Computer model of Tsunami vulnerability using machine learning and multispectral satellite imagery

Sri Yulianto Joko Prasetyo¹, Wiwin Sulistyo¹, Prihanto Ngesti Basuki¹, Kristoko Dwi Hartomo¹, Bistok Hasiholan²

¹Department of Informatic, Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia
²Agriculture and Bussiness Faculty, Satya Wacana Christian University, Salatiga, Indonesia

Article Info

ABSTRACT

This research aims to develop a tsunami vulnerability assessment model on land use and land cover using information on NDVI, NDWI, MDWI, MSAVI, and NDBI extracted from sentinel 2 A and ASTER satellite images. The optimization model using algorithms LASSO and linear regression. The validation test is MSE, ME, RMSE and MAE which show that the linear regression has a higher accuracy than the LASSO. The NDWI interpolation values are 0.00 - (-0.35) and MNDWI interpolation values are 0.00 - (-0.40) which are interpreted as the presence of water surfaces along a coast. MSAVI are values (-0.20) - (-0.35) which are interpreted as the presence of no vegetation. The NDBI interpolation values are values 0.15-0.20 which are interpreted as the presence of built-up lands with social and economic activities. While the NDVI interpolation values are 0.20-0.30 which are interpreted as the presence of vegetation densities, biomass growths from the photosynthesis process, and moderate to low levels of vegetation health. The digital elevation model ASTER analysis shows that all areas with high socioeconomic activities, low NDVI, high NDWI/MDWI, high MSAVI and high NDBI are in areas with low elevation (<10 meters) so they have a high vulnerability to tsunami waves.

Corresponding Author:

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

1. INTRODUCTION

The territory of Indonesia is in a geographical position with a very high disaster complexity, both tectonically, and volcanically. Literature studies show that historically Indonesia has a potential for large Tsunami disasters caused by tsunamigenic megathrust earthquakes, especially in Southern Java, Palu, and Eastern Indonesia [1], [2]. The term Tsunami comes from a Japanese word "*tsu*" (meaning harbor) and "*nami*" (meaning wave). Physically, a Tsunami is defined as a series of ocean waves caused by sudden displacement of a large amount of seawater. The extraordinary volume of sea water movement makes a Tsunami so destructive and it can destroy coastal areas far inland [3]. Literature studies also show that currently computer models of various aspects of tsunamis such as inundation, topography, geomorphology, land use, and coastal cover have been developed using digital elevation model (DEM) data [4], [5]. DEM is a digital cartographic model with regularly spaced intervals covering the x, y, and z axes that reference a vertical datum. DEM provides a 3-dimensional description of the topography of the earth's surface, precise orthorectification of satellite imageries, urban development studies, archeology, topography, Tsunami assessments, glacier observations, geomorphology, plant cover research, 3D spatial analysis, multi-criteria

decision support systems, hydrography modeling, and deformation monitoring [6]. DEM that was originally made through a field survey process, by considering the effectiveness, efficiency and accuracy, and DEM was developed again with remote sensing technology. Remote sensing technologies that are used to develop DEM data are photogrammetry, interferometric synthetic aperture radar (InSAR) and light detection and ranging (LiDAR) [7]. Currently, almost 80% of the world's territories have been mapped using DEM models and the ones that are available for free are the shuttle radar topographic mission (SRTM) and the advanced spaceborne thermal emission and reflection radiometer (ASTER) [8]. Literature studies show that in Japan, Indonesia, Portugal, Mediterranean, and Thailand, SRTM and ASTER images have been used to model various aspects and impacts of tsunamis such as geomorphology, topography, building damages, land elevations, run-up inundations, changes in land use land cover (LULC), and altimetry of tsunami-prone areas [9], [10]. The level of tsunami vulnerability in an area has a strong correlation with the dynamics of changes in LULC throughout the year so that LULC can be used as an indicator in determining areas that are safe for residents [11], [12]. Currently, there are many algorithms that are used to model LULC changes, such as cellular Automata-Markov chain, analytical hierarchy process, analytical equation-based models, statistical models, evolutionary models, cellular models, Markov models, hybrid models, expert system models, and multiagent models [13], [14]. Literature studies show that until now there have not been many publications made by experts on predictions of changes in LULC as a result of tsunamis.

Thus, this research is focused on providing solutions to 3 research problems, are: i) this research needs valid indicators to assess and monitor coastal areas based on land use functionally and economically and based on land cover which refers to biophysical characteristics of the earth's surface, including the distribution of vegetation, water, soil, and other physical features of a land; ii) this research needs computer models to accurately assess the dynamics of LULC based on the classification and prediction of VI data; iii) LULC dynamics can be important information in assessing the level of vulnerability of tsunami impacts.

The purpose of this research is to develop a tsunami vulnerability assessment model on land use and land cover using information of normalized differential vegetation index (NDVI), normalized difference water index (NDWI), modified difference water index (MDWI), modified soil-adjusted vegetation index (MSAVI), and normalized difference built-up index (NDBI) vegetation indices extracted from sentinel 2A and ASTER satellite imageries. Prediction and classification of vegetation indices information was carried out using the least absolute shrinkage and selection operator (LASSO) and linear regression algorithms. The vulnerability of the research area was identified using contour analysis of the ASTER DEM imageries. The research proposed conceptual framework of LULC vulnerability identification are: i) supervised classification methods on sentinel 2A images, ii) classification methods of ASTER with Contour and Hillshade, iii) supervised classification methods of VI to determine LULC, iv) mehods of LASSO and linear regression for vegetation index (VI) prediction, and v) interpolation with ordinary Kriging in determining the unknown data.

2. THEORETICAL FRAMEWORK

ASTER imagery is a product of cooperation between the Japan's ministry of economy, trade and industry (METI) and the U.S. national aeronautics and space administration (NASA). METI and NASA perform scheduling for data acquisition, instrument calibration, archiving, and distribution of data to users. The spatial resolution of the ASTER imagery is 15 m on horizontal plane [15]. ASTER has 3 channels of visible near infra-red (VNIR) spectrum, 6 channels of short wave infra-red (SWIR), and 5 channels of thermal infra-red (TIR) spectrum [16]. ASTER and SRTM imageries combined with VI analysis of sentinel 2A image will produce periodic LULC dynamics information. VI is a mathematical transformation that is combined linearly from the reflectance measurement data of different spectral bands, especially the visible light red, green, and blue (RGB) and near infrared (NIR) bands from remotely sensed images. The mathematical transformation process was carried out in the form of addition, division, and multiplication operations to produce a single value. VI is used for the classification process that separates vegetated and non-vegetated areas, water bodies and land areas, as well as vacant lands and built-up lands on land cover or biomass from image pixels [17], [18]. Visible light and NIR bands are used in VI because they have high sensitivity in detecting the presence of vegetation on the earth's surface [19]. NDVI is one of the VI data which represents the level of greenery or vegetation biomass as an indicator of the health and activities of photosynthesis. The NDVI value is calculated using the VNIR band with wavelengths of 705-865 nm and the red band with a wavelength of 665 nm. The value of the NDVI calculation is range of values -1 and 1. A value of 0 represents no photosynthetic activity or no vegetation and a value of 1 represents a high photosynthetic activity [20]. The NDWI is a type of VI which is used as an indicator to detect the surface of water bodies using the VNIR band with wavelengths of 705-865 nm and the green band with a wavelength of 560 nm and effectively represents the effective surface of water bodies and built lands [21]. The NDWI equation is shown in Table 1.

Table 1. Equation vegetation indices

References	Vegetation indices	Formula	Equation number
[22]	NDVI	$\dfrac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}}$	(1)
[23]	MNDWI	$\dfrac{\rho_{green} - \rho_{MIR}}{\rho_{green} + \rho_{MIR}}$	(2)
[24]	NDWI	$\dfrac{\rho_{green} - \rho_{NIR}}{\rho_{green} + \rho_{NIR}}$	(3)
[25]	NDBI	$\dfrac{\rho_{SWIR} - \rho_{NIR}}{\rho_{SWIR} + \rho_{NIR}}$	(4)
[26]	MSAVI	$\dfrac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red} + L_o}(1 + L_o)$	(5)

The NDBI is a type of VI which represents the built-up areas and other types of land cover. NDBI is determined using the VNIR band with wavelengths of 705-865 nm and the SWIR band with wavelengths of 940-2190 nm [27]. MSAVI is a type of VI applied to overcome the weaknesses of NDVI, especially in relation to large open lands. MSAVI has the ability to reduce the reflectance function of the soil surface and increase the reflectance function of the vegetation canopy density and soil slope. MSAVI requires information on soil brightness factor (L) with values ranging from 0 representing high vegetation cover and 1 representing low vegetation cover. Generally, researchers use medium level of soil brightness factor i.e., 0.5 [28], [29]. Literature study shows that machine learning (ML) is the effective algorithm for predicting and classifying vegetation growth which is represented in the form of VI [30]. ML is a dominant data processing paradigm for extracting information from remotely sensed data. ML works through model development based on prior knowledge as a limited number of labeled samples. However, the basic problem is that sample collection is very expensive and time consuming [31]. Various ML methods that are usually used to predict an earthquake as a tsunami trigger include the LASSO [32], [33]. LASSO is an ML algorithm that works in 2 steps, namely regularization and feature selection. Regularization is done by determining the absolute value of a parameter while regularization is done by shrinking the value of the variable coefficient set to a value of 0. If there is a feature that has a variable coefficient value of not 0, it will be in the regularization process and it will be selected as part of the model. The purpose of this process is to reduce prediction errors [34]. The LASSO equation is shown in (5).

$$min_{\beta_0,\beta}\{||y_i - \beta_0 1_N - x\beta||_2^2\}\, s.t\, ||\beta||_1 \leq \tag{6}$$

Where y_i is the ith observation data, β is the regression coefficient, the notation x is the data matrix, t is the iteration to validate the prediction result [31].

Linear regression is a method to predict the response variable by using two or more independent variables. Assuming the response variable is y, independent variable is x, then $x = (x_0, x_1, \dots x_n), x_0 = 1$, and the regression coefficient is $\beta = (\beta_0, \beta_1, \dots \alpha_n), x_0 = 1$. The regression coefficient is formulated with the (6) [6].

$$y = \beta_0 + \beta_1 x_1 \dots \beta_n x_n + \epsilon \tag{7}$$

Ordinary Kriging (OK) is an interpolation method that works using the spatial autocorrelation principle, which assumes that a point closer to the prediction point has a greater value than the sample point farther from the prediction point. OK is defined using the (7).

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

Where $\gamma(d)$ is the variogram value at d distance, $Z(x_i)$ is the variable value observed at (x_i) location, notation $N(d)$ is the total number of all observation points at d distance. The prediction of variable distribution is calculated using linier regression $Z * (x)$, as shown in (8).

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

Where $\lambda_i(x)$ is the weight, $m(x)$ and $m(x_i)$ are mathematical expectations on random variables $Z * (x)$ and $Z(x_i)$ [35].

3. PROPOSED FRAMEWORK

The main idea of this research is to develop a computer model in identifying the dynamics of LULC, based on changes in the physical environmental of coastal areas using VI indicators. VI consists of NDVI representing vegetation density, MNDWI and NDWI representing water bodies, MSAVI representing open land, and modified normalized difference built-up index (MNDBI) representing built-up land. LULC dynamics are predicted using LASSO and linear regression methods. The spatial distribution of LULC elements is predicted using the ordinary Kriging spatial interpolation method. The overlay of the results of the ordinary Kriging analysis with contour and hillshade results LULC classification of areas that have a high tsunami vulnerability. This concept and algorithms pseudocode are the novelty of this research as shown in Figure 1 (a) to (e). The proposed pseudocode algorithms 1:
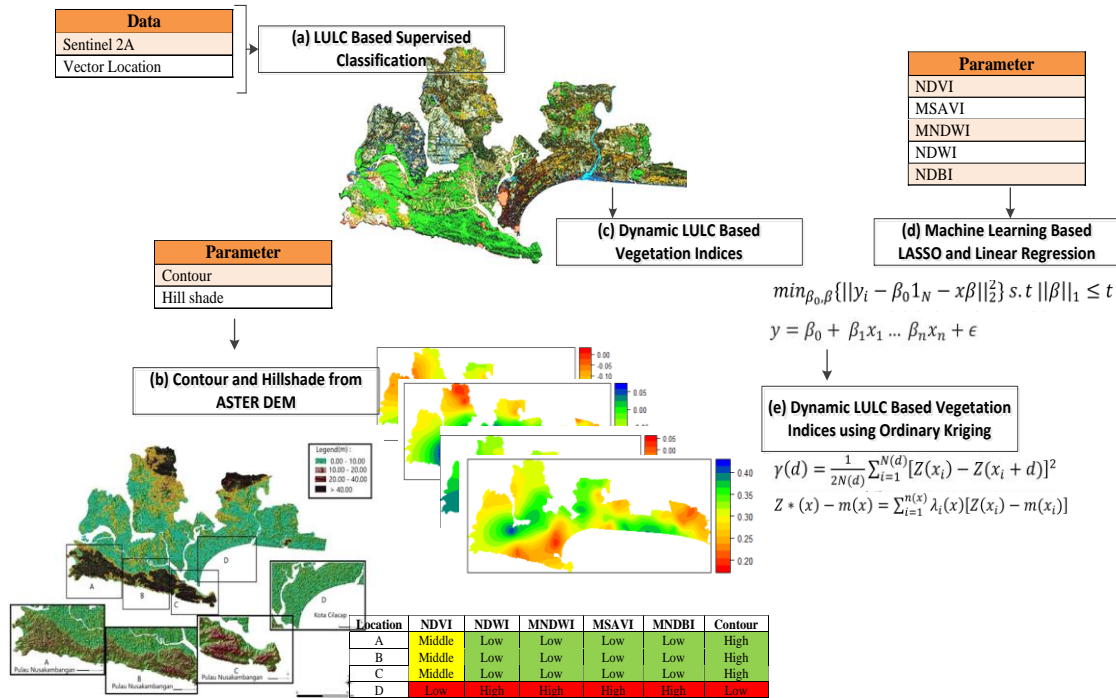
Figure 1. Proposed framework concept for LULC vulnerability identification; (a) the method of supervised classification on sentinel 2A images, (b) classification of ASTER G-DEM with contour and hillshade, (c) guided classification of VI to determine LULC, (d) the method of LASSO and linear regression for VI prediction, and (e) interpolation with ordinary Kriging in determining the unknown data

The proposed pseudocode algorithms 1

```
Begin
        Numeric Data Sentinel-2 ESA = {M_L, Q_L, A_L, M_ρQ_cal, A_ρ}
        Numeric Aster G-DEM = {contour, Hillshade}
        Radiometric correction process:
               Step 1: Conversion DN to TOA Radiance: L_λ = M_LQ_cal + A_L
               Step 2: Conversion DN to TOA Reflectance: ρλ`= M_ρQ_cal + A_ρ
        Atmospheric Correction process:
        ρ*(λ) =ρr(λ) + ρa(λ) + ρra(λ) + T(λ) ρg(λ) + t(λ) ρwc(λ) + t(λ) ρBOA(λ)
        Geometric Corection process
        Vegetation Indices process:
               NDVI = (ρ_NIR − ρ_red)/(ρ_NIR + ρ_red) ; MNDWI = (ρ_green − ρ_MIR)/(ρ_green + ρ_MIR) ; NDWI = (ρ_green − ρ_NIR)/(ρ_green + ρ_NIR) ; NDBI = (ρ_SWIR − ρ_NIR)/(ρ_SWIR + ρ_NIR)
               MSAVI = (ρ_NIR − ρ_red)/(ρ_NIR + ρ_red+L_o) (1 + L_o)
        Lasso process: min_{β_0,β}{||y_i − β_0 1_N − xβ||_2^2} s.t ||β||_1 ≤ t
        Linear Regression: y = β_0 + β_1 x_1 … β_n x_n + ϵ
        Ordinary Kriging: (d) = 1/(2N(d)) Σ_{i=1}^{N(d)}[Z(x_i) − Z(x_i + d)]^2 ; Z*(x) − m(x) = Σ_{i=1}^{n(x)} λ_i(x)[Z(x_i) − m(x_i)]

        Accuration Prediction Count: MSE = Σ (Y'−Y)^2/n ; MAE = Σ |Y'−Y|/n ; RMSE = √((Y'−Y)^2/n)
        Spatial Interpolatioan ← Aster G-DEM
        Display matrix Higher Vulnerability, Middle Vulnerability, Lower Vulnerability
```

4. RESEARCH METHOD

4.1. Research location

This research was conducted in areas that have a high Tsunami vulnerability, namely the areas of Cilacap Regency of Central Java Province, Indonesia. Those areas are shown in Table 2. The determination of high tsunami susceptibility areas is based on the regulation of Peraturan Daerah Kabupaten Cilacap Nomor 9 Tahun 2011 tentang rencana tata ruang Wilayah Kabupaten Cilacap tahun 2011-2031 (in Indonesian). The existing document informs that the total area with a high Tsunami risk is 5,856 Ha. The Tsunami-prone area in this research is defined as an area with a low elevation beach and/or an area that has capability or has experienced a Tsunami which includes areas in 7 sub-districts as shown in Table 2 and Figure 2. The research area has various land uses, namely: vegetation, industrial areas, rice fields, fisheries, residential areas, and beaches.

Table 2. Areas with a high Tsunami vulnerability in Cilacap Regency of Central Java Province, Indonesia

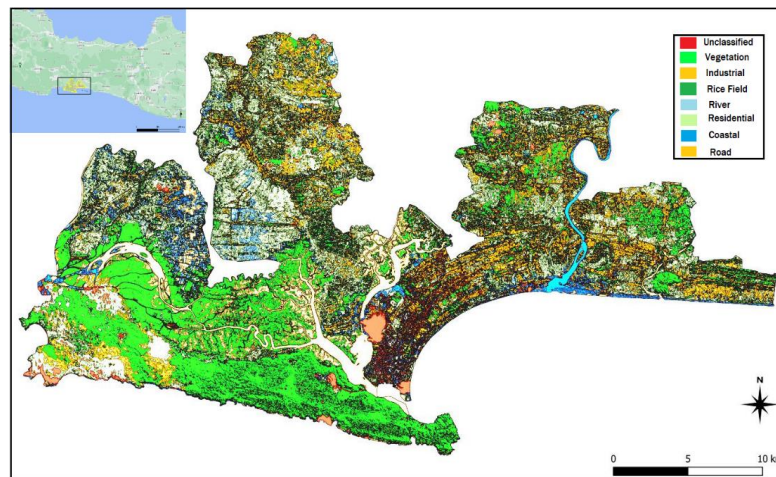| No	Name of Sub-District	Latitude	Longitude
1	Adipala Sub-District	-7.6680	109.1642
2	Kesugihan Sub-District	-7.6201	109.0810
3	Cilacap Utara Sub-District	-7.6690	109.0275
4	Cilacap Tengah Sub-District	-7.6865	108.9859
5	Cilacap Selatan Sub-District	-7.7497	108.9621
6	Kawunganten Sub-District	-7.5961	108.9265
7	Kampunglaut Sub-District	-7.6712	108.8670

Figure 2. Research area consisting 7 sub-districts in Cilacap Regency with a high Tsunami vulnerability

4.2. Research data

Research data is classified into two types of DEM data, namely: i) ASTER. ASTER data is used to determine the aspect, slope, and contour, ii) VI data is extracted from sentinel 2A images. Data VI consists of 5 types, namely NDVI, NDWI, MNDWI, NDBI, and MSAVI. Data VI is used to determine the characteristics and dynamics of LULC patterns.

4.3. Stages of experimental procedure

Figure 3 shows stages of experimental procedure by; i) DEM data preprocessing. This stage is performed by downloading the ASTER image data through the portal https://earthexplorer.usgs.gov/ and selecting the research area according to its coordinates, ii) preprocessing satellite image data. This stage is performed by downloading sentinel 2 A image data from portal https://earthexplorer.usgs.gov/, selecting the research area according to its coordinates and performing atmospheric, radiometric, and geometric corrections, iii) analysis and interpretation of DEM data. This stage is performed by determining hillshade and elevation, iv) the VI time series data extraction. This stage is carried out using NDVI, NDWI, MNDWI, NDBI, and MSAVI algorithms, v) prediction of VI data. This stage is done using LASSO algorithm and linear regression, vi) the accuracy test of the prediction results. This stage is performed using MSE, ME, RMSE, and MAE methods, vii) prediction VI results are mapped in areas with high tsunami susceptibility using the ordinary Kriging method.
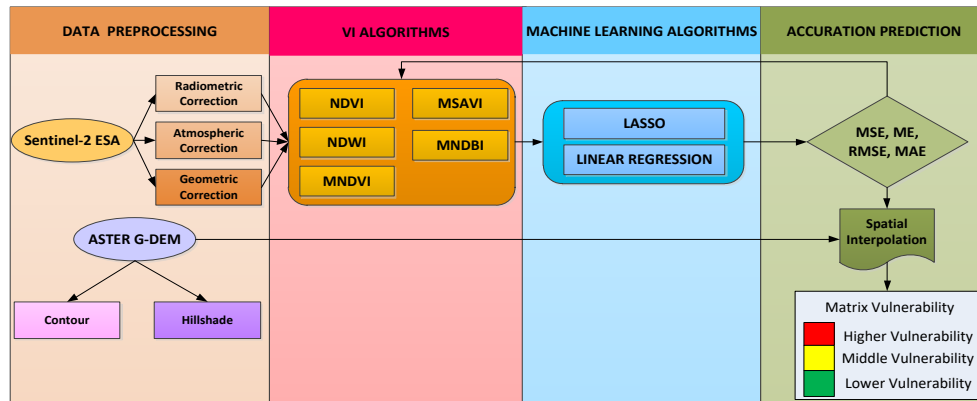
Figure 3. Experimental procedure of the satellite imagery for Tsunami inundation framework based on vegetation indices

5. RESULTS AND DISCUSSION

The composition and structure of vegetation in LULC is dynamic, which they fluctuate over a long period of time following seasonal patterns and they can also fluctuate in a short time according to anthropogenic pressures. The computational experiment of sentinel 2A band images in time series show that the composition and structure of vegetation in LULC are determined using NDVI, NDWI, MNDWI, MSAVI, dan NDBI [8]. Reference source shows that VI and terrain aspects (slope, elevation, and aspect) are the essential parameters used to monitor the intensity of anthropogenic dynamics of coastal beaches which are caused by anthropogenic activities [9], [36].

The changes in NDVI value are in line with the long-term seasonal pattern observational data, LASSO prediction and linear regression prediction (Figures 4 (a) to (c)) indicates a change in land use which affects on vegetation density. The changes in vegetation density caused by natural cycle of plant seasonal and society's socio-economy activities.
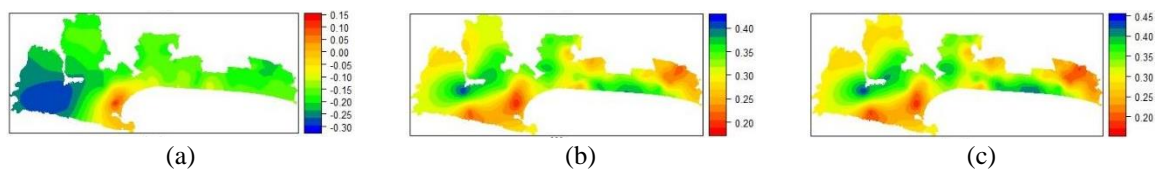
Figures 4. The NDVI spatial pattern for (a) observational data, (b) LASSO prediction data, and (c) linear regression prediction data are indicates a change in land use which affects on vegetation density caused by natural cycle of plant seasonal and society's socio-economy activities

The results show that the NDVI value is in the range of 0.20-0.30 which is interpreted that the study area has vegetation density, biomass growth from photosynthesis, various canopy formations, and has a moderate to low vegetation health. Tsunami propagation was affected by elevation, coastal topography and coast-to-land distance which can be extracted from DEM images [37]. The coastal topography is dominated by mangroves or pine forests and the built environment so that to a certain extent it has an influence on propagation, inundation, and absorption of Tsunami energy [38]. Coastal topographic structures can be identified and predicted using VI, namely NDVI, MNDWI, NDWI, MSAVI, and NDBI. The use of VI is based on the following considerations: i) VI focuses on certain types of land cover (vegetation type, water body, and soil surface), ii) the disturbing background effects can be reduced, and iii) atmospheric distortion effect caused by the angle of the sun and the angle of view of the sensor can be reduced [37].

The MNDWI indices in observational data, LASSO prediction data and linear regression data (Figures 5(a) to (c)) are calculated using green band and SWIR band spectrums. The MNDWI value is in a range of -1 to +1 where water body is indicated by pixels between 0 and +1. The separation of water pixels from non-water pixels is made using the image binarization method, namely the use of the optimal threshold for converting the spectrum index into two classes, namely discrete 0, binary 1. The algorithm converts the pixel value of an image less than 0 with a value of 0 and converts the pixel value of an image more than 1 with a value of 1. With this concept, all pixels with a code of 1 represent water bodies and those pixels with a code of 0 represent non-water bodies [39].

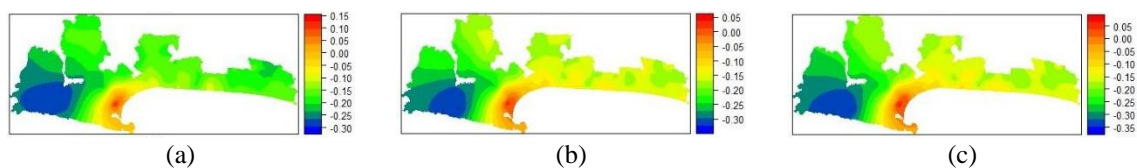(a)   (b)   (c)
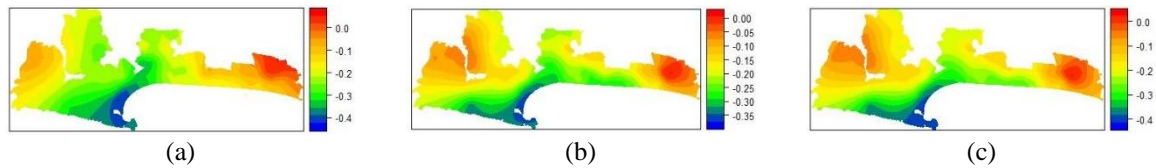
Figures 5. The NDWI spatial pattern for (a) observational data, (b) LASSO prediction data, and (c) linear regression data are affected by long-term seasonal pattern which occurred naturally, short-term seasonal pattern which occurred because of socio-economy activities of the coastal society

Figures 5(a) to (c) show blue, green, yellow to red colors with an interpolation value of 0 - (-0.3) which represent the water surface of shrimp ponds and rice fields along a coast. The changes in NDWI and MNDWI values are also affected by long-term seasonal pattern which occurred naturally, short-term seasonal pattern which occurred because of socio-economy activities of the coastal society, such as in stocking season and harvest season in fishery industry as shown in Figures 6(a) to (c). The NDWI is one of the indicators to determine if there is a wide water body and water puddles such as rivers and rice fields during the rainy season. Figures 6(a) to (c) are in the range of 0.00 - (-0.40) and NDWI is in the range of 0.00 - (-0.35) which represent water surfaces of shrimp ponds and rice fields. All coastal areas show land uses for economic, social, and residential activities.

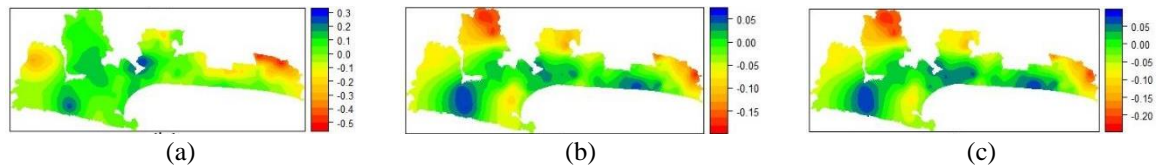(a)   (b)   (c)
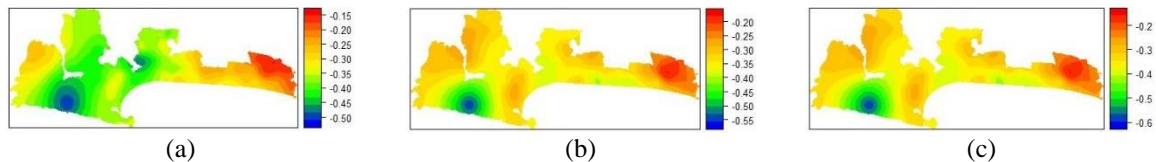
Figures 6. The MNDWI spatial pattern for (a) observational data, (b) LASSO prediction data, and (c) linear regression prediction data are also affected by long-term seasonal pattern which occurred naturally, short-term seasonal pattern which occurred because of socio-economy activities such as in stocking season and harvest season in fishery industry

The changes in short-term NDBI value shows the changing in the land use which is massively conducted in the coastal area as physical development activities on the coastal area as shown in Figures 7(a) to (c). The physical development is the land-use change from open land to buildings, mesh networks, and fishery industry, such as shrimp embankments.

(a)   (b)   (c)

Figures 7. The NDBI spatial pattern for (a) observational data, (b) LASSO prediction data, and (c) linear regression prediction data. The changes in short-term NDBI value shows the changing in the land use which is massively conducted in the coastal area as physical development activities on the coastal area

MSAVI in Figures 8(a) to (c) is the vegetation index used as a correction for NDVI index to remove the effect of a background in the calculation of VI. MSAVI has a higher dynamic response of vegetation and lower background variation when compared to NDVI and SAVI. MSAVI is determined using band 3 (red) and band 4 (NIR) reflectance. A high MSAVI value indicates a good and healthy vegetation, while a value close to or equal to -1 indicates no vegetation, i.e., an open land or barren land [39], [40].

Figures 8. The MSAVI spatial pattern for (a) observational data, (b) LASSO prediction data, and (c) linear regression prediction data is the vegetation index used as a correction for NDVI index to remove the effect of a background in the calculation of VI

MSAVI in Figures 8(a) to (c) shows that it is in the range -0.20 to -0.35 which indicating low values, and it is interpreted that the study areas are open lands and have no vegetation. NDBI is the VI used to map built-up lands with the range of values 0.15-0.20 indicating low values, and it is interpreted that most of the study areas are built-up lands. The NDBI value lies between -1 to +1, where a negative value represents a water body and a positive value represents a development area.

Based on the VI historical data, the following experiment is a prediction to find out the VI data in one period in the future using the linear regression method. In some references, the linear regression method shows good accuracy in predicting VI from sentinel images in its relations to the land coverage [41]. The data results of VI prediction are classified based on the grouping or data class using the LASSO algorithm. The classification and prediction results are tested for their accuracy and validation using various statistical methods, such as the MSE, ME, RMSE, and MAE. The MSE, ME, and MAE testing results which are close to zero will have higher accuracy, on the other hand, the results which are not close to zero have lower accuracy level (Table 3 and Table 4). In order to find out the spatial distribution pattern in every VI, the ordinary Kriging (OK) method is used. OK is a method to determine the unknown VI value because it is not the sample points [42]. The visualization of the spatial pattern using the OK method in the VI data and VI prediction data is presented in Figure 4(a), (b), (c) to Figure 8(a), (b), (c).

Table 3. Results of accuracy and validation tests of LASSO algorithm using MSE, ME, RMSE, and MAE

	NDVI	MSAVI	NDWI	MNDWI	NDBI
Minimum	0.172	-0.183	-0.605	-0.395	-0.324
Average	0.296	-0.036	-0.313	-0.171	-0.159
Maximum	0.433	0.058	-0.147	0.072	0.038
MSE	0.004	0.018	0.003	0.004	0.001
ME	0.005	0.003	-0.001	-0.002	0.002
RMSE	0.065	0.135	0.055	0.067	0.027
MAE	0.050	0.109	0.046	0.055	0.019

Table 4. Results of accuracy and validation tests of linier regression algorithm using MSE, ME, RMSE, and MAE

	NDVI	MSAVI	NDWI	MNDWI	NDBI
Minimum	0.155	-0.226	-0.651	-0.440	-0.350
Average	0.297	-0.036	-0.313	-0.171	-0.159
Maximum	0.454	0.077	-0.124	0.098	0.067
MSE	0.004	0.018	0.003	0.004	0.001
ME	0.004	0.003	0.000	-0.002	0.002
RMSE	0.066	0.134	0.056	0.067	0.025
MAE	0.048	0.108	0.044	0.053	0.018

Coastal topography in the study area can be visualized using hillshade method as shown in Figures 9. Hillshade is a method for visualizing the relief of the earth's surface with raster data sources DEM in 2-dimensional format by adding light to make it appears as a 3-dimensional object. The purpose of using the hillshade method is to sharpen relief visualization of the earth's surface. The hill shade and elevation visualizations are shown in Figure 9(A, B, and C). representing the characteristic model of the relief and elevation of earth's surface in 3D format as the background where vegetations NDVI, water bodies MNDWI, open lands MSAVI, and built lands MNDBI are located. Figures 9(A, B, and C) shows a visualization of the terrain on the coast of Nusakambangan Island which is dominated by hills which theoretically can protect the island from high tsunami waves. Elevations in A, B, and C indicate that the points are at an altitude > 40 meters above sea level Figure 5. These areas are more protected from a tsunami because these areas are hilly and heavily forested. Land use and socio-economic activities on Nusakambangan Island are lower when compared to that of Cilacap City (D). Cilacap City and its coastal areas have high land use and socio-economic activities. However, Cilacap City shows the existence of open coastal terrains without natural protection of hills. The tsunami barriers of Cilacap City are buildings and city park's vegetations. The correlation between VI and the studied areas' elevation can be represented in a relation matrix of the components to determine the affected areas, as seen in Table 5. The Nusakambangan Island areas (A, B, and C) have middle to high vegetation levels, although the island's water surface, open space, and buildings are on the low level, and it has high elevations in the form of hills. Area D is Cilacap City which has low vegetation level, but it has a wide water surface, open space, and built-up lands.

Table 5. Relation matrix of location, VI and contour components to determine the vulnerability degree of the affected areas

Location	NDVI	NDWI	MNDWI	MSAVI	MNDBI	Contour
A	Middle	Low	Low	Low	Low	High
B	Middle	Low	Low	Low	Low	High
C	Middle	Low	Low	Low	Low	High
D	Low	High	High	High	High	Low
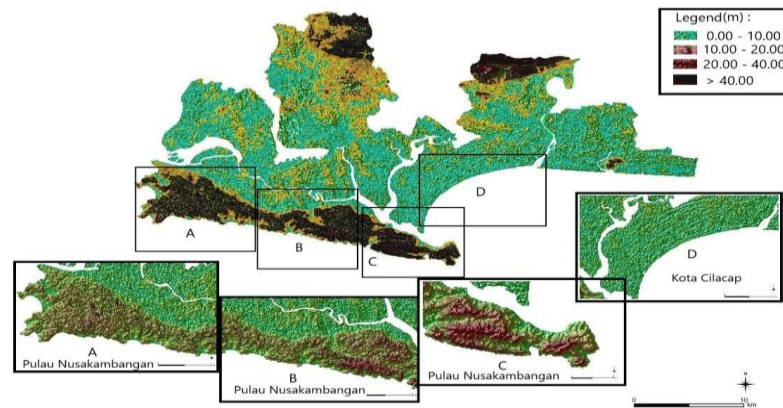
Figure 9. Characteristic model of the earth's surface relief and elevation at 4 observational points

6.   CONCLUSION

The tsunami vulnerability assessment model on land use and land cover using information of NDVI, NDWI, MDWI, MSAVI, and NDBI vegetation indices extracted from sentinel 2 A and ASTER satellite imageries shows an effective performance. The test of machine learning algorithm, namely LASSO and linear regression and the test of spatial interpolation, namely Universal Kriging by using MSE, ME, RMSE, and MAE statistical methods show a high accuracy. Spatial interpolation analysis using Universal Kriging shows that the linear regression algorithm is closer to the observation data than that of the LASSO algorithm. Interpolation NDWI 0.00-(-0.35) and MNDWI 0.00-(-0.40) are interpreted as water body or water surface along a coast. MSAVI (-0.20)-(-0.35) is interpreted as open land and no vegetation. NDBI 0.15-0.20 is interpreted as built up land with social and economic activities. NDVI 0.20-0.30 is interpreted as vegetation density, biomass growth from photosynthesis, various canopy formations, and moderate to low level of vegetation health. The DEM ASTER analysis shows that there are high socioeconomic activities, low vegetation densities, large water bodies, wide open lands, and large buildings located in an area with low elevation (<10 meters) so that it has a high vulnerability to tsunami waves. The relations of VI and the studied area elevation can be represented in a relation matrix between components. Nusakambangan Island (A, B, and C) has middle to high levels of vegetation, but its water surface, open space and buildings are in the low levels, and it has high elevation level in form of hills. The D area is Cilacap City which has low vegetation density level, but there are wide areas of water surface, open space and built-lands.

ACKNOWLEDGEMENTS

REFERENCES

[1]   I. R. Pranantyo, M. Heidarzadeh, and P. R. Cummins, "Complex tsunami hazards in eastern Indonesia from seismic and non-seismic sources : Deterministic modelling based on historical and modern data," *Geosci. Lett.*, 2021, doi: 10.1186/s40562-021-00190-y.
[2]   H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3025–3033, 2006, doi: 10.1080/01431160600589179.
[3]   F. Yamazaki, K. Kouchi, and M. Matsuoka, "Tsunami damage detection using moderate-resolution satellite imagery," *8th US Natl. Conf. Earthq. Eng. 2006*, vol. 13, no. 465, pp. 7532–7541, 2006.
[4]   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.

[5]   C. Wang, J. Qi, and M. Cochrane, "Assessment of tropical forest degradation with canopy fractional cover from Landsat ETM+ and IKONOS imagery," *Earth Interact.*, vol. 9, no. 22, pp. 1–18, 2005, doi: 10.1175/EI133.1.

[6]   Y. M. Choo, K. H. Chun, H. S. Jeon, and S. B. Sim, "A predictive model for estimating damage from wind waves during coastal storms," *Water (Switzerland)*, vol. 13, no. 9, 2021, doi: 10.3390/w13091322.

[7]   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.

[8]   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.

[9]   A. S. Alademomi, C. J. Okolie, O. E. Daramola, R. O. Agboola, and T. J. Salami, "Assessing the relationship of LST, NDVI and EVI with land cover changes in the Lagos Lagoon environment," *Quaest. Geogr.*, vol. 39, no. 3, pp. 87–109, 2020, doi: 10.2478/quageo-2020-0025.

[10]   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.

[11]   K. G. Nikolakopoulos, "Accuracy assessment of ALOS AW3D30 DSM and comparison to ALOS PRISM DSM created with classical photogrammetric techniques," *Eur. J. Remote Sens.*, vol. 53, no. sup2, pp. 39–52, 2020, doi: 10.1080/22797254.2020.1774424.

[12]   S. Islam, A. H. Tanim, and R. A. Mullick, "Land Use and Land Cover Classification of Coastal Districts of Bangladesh in a 10m Resolution of Sentinel-2 Satellite Image," *Proc. Int. Conf. Planning, Archit. Civ. Eng. 9 - 11 Febr. 2019, Rajshahi Univ. Eng. Technol. Rajshahi, Bangladesh*, no. February, pp. 9–11, 2019.

[13]   B. Nath, Z. Wang, Y. Ge, K. Islam, R. P. Singh, and Z. Niu, "Land use and land cover change modeling and future potential landscape risk assessment using Markov-CA model and analytical hierarchy process," *ISPRS Int. J. Geo-Information*, vol. 9, no. 2, 2020, doi: 10.3390/ijgi9020134.

[14]   C. Liping, S. Yujun, and S. Saeed, "Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques—A case study of a hilly area, Jiangle, China," *PLoS One*, vol. 13, no. 7, pp. 1–23, 2018, doi: 10.1371/journal.pone.0200493.

[15]   M. Abrams, R. Crippen, and H. Fujisada, "ASTER Global Digital Elevation Model (GDEM) and ASTER Global Water Body Dataset (ASTWBD)," *Remote Sens.*, vol. 12, no. 7, pp. 1–12, 2020, doi: 10.3390/rs12071156.

[16]   K. Iwao *et al.*, "Validating Global Digital Elevation Models With Degree Confluence Project Information and Aster-Dem on Geo Grid," *Archives*, pp. 1847–1852, 1847.

[17]   C. K. Min, A. Muchtar, A. Bahar, and W. S. Udin, "Landslide Assessment Using Normalized Difference Vegetation Index (NDVI)," vol. 4, pp. 98–104, 2016.

[18]   R. Xu, J. Liu, and J. Xu, "Extraction of high-precision urban impervious surfaces from sentinel-2 multispectral imagery via modified linear spectral mixture analysis," *Sensors (Switzerland)*, vol. 18, no. 9, pp. 1–15, 2018, doi: 10.3390/s18092873.

[19]   A. Ayanlade, "Remote Sensing Vegetation Dynamics Analytical Methods: A Review Of Vegetation Indices Techniques," *Geoinformatica Pol.*, vol. 16, pp. 7–17, 2017, doi: 10.4467/21995923GP.17.001.7188.

[20]   B. Satyanarayana, K. A. Mohamad, I. F. Idris, M. L. Husain, and F. Dahdouh-Guebas, "Assessment of mangrove vegetation based on remote sensing and ground-truth measurements at Tumpat, Kelantan Delta, East Coast of Peninsular Malaysia," *Int. J. Remote Sens.*, vol. 32, no. 6, pp. 1635–1650, 2011, doi: 10.1080/01431160903586781.

[21]   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.

[22]   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.

[23]   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.

[24]   T. D. Acharya, A. Subedi, and D. H. Lee, "Evaluation of water indices for surface water extraction in a landsat 8 scene of Nepal," *Sensors (Switzerland)*, vol. 18, no. 8, pp. 1–15, 2018, doi: 10.3390/s18082580.

[25]   Y. XI, N. X. Thinh, and C. LI, "Preliminary comparative assessment of various spectral indices for built-up land derived from Landsat-8 OLI and Sentinel-2A MSI imageries," *Eur. J. Remote Sens.*, vol. 52, no. 1, pp. 240–252, 2019, doi: 10.1080/22797254.2019.1584737.

[26]   M. A. Zaraza Aguilera, "Classification of land-cover through machine learning algorithms for fusion of sentinel-2A and planetscope imagery," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 42, no. 3/W12, pp. 361–368, 2020, doi: 10.5194/isprs-archives-XLII-3-W12-2020-361-2020.

[27]   S. Yulianto, J. Prasetyo, B. H. Simanjuntak, K. D. Hartomo, and W. Sulistyo, "Computer model for tsunami vulnerability using sentinel 2A and SRTM images optimized by machine learning," vol. 10, no. 5, pp. 2821–2835, 2021, doi: 10.11591/eei.v10i5.3100.

[28]   S. Papaiordanidis, M. Tompoulidou, P. Lefakis, and I. Gitas, "Evaluation of Spectral Indices Efficiency In Burned Area Mapping Using Object- Based Image Analysis," *Geoscience*, vol. 2, pp. 065–072, 2017, doi: 10.3390/rs6076709.

[29]   Farooq Ahmad, "Spectral vegetation indices performance evaluated for Cholistan Desert," *J. Geogr. Reg. Plan.*, vol. 5, no. 6, pp. 165–172, 2012, doi: 10.5897/jgrp11.098.

[30]   X. Li, W. Yuan, and W. Dong, "A machine learning method for predicting vegetation indices in china," *Remote Sens.*, vol. 13, no. 6, 2021, doi: 10.3390/rs13061147.

[31]   S. Bakken, G. Johnsen, and T. A. Johansen, "ANALYSIS AND MODEL DEVELOPMENT OF DIRECT HYPERSPECTRAL CHLOROPHYLL-A ESTIMATION FOR REMOTE SENSING SATELLITES Center for Autonomous Marine Operations and Systems Department of Engineering Cybernetics (1) & Department of Biology (2) Norwegian Universit," no. 1, pp. 2–6.

[32]   I. M. Murwantara, P. Yugopuspito, and R. Hermawan, "Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 18, no. 3, pp. 1331–1342, 2020, doi: 10.12928/TELKOMNIKA.v18i3.14756.

[33]   T. KURITA, "Tsunami Inversion of Fault Displacements Based on Sparse Modeling," *J. Japan Assoc. Earthq. Eng.*, vol. 19, no. 8, pp. 8_1-8_18, 2019, doi: 10.5610/jaee.19.8_1.

[34]   M. Afshar and H. Usefi, "Dimensionality reduction using singular vectors," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021, doi: 10.1038/s41598-021-83150-y.

[35]   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.

[36]   S. GüNthert, M. Wieland, and A. S. Heidelberg, "Change detection analysis for assessing the vulnerability and protective effect of beach forests in case of the Tsunami 2004 in Thailand," *Photogramm. Fernerkundung, Geoinf.*, vol. 2011, no. 4, pp. 247–260, 2011, doi: 10.1127/1432-8364/2011/0086.

[37]   J. Song and K. Goda, "Influence of Elevation Data Resolution on Tsunami Loss Estimation and Insurance Rate-Making," *Front. Earth Sci.*, vol. 7, 2019, doi: 10.3389/feart.2019.00246.

[38]   G. Kaiser, L. Scheele, A. Kortenhaus, F. Løvholt, H. Römer, and S. Leschka, "The influence of land cover roughness on the results of high resolution tsunami inundation modeling," *Nat. Hazards Earth Syst. Sci.*, vol. 11, no. 9, pp. 2521–2540, 2011, doi: 10.5194/nhess-11-2521-2011.

[39]   S. Sharifzadeh, "A Support Vector Machine-Based Water Detection Analysis in a Heterogeneous Landscape Using Landsat TM Imagery Saeideh," *Calif. Geogr.*, vol. 59, 2020.

[40]   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," vol. II, no. 1, pp. 69–75, 2018.

[41]   S. Emami, "Detecting and predicting vegetation cover changes using sentinel 2 Data (A Case Study : Andika Region)," *J. Radar Opt. Remote Sens.*, vol. 2, pp. 38–54, 2018.

[42]   T. G. Pham, M. Kappas, C. Van Huynh, and L. H. K. Nguyen, "Application of ordinary kriging and regression kriging method for soil properties mapping in hilly region of central Vietnam," *ISPRS Int. J. Geo-Information*, vol. 8, no. 3, 2019, doi: 10.3390/ijgi8030147.

BIOGRAPHIES OF AUTHORS

Sri Yulianto Joko Prasetyo 🆔 📇 SC Ⓟ 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 in SIJR 0.8. His main area of interest focuses on geospatial computing. His area of expertise includes machine learning, remote sensing, spatial modeling, and software engineering. His founder of Qua-edutehcno which is a technology based start up for software products higher education quality manajement and technology. He can be contacted at email: sri.yulianto@uksw.edu.

Wiwin Sulistyo 🆔 📇 SC Ⓟ 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. His main area of interest focuses on geospatial computing. His area of expertise includes machine learning, remote sensing, spatial modeling, and computer network management. He can be contacted at email: wiwin.sulistyo@uksw.edu.

Kristoko Dwi Hartomo 🆔 📇 SC Ⓟ 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. His main area of interest focuses on geospatial computing. His area of expertise includes machine learning, remote sensing, spatial modeling, and software engineering. He can be contacted at email: Kristoko@uksw.edu.

Prihanto Ngesti Basuki [ID] [SC] [P] he is student of Doctorate Program of Computer Science, Faculty of Information Technology, Satya Wacana Christian University, in 2019. He has been active on research since 2019 until now on computer model. He can be contacted at email: ngesti@uksw.edu.

Bistok Hasiholan Simanjuntak [ID] [SC] [P] is currently working as Lecturer and Researcher at Department of Agrotechnology, Faculty of Agriculture, Satya Wacana Christian University, Salatiga, Indonesia. He has completed his Ph.D. in Soil Science from University of Brawijaya, Indonesia. His main area of interest focuses on plant and soil sciences. His area of expertise includes soil management, land evaluation, geographyc information system, and remote sensing, soil organic matter, soil conservation, and organic farming. Bistok has 20 publications in Scopus journals as author/co-author. He can be contacted at bhasiholans@uksw.edu.