability and risk

Interpretation	Range of NDWI/MNDWI	Vulnerability	Color
Open Water and Water Body	> 0.1	High	Red
Built-up Land and Soil Surface	(-0.1) – (-0.3)	Intermediate	Yellow
Vegetation Land	> (-0.3)	Low	Blue

NDBI is referred to as the urban index or index which represents an area that functions as a place of settlement, concentration and distribution of government services, social services, and economic activities. The results show that the range of NDBI values is from (-0.2) to (-0.6) with the interpretation that most of the study area are built-up lands (Figure 5 (e), (j), (o) and (t)). As seen in Figures 5 (f), (k) and (p), it can be seen that in 2021 most of the study area are built-up land which can be concluded in Table 5. Topographic characteristics show that most of the areas are built-up lands with a high risk of tsunami waves (> 0.1), other part of the areas are built-up lands and land surfaces with a range of values from (-0.1) to (-0.3) with moderate risk of tsunami waves, and other part of the areas are vegetation lands with a range of values of > (-0.3) and low risk of tsunami waves. The acombine of algorithms for generate new values (p_{pred}) of vegetation indices are:

```
create number tree (c)
for each c
        for i = 1...n to p do
        select i training set data from data sample
        grow tree from training set data (p)
        select predictor set data from tree node (p_pred)
        build tree node (p_pred)
Build tree node (p_pred)
        else
                bulid m new node in r tree node, r1, r2, ... rn
        for i = 1...n to r do
                values match p_pred1 , p_pred2 ... p_predn with r1, r2, ... rn
                Build tree node class (p)
                Build tree node class (r)
        end
for each c
        for i = 1...n to c do
        calculate distance d ( c, s)
end
        for i = 1...n to f do
        calculate set f for d1 < d ( c, s)
end
        for each c
                for i = 1...n to p do
        build tree node (p)
        Node (1d - 1) + Node(1d)+...Node(1d+1)=1-1d+1
        else
        Node(1i+1)=-1
        Node(1i+1)=-1
End
```

The DEM in this study was used to predict the slope angle and assess the stability of the coastal slope. Altitude is a variable in the elevation model so it is the first derivative of elevation, which is calculated to measure elevation variations over a distance [54]. SRTM DEM is the result of the transformation from geographic coordinates to Cartesian coordinates in the UTM projection system. All DEM pixels are classified into 6 classes as shown in Table 9. As seen in Table 9, each slope class has a different LULC so that it has a different level of vulnerability. The VI distribution is a representation of the LULC analysis based on land

slope as shown in Figure 6. Most of the coastal areas in the study area have a slope of 5-10% with a land composition consists of: green vegetation, open land, water bodies in the form of beaches and shrimp ponds, and built-up land (Figure 6). Other coastal areas with a slope of 10-15% which are located at a distance of 1 to 2.5 km from the coastline, have a land composition consists of limestone hills, low plant vegetation, and built-up lands in the form of houses for residents. Other coastal areas with a slope of > 15% have a land composition consists of limestone hills that overgrown with low plant vegetation to forest and built-up lands in the form of houses for residents.
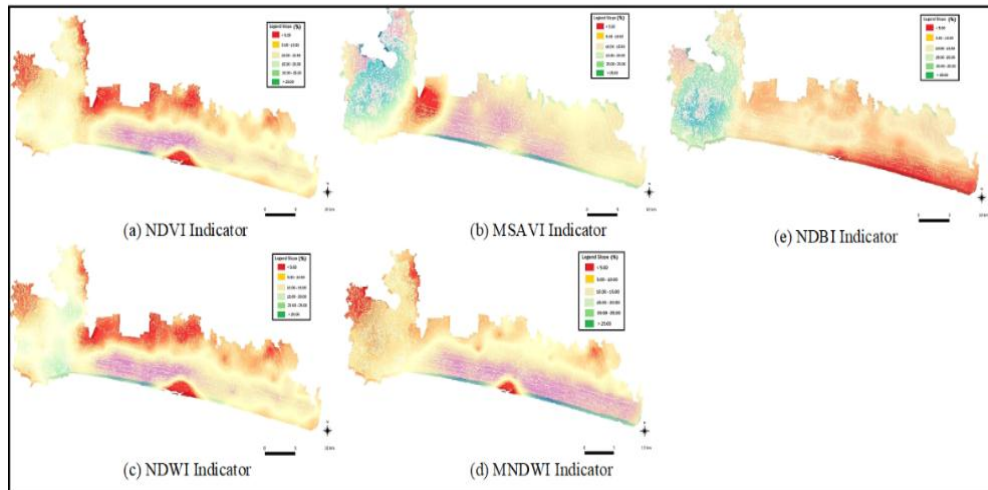

Figure 6. The LULC distribution using the; (a) NDVI, (b) MSAVI, (c) NDWI, (d) MNDWI, (e) NDBI indicators in the land slope classes

5. CONCLUSION

The physical component of vulnerability shows that in LULC the vegetation class in the study area in 2020 decreased to 67% from the original area in 2017, which was the size of 135 km^2. Forest vegetation covering an area of 116 km^2 in 2017 had decreased by 45% to 53 km^2 in 2020. This decrease was caused by changes in land cover from coastal forests to be built-up lands for social, cultural and economic functions with the size of 187 km^2 in 2017. Although in 2020 the built-up lands only covered a size of 51 km^2, there was an increase of open lands covering an area of 86 km^2. On the other hand, the land use for fisheries from an area of 24 km^2 in 2017 increased to a size of 86 km^2 in 2020. The prediction of future changes in LULC was carried out by predicting VI using MARS, CART, RF, and k-nn algorithms. Testing the accuracy of the prediction results shows that the most optimal algorithms for classification and prediction are RF and k-nn. The VI spatial distribution was predicted using the Ordinary Kriging method. The prediction results show that in 2021 most of the area will have a vegetation density from (-0.1) to (0.3) with moderate to high vulnerability and risk of tsunami in LULC as a result of the decrease of vegetation areas. The prediction results in 2021 show a low density of vegetation canopy and a high degree of land surface slope. The prediction results in 2021 shows that there are built-up lands with a high vulnerability of tsunami waves in most of the study area (> 0.1).

REFERENCES

[1] N. Chrysoulakis, E. Flouri, E. Diamandakis, V. Dougalis, C. E. Synolakis, and S. Foteinis, "Remote sensing in support of tsunami mitigation planning in the Mediterranean," *First Int. Conf. Disaster Manag. Emerg. Response Mediterrannean Reg.*, no. August 2014, pp. 1–10, 2008.
[2] A. B. Sambah and F. Miura, "Remote sensing and spatial multi-criteria analysis for tsunami vulnerability assessment," *Disaster Prev. Manag. An Int. J.*, vol. 23, no. 3, pp. 271–295, 2014, doi: 10.1108/DPM-05-2013-0082.

[3]   L. Kurniawan, S. Triutomo, R. Yunus, R. Amri, and A. A. Hantyanto, *Indeks Risiko Bencana Indonesia*, vol. 53. 2013.

[4]   M. Wronna, M. A. Baptista, and J. Götz, "On the construction and use of a Paleo-DEM to reproduce tsunami inundation in a historical urban environment–the case of the 1755 Lisbon tsunami in Cascais," *Geomatics, Nat. Hazards Risk*, vol. 8, no. 2, pp. 841–862, 2017, doi: 10.1080/19475705.2016.1271832.

[5]   C. Chen, S. Yang, and Y. Li, "Accuracy assessment and correction of SRTM dem using icesat/glas data under data coregistration," *Remote Sens.*, vol. 12, no. 20, pp. 1–23, 2020, doi: 10.3390/rs12203435.

[6]   H. Römer *et al.*, "Potential of remote sensing techniques for tsunami hazard and vulnerability analysis-a case study from Phang-Nga province, Thailand," *Nat. Hazards Earth Syst. Sci.*, vol. 12, no. 6, pp. 2103–2126, 2012, doi: 10.5194/nhess-12-2103-2012.

[7]   T. V Bhambure, K. D. Gharde, D. M. Mahale, S. B. Nandgude, M. S. Mane, and T. Bhattacharyya, "Comparative Study of Different Vegetation Indices for Savitri Basin using Remote Sensing Data," *Advanced Agricultural Research & Technology Journal,* vol. II, no. 1, pp. 69–75, 2018.

[8]   M. Viana and J. Rocha, "Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open Source Data in a Rural Region," 2019.

[9]   K. Abolfathi, M. Alikhah-Asl, M. Rezvani, and M. Namdar, "Evaluation of Land Cover Changes Ysing Remote Sensing Technique (Case study: Hableh Rood Subwatershed of Shahrabad Basin)," *Int. J. Agric. Manag. Dev.*, vol. 6, no. 1, pp. 33–42, 2016, doi: 10.22004/ag.econ.262535.

[10]   G. Pasau, G. H. Tamuntuan, and A. Tanauma, "Numerical modelling for tsunami wave propagation (case study: Manado bays)," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 567, no. 1, 2019, doi: 10.1088/1757-899X/567/1/012005.

[11]   K. K. Singh, M. J. Nigam, and K. Pal, "Detection of 2011 Tohoku tsunami inundated areas in Ishinomaki city using generalized improved fuzzy Kohonen clustering network," *Eur. J. Remote Sens.*, vol. 47, no. 1, pp. 461–475, 2014, doi: 10.5721/EuJRS20144726.

[12]   S. Sarun, P. Vineetha, R. Kumar, and V. Jayalekshmi, "Spatial Analysis of Land Use and Land Cover Changes Using Spectral Indices in the Tsunami Affected Areas in Kerala, India," *J. Geogr. Environ. Earth Sci. Int.*, vol. 15, no. 4, pp. 1–11, 2018, doi: 10.9734/jgeesi/2018/41927.

[13]   P. Mondal, X. Liu, T. E. Fatoyinbo, and D. Lagomasino, "Evaluating combinations of sentinel-2 data and machine-learning algorithms for mangrove mapping in West Africa," *Remote Sens.*, vol. 11, no. 24, 2019, doi: 10.3390/rs11242928.

[14]   M. Majidi Nezhad, A. Heydari, L. Fusilli, and G. Laneve, "Land cover classification by using sentinel-2 images: A case study in the city of Rome," *World Congr. Civil, Struct. Environ. Eng.*, vol. 0, no. Cm, pp. 2–9, 2019, doi: 10.11159/iceptp19.158.

[15]   W. Gao, J. Fan, R. C. Easter, Q. Yang, C. Zhao, and S. J. Ghan, "An innovative approach to improve SRTM DEM using multispectral imagery and artificial neural network," *J. Adv. Model. Earth Syst.*, vol. 8, pp. 1289–1309, 2016, doi: 10.1002/2015MS000536.Received.

[16]   Y. Jin, C. L. Yi, and P. Fu, "Evaluation of the Accuracy of SRTM3 and ASTER GDEM in the Tibetan Plateau Mountain Ranges," *E3S Web Conf.*, vol. 206, pp. 1–9, 2020, doi: 10.1051/e3sconf/202020601027.

[17]   Y. Chen *et al.*, "Estimating salt content of vegetated soil at different depths with Sentinel-2 data," *PeerJ*, vol. 8, pp. 1–26, 2020, doi: 10.7717/peerj.10585.

[18]   I. Caballero, J. Ruiz, and G. Navarro, "Sentinel-2 satellites provide near-real time evaluation of catastrophic floods in the West Mediterranean," *Water (Switzerland)*, vol. 11, no. 12, 2019, doi: 10.3390/w11122499.

[19]   S. Pratic, F. Solano, and S. Di Fazio, "Machine Learning Classification of Mediterranean Forest Habitats in Google Earth Engine Based on Seasonal Sentinel-2 Time-Series and Input Image Composition Optimisation," 2021.

[20]   F. Ngadze, K. S. Mpakairi, B. Kavhu, H. Ndaimani, and M. S. Maremba, "Exploring the utility of Sentinel-2 MSI and Landsat 8 OLI in burned area mapping for a heterogenous savannah landscape," *PLoS One*, vol. 15, no. 5, pp. 1–13, 2020, doi: 10.1371/journal.pone.0232962.

[21]   W. Ge, Q. Cheng, L. Jing, F. Wang, M. Zhao, and H. Ding, "Assessment of the capability of sentinel-2 imagery for iron-bearing minerals mapping: A case study in the cuprite area, nevada," *Remote Sens.*, vol. 12, no. 18, 2020, doi: 10.3390/RS12183028.

[22]   L. Hawker, P. Bates, J. Neal, and J. Rougier, "Perspectives on Digital Elevation Model (DEM) Simulation for Flood Modeling in the Absence of a High-Accuracy Open Access Global DEM," *Front. Earth Sci.*, vol. 6, 2018, doi: 10.3389/feart.2018.00233.

[23]   Y. Gorokhovich and A. Voustianiouk, "Accuracy assessment of the processed SRTM-based elevation data by CGIAR using field data from USA and Thailand and its relation to the terrain characteristics," *Remote Sens. Environ.*, vol. 104, no. 4, pp. 409–415, 2006, doi: 10.1016/j.rse.2006.05.012.

[24]   C. G. Brown, K. Sarabandi, and L. E. Pierce, "Validation of the shuttle radar topography mission height data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1707–1715, 2005, doi: 10.1109/TGRS.2005.851789.

[25]   C. S. Gherghina, C. Maftei, and C. Filip, "Assessment of Multi-spectral Vegetation Indices using Remote Sensing and Grid Computing," *Int. J. Comput.*, vol. 5, no. 4, pp. 468–475, 2011.

[26]   A. Karakacan Kuzucu and F. Bektas Balcik, "Testing The Potential of Vegetation Indices For Land Use/Cover Classification Using High Resolution Data," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 4, no. 4W4, pp. 279–283, 2017, doi: 10.5194/isprs-annals-IV-4-W4-279-2017.

[27]   M. Arekhi, C. Goksel, F. B. Sanli, and G. Senel, "Comparative evaluation of the spectral and spatial consistency of Sentinel-2 and Landsat-8 OLI data for Igneada longos forest," *ISPRS Int. J. Geo-Information*, vol. 8, no. 2, 2019, doi: 10.3390/ijgi8020056.

[28]   G. M. Gandhi, S. Parthiban, N. Thummalu, and A. Christy, "Ndvi : Vegetation change detection using remote

sensing and gis – A case study of Vellore District," *Procedia - Procedia Comput. Sci.*, vol. 57, pp. 1199–1210, 2015, doi: 10.1016/j.procs.2015.07.415.

[29] Y. Du, Y. Zhang, F. Ling, Q. Wang, W. Li, and X. Li, "Water bodies' mapping from Sentinel-2 imagery with Modified Normalized Difference Water Index at 10-m spatial resolution produced by sharpening the swir band," *Remote Sens.*, vol. 8, no. 4, 2016, doi: 10.3390/rs8040354.

[30] E. Özelkan, "Water Body Detection Analysis Using NDWI Indices Derived from Landsat-8 OLI," *Polish J. Environ. Stud.*, vol. 29, no. 2, pp. 1759–1769, 2020, doi: 10.15244/pjoes/110447.

[31] D. Yan, C. Huang, N. Ma, and Y. Zhang, "Improved landsat-based water and snow indices for extracting lake and snow cover/glacier in the Tibetan plateau," *Water (Switzerland)*, vol. 12, no. 5, 2020, doi: 10.3390/W12051339.

[32] X. Yang, S. Zhao, X. Qin, N. Zhao, and L. Liang, "Mapping of urban surface water bodies from sentinel-2 MSI imagery at 10 m resolution via NDWI-based image sharpening," *Remote Sens.*, vol. 9, no. 6, pp. 1–19, 2017, doi: 10.3390/rs9060596.

[33] M. S. Malik, J. P. Shukla, and S. Mishra, "Relationship of LST, NDBI and NDVI using landsat-8 data in Kandaihimmat watershed, Hoshangabad, India," *Indian J. Geo-Marine Sci.*, vol. 48, no. 1, pp. 25–31, 2019.

[34] G. Kuc and J. Chormański, "Sentinel-2 imagery for mapping and monitoring imperviousness in urban areas," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 42, no. 1/W2, pp. 43–47, 2019, doi: 10.5194/isprs-archives-XLII-1-W2-43-2019.

[35] Z. Jiang, "Interpretation of the modified soil-adjusted vegetation index isolines in red-NIR reflectance space," *J. Appl. Remote Sens.*, vol. 1, no. 1, p. 013503, 2007, doi: 10.1117/1.2709702.

[36] L. Halounová, "Reclamation areas and their development studied by vegetation indices," *Int. J. Digit. Earth*, vol. 1, no. 1, pp. 155–164, 2008, doi: 10.1080/17538940701782627.

[37] A. J. Prins and A. Van Niekerk, "Crop type mapping using LiDAR, Sentinel-2 and aerial imagery with machine learning algorithms," *Geo-Spatial Inf. Sci.*, vol. 00, no. 00, pp. 1–13, 2020, doi: 10.1080/10095020.2020.1782776.

[38] Ö. Akar and O. Güngör, "Classification of multispectral images using Random Forest algorithm," *J. Geod. Geoinf.*, vol. 1, no. 2, pp. 105–112, 2012, doi: 10.9733/jgg.241212.1.

[39] W. Y. Loh, "Classification and regression trees," *Wiley Interdisc. Rev. Data Min. Knowl. Discov.*, vol. 1, no. 1, pp. 14–23, 2011, doi: 10.1002/widm.8.

[40] A. Htitiou, A. Boudhar, and Y. Lebrini, "The Performance of Random Forest Classification Based on Phenological Metrics Derived from Sentinel-2 and Landsat 8 to Map Crop Cover in an Irrigated Semi-arid Region," 2019.

[41] C. Dewi and R. C. Chen, "Random forest and support vector machine on features selection for regression analysis," *Int. J. Innov. Comput. Inf. Control*, vol. 15, no. 6, pp. 2027–2037, 2019, doi: 10.24507/ijicic.15.06.2027.

[42] S. Park, S. Y. Hamm, H. T. Jeon, and J. Kim, "Evaluation of logistic regression and multivariate adaptive regression spline models for groundwater potential mapping using R and GIS," *Sustain.*, vol. 9, no. 7, 2017, doi: 10.3390/su9071157.

[43] S. Kalbi, A. Fallah, and S. Shataee, "Forest Stand Types Classification Using Tree-Based Algorithms and SPOT-HRG Data," *Int. J. Environ. Resour. Res.*, vol. 2, no. 1, pp. 263–278, 2014.

[44] U. Sakarya, "Hyperspectral dimension reduction using global and local information based linear discriminant analysis," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. II–7, no. October, pp. 61–66, 2014, doi: 10.5194/isprsannals-ii-7-61-2014.

[45] P. Thanh Noi and M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery," *Sensors (Basel).*, vol. 18, no. 1, 2017, doi: 10.3390/s18010018.

[46] S. M. Jog, M. Dixit, A. Rajgopalan, and S. D. Ranade, "KNN and SVM based Satellite Image Classification," *Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng.*, vol. 4, no. 6, pp. 158–162, 2016, doi: 10.17148/IJIREEICE.2016.4637.

[47] S. Zarco-Perello and N. Simões, "Ordinary kriging vs inverse distance weighting: Spatial interpolation of the sessile community of Madagascar reef, Gulf of Mexico," *PeerJ*, vol. 2017, no. 11, 2017, doi: 10.7717/peerj.4078.

[48] V. Yakushev, A. Petrushin, O. Mitrofanova, E. Mitrofanov, V. Terleev, and A. Nikonorov, "Spatial distribution prediction of agro-ecological parameter using kriging," *E3S Web Conf.*, vol. 164, pp. 1–8, 2020, doi: 10.1051/e3sconf/202016406030.

[49] S. S. Baboo and S. Thirunavukkarasu, "Geometric Correction in High Resolution Satellite Imagery using Mathematical Methods: A Case Study in Kiliyar Sub Basin," *Double Blind Peer Rev. Int. Res. J.*, vol. 14, no. 1, pp. 34–40, 2014.

[50] R. Minăˇ, "Radiometric and Atmospheric Corrections of Multispectral μ MCA Camera for UAV Spectroscopy," 2019.

[51] R. S. Ajin, M. K. Jacob, and P. G. Vinod, "Tsunami vulnerability mapping using remote sensing and GIS techniques: A case study of Kollam District, Kerala, India," in *Iranian Journal of Earth Sciences*, 2014, vol. 6, no. 1, pp. 43–50.

[52] M. M. Aburas, S. H. Abdullah, M. F. Ramli, and Z. H. Ash'aari, "Measuring Land Cover Change in Seremban, Malaysia Using NDVI Index," *Procedia Environ. Sci.*, vol. 30, pp. 238–243, 2015, doi: 10.1016/j.proenv.2015.10.043.

[53] H. Baik, Y. S. Son, and K. E. Kim, "Detection of liquefaction phenomena from the 2017 Pohang (Korea) earthquake using remote sensing data," *Remote Sens.*, vol. 11, no. 18, pp. 1–13, 2019, doi: 10.3390/rs11182184.

[54] N. Saleem, M. Enamul Huq, N. Y. D. Twumasi, A. Javed, and A. Sajjad, "Parameters derived from and/or used with digital elevation models (DEMs) for landslide susceptibility mapping and landslide risk assessment: A review," *ISPRS Int. J. Geo-Information*, vol. 8, no. 12, 2019, doi: 10.3390/ijgi8120545.

BIOGRAPHIES OF AUTHORS

Sri Yulianto Joko Prasetyo, completed his doctorate degree on the Doctorate Program of Computer Science, Science Faculty of Gadjah Mada University in 2013. He has been active on research since 2008 until now on the Spatial Data Processing and Remote Sensing. He has published his papers on international journals Scopus Indexed.

Bistok Hasiholan Simanjuntak, completed his doctorate degree on the Doctorate Program of Soil Science, Brawijaya University in 2007. He has been active on research since 2008 until now on the GIS and Remote Sensing Implementation in Agriculture. He has published his papers on international journals Scopus Indexed.

Kristoko Dwi Hartomo, completed his study on the Doctorate Program of Computer Science, Science Faculty of Gadjah Mada University Yogyakarta in 2019. He has been active on research since 2008 until now on Geography Information System. He has published his papers on international journals Scopus Indexed.

Wiwin Sulistyo, completed his study on the Doctorate Program of Computer Science, Science Faculty of Gadjah Mada University Yogyakarta in 2019. He has been active on research since 2018 until now on Geography Information System. He has published his papers on international journals.