

Tsunami Vulnerability and Risk Assessment Using Machine Learning and Landsat 8

Gallen Cakra Adhi Wibowo , Sri Yulianto Joko Prasetyo , Irwan Sembiring
Universitas Kristen Satya Wacana, Salatiga, Indonesia

Article Info

Article history:

Received January 11, 2023
Revised February 24, 2023
Accepted March 20, 2023

Keywords:

Landsat 8
Machine Learning
Tsunami
Vegetation Indices

ABSTRACT

Tsunami is a disaster that often occurs in Indonesia, and there are no valid indicators to assess and monitor coastal areas based on functional land use and based on the land cover, which refers to the biophysical characteristics of the earth's surface. One of the recommended methods is the vegetation index. The vegetation index is a method from LULC that can provide information on how severe the tsunami's impact was on the area. In this study, an increase in the vegetation index was carried out using machine learning. This study aimed to develop a tsunami vulnerability assessment model using the Vegetation Index extracted from Landsat 8 satellite imagery optimized with KNN, Random Forest, and SVM. The stages of the study were 1) extraction of Landsat 8 images using algorithms NDVI, NDBI, NDWI, MSAVI, and MNDWI; 2) prediction of vegetation indices using KNN, Random Forest, and SVM algorithms. 3) accuracy testing using the MSE, RMSE, and MAE, 4) spatial prediction using the Kriging function, and 5) tsunami modeling vulnerability indicators. The results of this study indicated that the NDVI interpolation value is 0 - 0.1, defined as vegetation density, biomass growth, and moderate to low vegetation health. The NDWI value was 0.02 - 0.08, and the MNDWI value was 0.02 - 0.09, interpreted as surface water along the coast. MSAVI is a value of 0.1 - 0, defined as vegetation's absence. The NDBI interpolation value was -0.05 - (-0.08), interpreted as the existence of built-up land with social and economic activities. From the research results on the ten areas studied, there were three areas with conditions with a high level of tsunami vulnerability, two areas with medium vulnerability, and five areas with low vulnerability to tsunamis.

Copyright ©2022 The Authors.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Gallen Cakra Adhi Wibowo, +6282266554783
Faculty of Information Technology, Information System,
Universitas Kristen Satya Wacana, Salatiga, Indonesia,
Email: gallenbale11@gmail.com

How to Cite: G. wibowo, S. Prasetyo, and I. Sembiring, "Tsunami Vulnerability and Risk Assessment in Banyuwangi District using machine learning and Landsat 8 image data", *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, no. 2, pp. 365-380, Mar. 2023.

This is an open access article under the CC BY-SA license (<https://creativecommons.org/licenses/by-sa/4.0/>)

1. INTRODUCTION

In 1994 there was a 13.9-meter-high Tsunami that hit the Banyuwangi area, with victims reaching 222 thousand of people [1]. This natural disaster occurred due to an earthquake in the Mw 7.8 megathrust zone. Based on historical records, the tsunami was not a new disaster in the lives of people in Indonesia. From 1600 to 2007, Indonesia experienced several natural disasters, including tsunamis [2]. Indonesia experienced approximately 172 tsunami natural disasters [3]. Physically, a tsunami is described as a chain of ocean waves due to entrapping huge quantities of seawater. Abnormal amounts of sea wind movement could turn into devastating tsunamis, devastating coastal areas far inland [4, 5]. Combating the vulnerability level of an area is important for areas adjacent to the coastline and areas that have the potential to generate a tsunami. To determine tsunami risk, digital elevation model data (DEM) can be developed using tsunami aspects such as vegetation, geomorphology, land cover, land use, and topography. DEM is digital data that describes the geometry of the earth's surface [6]. DEM has important data in tsunami modeling because it can provide a geomorphological, hydrological, and ecological overview of tsunami-prone areas. The DEM model used in this study is a space-based reflected and thermal radiation meter (ASTER). In some countries, ASTER was used to model different aspects and effects of tsunami, such as topography and topography, structural damage, soil elevation, water level rise and layer changes. land cover (LULC) [7].

The magnitude of local tsunami vulnerability is highly correlated with annual land use and land cover change (LULC) dynamics. As a result, LULC can be used as an indicator to determine safe areas for local populations. In addition, LULC can be used to provide information on how severe the tsunami impact was on the area. Therefore, analysis of the dynamics of LULC change is of great importance for policymaking, planning, and implementation of coastal tsunami risk assessment. Currently, several algorithms are used to model LULC change [8]. VI is a new LULC analysis method based on infrared and visible spectral irradiance measurement with leaf area indices, ground cover, chlorophyll content, and plant biomass [9]. LULC temporal dynamics, identified using the Vegetation Index (VI) indicator, are the normalized vegetation difference index (NDVI), the modified terrestrial adaptation vegetation index (MSAVI), and the normalized water difference index (NDWI), Modified Normalized Hydration Index (MNDWI) and Normalized Difference Composition Index (NDBI) [10].

To get highly accurate results, it is necessary to use machine learning. An application of artificial intelligence is machine learning. The main capability of machine learning is handling high-dimensional data, such as remote data sensing, and its complexity into several classes with complex characteristics [11, 12] by using methods in machine learning such as k-Nearest Neighbor (KNN), Random Forest, and SVM. One of the advantages of the random forest method is its resistance to overtraining and creating a large number of random forest branches, which does not cause the risk of overfitting [13]. KNN has several advantages, including very fast training, simple and easy to learn, anti-noise training data, and efficiency when training big data [14, 15]. The concept of a Support Vector Machine (SVM) sends high-dimensional data to low-dimensional vectors. The Support Vector Machine (SVM) method classifies the extracted features as an image classification method [16].

The vegetation index has many indicators that can be used in researching tsunamis. The study [17] examined vegetation density due to the tsunami disaster using the Normalized Difference Vegetation Index (NDVI) indicator. In this study, NDVI was able to identify the density of vegetation in the area. Therefore, vegetation density can be used as an indicator to determine the severity of the area affected by the tsunami. In research, [18] used NDWI and MNDWI indicators to obtain areas that have high-standing water, such as rivers and lakes. The tsunami risk level increases in areas with rivers or lakes because the stagnant water can overflow if a tsunami occurs. In research [19] conducted research to measure soil density, the indicator used in this study was the Modified Soil Adjusted Vegetation Index (MSAVI). MSAVI is proven to be used to study vegetation density which can be used as an indicator of tsunami vulnerability. Another indicator that can be used as an indicator of tsunami vulnerability is the Normalized Difference Built-up Index (NDBI). In this study [20], this indicator was used to measure the density of built-up land. Several previous studies have applied various approaches to improve NDVI, NDWI, MNDWI, NDBI, and MSAVI. However, there are weaknesses in previous studies; namely, the accuracy of the estimation method is still in the range of 70% to 85%, so there is fitness to increase the accuracy. So, this study proposes the Ordinary Kriging method in the application of KNN, Random Forest, and SVM. Ordinary Kriging is a good way to avoid the drawbacks of KNN, Random Forest, and SVM techniques in generating vegetation index data. Ordinary Kriging is an interpolation method that works using the Spatial Autocorrelation principle to obtain data that has no value based on points close to the predicted points.

This study aimed to develop a tsunami vulnerability assessment model using the Vegetation Index extracted from Landsat 8 satellite imagery optimized with KNN, Random Forest, and SVM.

2. RESEARCH METHOD

2.1. Research location

The study was conducted in an area with high tsunami vulnerability, namely Banyuwangi District, East Java, Indonesia. The area is shown in Figure 1. The determination of high tsunami-prone areas is based on the Banyuwangi Regency Regional Regulation Number 8 of 2012 concerning the spatial plan for the Banyuwangi Region 2012-2032 (Indonesian language). Existing documents inform that the total area with high Tsunami risk is 302.86 km². Tsunami-prone areas in this study are defined as areas with low coastal elevations and areas that have the potential or have been affected by a tsunami, including areas in four subdistricts. Ten villages were taken from these sub-districts: Sarongan, Kandangan, Sumberagung, Pesanggaran, Buluagung, Temurejo, Grajagan, Sumberasri, Purwoasri, and Kendalrejo. The study results have various land uses: industrial areas, vegetation, rice fields, fisheries, beaches, and settlements.

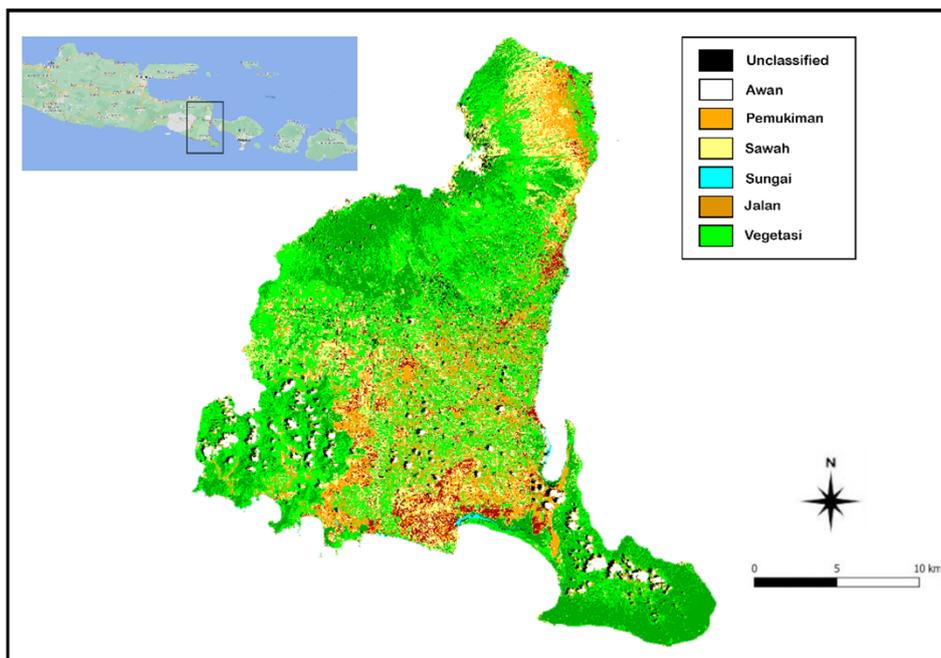


Figure 1. The research area is Banyuwangi Regency, East Java Province, Indonesia

1. Research data

Research data can be categorized into two types of DEM data: i) Aster uses ASTER data to determine contours, slopes, and aspects. ii) VI data is extracted from the runway eight imagery. VI data comprises five types: MSAVI, NDVI, NDBI, NDWI, and MNDWI. VI data is used to determine the characteristics and dynamics of LULC patterns.

2.2. Machine Learning

Data classification methods using ML algorithms are of interest to much remote analytics. Sensing researchers because it can model unusual high-dimensional nonparametric data large scale with high accuracy compared to traditional classification methods number of predictors. Random Forest (RF) is a method to increase the accuracy value. This method aims to build a decision tree consisting of a root, internal, and leaf nodes by randomly taking attributes and data according to the provisions [21]. The decision tree begins by calculating the entropy value as a determinant of the level of attribute impurity and the value of information gain. To calculate the entropy value, use the formula in equation (1).

$$Entropy(Y) = \sum_i p(c|Y) \log_2 p(c|Y) \quad (1)$$

Y is the case set, and $P(c|Y)$ = The proportion of the value of Y to class c .

K-Nearest Neighbor is an algorithm that makes predictions by classifying them to designate something based on similarities with previous data [22]. The k-nn algorithm is intrinsically nonparametric and generally applied for pattern recognition. The k-nn algorithm works by calculating the distance between the unknown sample and the nearest known sample. Then, the unknown sample is assigned a label or class calculated from the average distance of the k-nearest neighbor variables. The formula commonly used to determine distances using Euclidean is as in the following equation (2).

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (2)$$

The use of the k-NN algorithm as a prediction system algorithm is due to the characteristics of the data set. Where five attributes are numeric, and one target data is nominal. Feature data is based on data mining role theory [23], which can be used as a classification that can accommodate both numeric and nominal attributes, but targets must be in nominal form.

The Support Vector Machine (SVM) algorithm is a discrimination-based algorithm that aims to find optimal separation boundaries called hyperplanes to distinguish classes from one another. The sample closest to these hyperplanes is called the support vector, and the difference is expressed as the sum of the weights of the sample subsets, which limits the complexity of the problem [24]. To separate the two classifications through the optimal hyperplane, use the following equation (3).

$$f(x) = \sum_{i=1}^n a_i y_i K(x, x_i) + b \quad (3)$$

Ordinary Kriging (OK) is an interpolation method that works using the Spatial Autocorrelation principle, which assumes that points that are closer to the prediction points have a greater value than sample points that are farther from the prediction points using equation (4) [25].

$$\gamma(d) = \frac{1}{2N(d)} \sum_{i=1}^{N(d)} [Z(x_i + d)]^2 \quad (4)$$

Where $\gamma(d)$ is the value of the variogram at a distance d , $Z(x_i)$ is the value of the variable observed at the location (i), notation $N(d)$ is the sum of all observation points within the range of distance d . The variables' distribution predictions were calculated using linear regression $Z^*(x)$, as shown in equation (5). Where $\lambda_i(x)$ is the weight, $m(x)$ and $m(i)$ are the mathematical expectations of the random variables $Z^*(x)$ and $Z^*(i)$.

$$Z^*(x) - m(x) = \sum_{i=1}^{N(x)} \gamma_i(x) [Z(x_i) - m(x_i)] \quad (5)$$

2.3. Validation test

Mean squared error (MSE) measures the amount of error in statistical models. It assesses the average squared difference between the observed and predicted values using equation (6). When a model has no error, the MSE equals zero. As model error increases, its value increases. The mean squared error is also known as the mean squared deviation (MSD) [26].

$$MSE = \sum_{i=1}^N \frac{(Actual_i - Predicted_i)^2}{N} \quad (6)$$

RMSE is an acronym for Root Mean Square Error, which is the square root of the value obtained from the Mean Square Error function. Using RMSE, we can easily plot a difference between the estimated and actual values of a parameter of the model. The RMSE has been used as a standard statistical metric to measure model performance in meteorology, air quality, and climate research studies [27]. RMSE is shown in equation (7).

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Actual_i - Predicted_i)^2}{N}} \quad (7)$$

Mean Absolute Error calculates the average difference between the calculated values and actual values. It is also known as scale-dependent accuracy, as it calculates errors in observations taken on the same scale. It is used as an evaluation metric for regression models in machine learning. It calculates errors between actual values and values predicted by the model. It is used to predict the accuracy of the machine learning model using equation (8) [28].

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x| \quad (8)$$

2.4. Stages of the experimental procedure

Figure 2 shows the experimental stage. 1) Initial processing of DEM data. This phase is performed by downloading the ASTER image from the portal <https://earthexplorer.usgs.gov/> and selecting the survey area by coordinates. 2) Preprocessing of satellite image data. In this step, the Landsat 8 imagery is transformed from a portal <https://earthexplorer.usgs.gov/>, atmospheric correction, radiometric correction, geometric correction coordinates and selection of study areas according to performance, and 3) Analyze and interpretation of DEM data. This stage is done by determining the hillshade and elevation. 4) Time series data mining VI. This step is performed using data projections MSAVI, NDVI, NDBI, NDWI, and MNDWI 5) VI. This stage is performed using the KNN algorithm, random forest, and SVM 6) to check the accuracy of prediction results. This stage is performed using the RMSE, MAE, and MSE methods. 7) Forecast VI results mapped at high tsunami risk areas by Kriging Method.

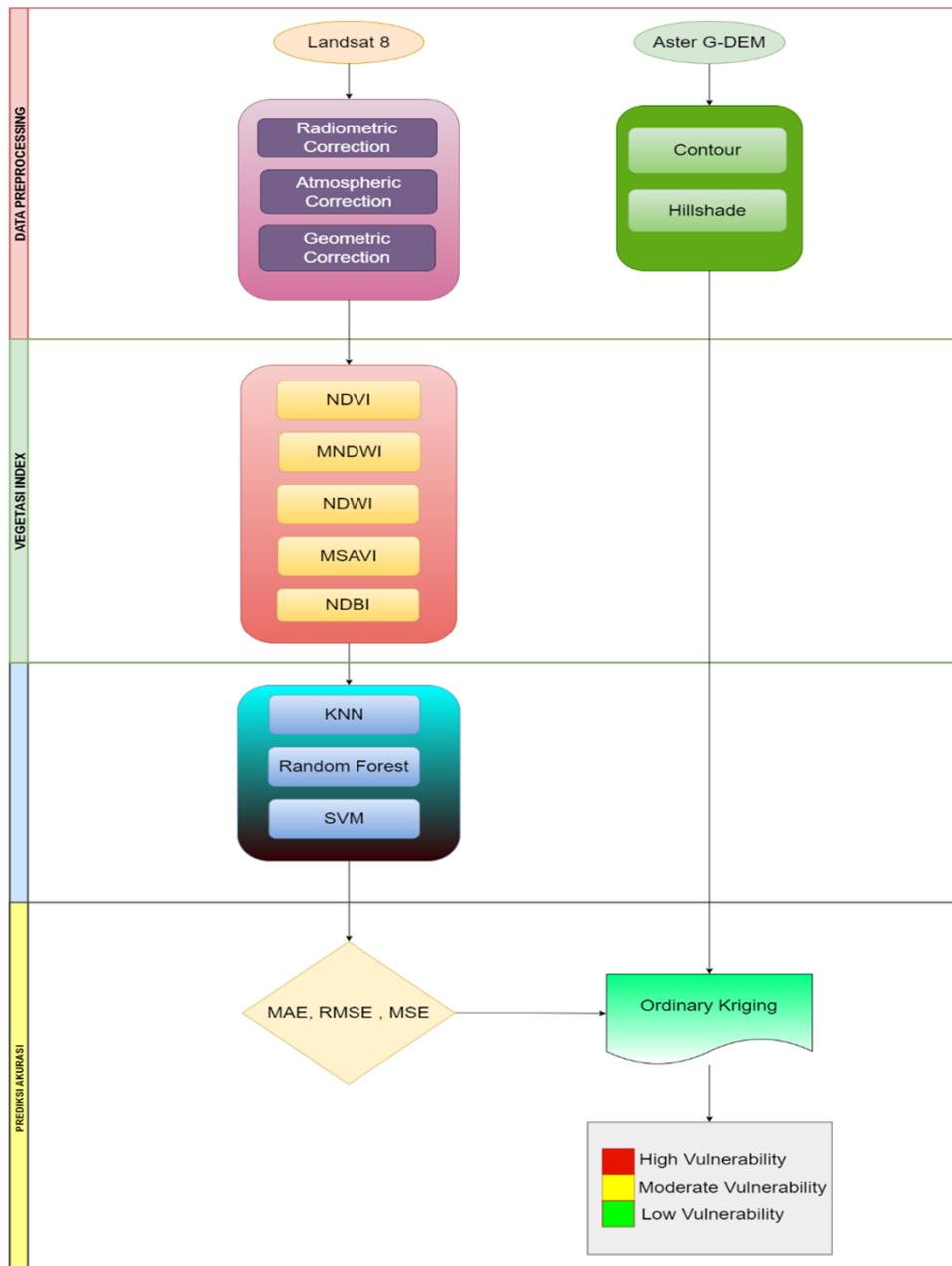


Figure 2. Computer model framework for tsunami vulnerability

3. RESULT AND ANALYSIS

To assess tsunami vulnerability using machine learning, we process eight satellite imagery data on a model built using the vegetation index formula to produce NDVI, NDWI, MNDWI, NDBI, and MSAVI. Next, the results of this vegetation index will be used for the machine learning process to produce a more accurate vegetation index. The resulting data from machine learning will then be entered into the kriging interpolation process. Finally, the results of the vegetation index from the kriging process will be used to determine the tsunami vulnerability level (see Figure 3).

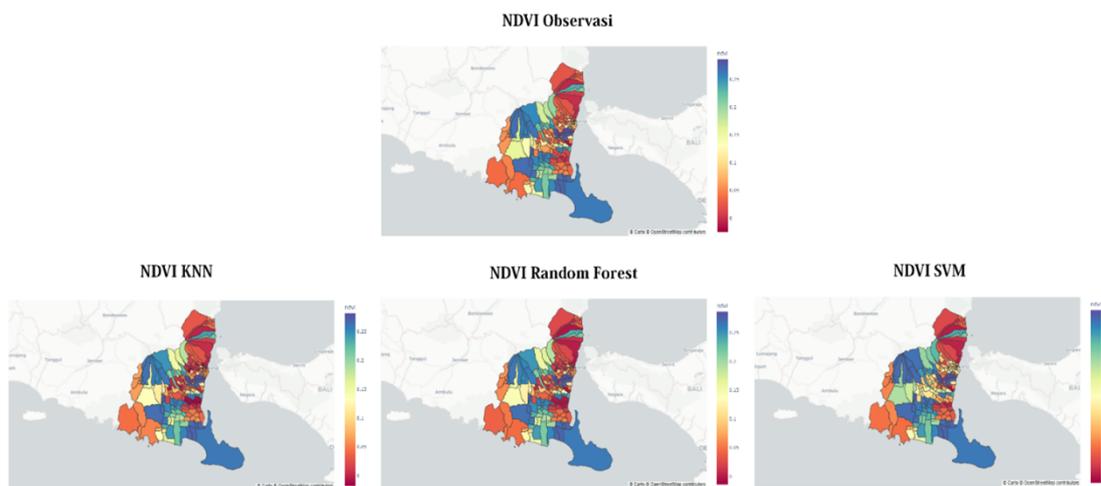


Figure 3. NDVI results of KNN, random forest, and SVM kriging interpolation

Figure 3 shows a spatial prediction model of potential tsunami impacts in each study area using machine learning combined with Kriging interpolation. Areas with a high NDVI represent a high vegetation density and are theoretically less likely to be damaged by a tsunami. The results showed that the NDVI values ranged from 0 to 0.3, meaning that the studies show vegetation density, photosynthetic biomass growth, diverse canopy formations, and high and low vegetation health. Green indicates a high vegetation level, yellow indicates that the vegetation level in the area is within normal limits, and red indicates that part of the land is used for construction. Hence, the area has almost no vegetation level, indicating that no Coastal topography can be recognized and forecasted using the VIs, namely MSAVI, NDVI, NDWI, NDBI, and MNDWI. The usefulness of VIs is based on the following considerations: i) VI focuses on specific types of land cover (vegetation types, water bodies, and soil surfaces). ii) It can reduce background clutter effects. iii) It can reduce the effects of atmospheric distortion caused by the sun's angle and the sensor's viewing angle [29]. The diagram above is categorized as shown in Table 1.

Table 1. NDVI Classification of Banyuwangi Regency [10]

Interpretation	Value Range	Vulnerability level	Color
LULC green vegetation density is low	0 0.1	High	Red
LULC green vegetation density is medium	0.1 0.2	Medium	Yellow
LULC green vegetation density is high	0.2 0.3	Low	Green

In Table 1, the NDVI values LULC are classified into three categories based on their level of vulnerability. Green vegetation. The NDVI number produced indicates the amount of tsunami vulnerability [30]. The high LULC density of green vegetation, which is denoted in green, indicates a high level of vegetation in the area. Green areas in coastal topography, dominated by mangroves, pine forests, and built environments, are affected to some extent by tsunami energy's inundation, propagation, and absorption [31].

Based on Figure 3, the authors took a sample of 10 areas on the coast to see the tsunami risk level, where the classification of tsunami risk levels is taken in Table 1. In the kriging test, NDVI data were obtained from 3 machine learning algorithms: KNN, Random Forest, and SVM. The collected data are in the form of NDVI values, measured by the classifications in Table 1, resulting in tsunami vulnerability scores. Table 2 shows the classification results for 10 sample regions.

Table 2. NDVI Results for 10 Regions

No.	Region	NDVI Results			Vulnerability
		KNN	RF	SVM	
1	Sarongan	0.04	0.03	0.02	High
2	Kandangan	0.05	0.05	0.03	High
3	Sumberagung	0.05	0.04	0.02	High
4	Pesanggaran	0.13	0.12	0.11	Medium
5	Buluagung	0.15	0.15	0.14	Medium
6	Tumurejo	0.21	0.21	0.21	Low
7	Grajagan	0.25	0.25	0.25	Low
8	Sumberasri	0.27	0.27	0.27	Low
9	Purwoasri	0.27	0.27	0.27	Low
10	Kendalrejo	0.26	0.26	0.26	Low

In Table 2, the vulnerability level of an area is taken based on three measurement indicators, namely KNN, Random Forest, and SVM. The level of vulnerability is calculated based on the results of the three indicator values for each region, and the average value is taken to determine the level of tsunami vulnerability based on Table 1. Figures 5 and 4 are a spatial prediction model of the potential impact of the tsunami in each study area using machine learning combined with Kriging interpolation. In areas with high MNDWI/NDWI, the color denoted in blue represents the surface or body of water, which in theory, increases the potential for damaging impacts in the event of a tsunami. Figures 4 and 5 show the colors blue, green, and yellow with an interpolation value of 0.1 - (-0.25), which represents the water surface and paddy fields along the coast, built-up land, and vegetation land.

Changes in NDWI and MNDWI values are also impacted by naturally occurring long-term seasonal patterns, as well as short-term seasonal trends that arise as a result of socioeconomic activity in coastal communities, such as the stocking season, variety, and harvest season in the fishing business. NDWI and MNDWI are the indicators to determine whether there are large stagnant bodies of water and stagnant water, such as rivers and rice fields, during the rainy season. In Figures 4 and 5, green indicates vegetation land with low tsunami vulnerability, yellow indicates built-up land and soil surface with moderate vulnerability, and red indicates high open water with high tsunami vulnerability. All coastal areas indicate land use for economic and social activities, and settlements. From the picture above, a classification is produced as in Table 3.

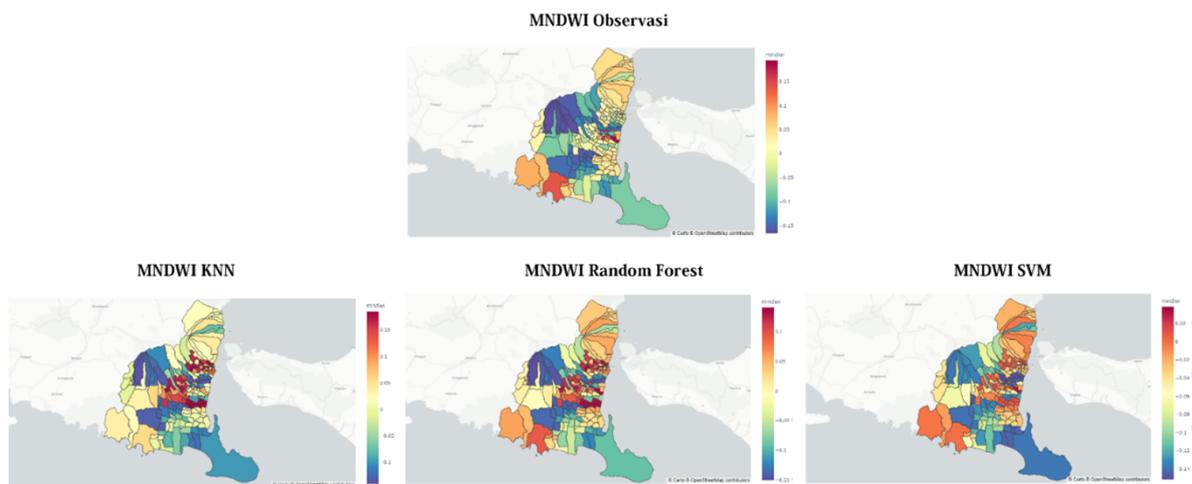


Figure 4. MNDWI interpolation of KNN, random forest, and SVM kriging interpolations

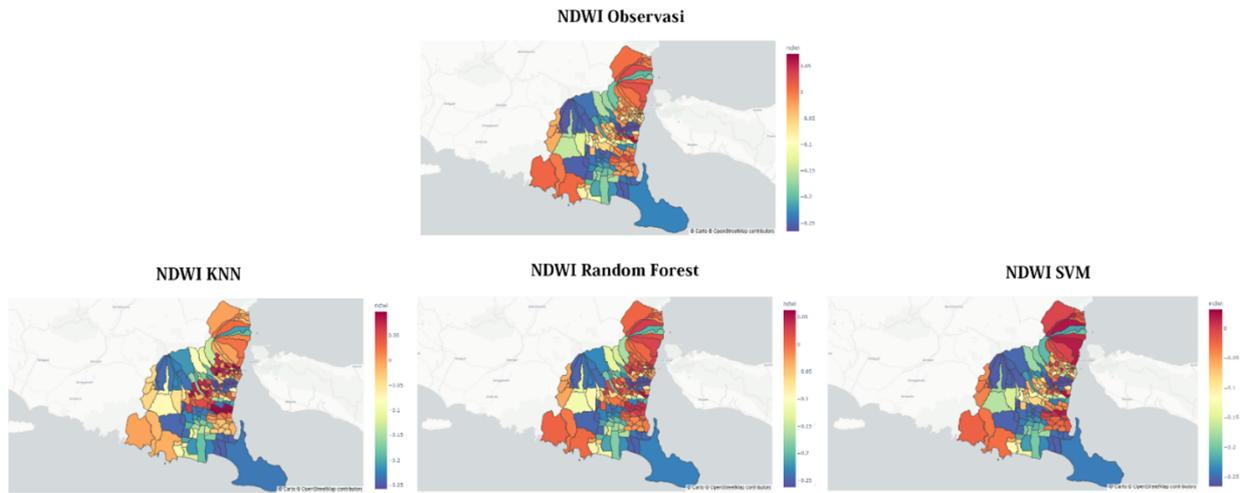


Figure 5. NDWI Interpolation of KNN, Random Forest, and SVM Kriging Interpolation

Table 3. Classification of NDWI/MNDWI Banyuwangi Regency [9]

Interpretation	Value Range	Vulnerability level	Color
Open Water	> 0.02	High	Red
Built-up Land and Ground Surface	0.02 (-0.05)	Medium	Yellow
Vegetation Land	> -0.05	Low	Green

In Table 3, the NDWI/MNDWI values are classified into three categories based on the level of open water, built-up land, and vegetation land. The level of tsunami vulnerability can be seen from the NDWI, and MNDWI values obtained. NDWI is an indicator to determine whether there are large bodies of water and stagnant water, such as rivers and rice fields [32]. The river itself is one of the factors that increase the vulnerability of the study area to tsunamis. When a tsunami occurs, tidal waves can enter a land through river channels so that areas close to rivers have a high level of vulnerability compared to areas without rivers [19].

Based on the results of Figures 4 and 5, the authors took a sample of 10 areas on the coast to see the tsunami risk level, where the tsunami risk level classification is taken in Table 3. In the kriging test, NDWI and MNDWI data were obtained from 3 machine learning algorithms: KNN, Random Forest, and SVM. The extracted data is in the form of NDWI and MNDWI values, which are then assessed using the categories in Tables 4 (MNDWI) and 5 (NDWI) to yield tsunami vulnerability ratings. Table 4 and Table 5 show the results of the ten classifications in the sample area.

Table 4. MNDWI results for 10 regions

No.	Region	MNDWI Results			Vulnerability
		KNN	RF	SVM	
1	Sarongan	0.03	0.05	0.04	High
2	Kandangan	0.02	0.04	0.02	High
3	Sumberagung	0.05	0.09	0.02	High
4	Pesanggaran	0.01	0.02	0.01	Medium
5	Buluagung	-0.01	0.005	0.006	Medium
6	Tumurejo	-0.06	-0.04	-0.1	Low
7	Grajagan	-0.09	-0.08	-0.13	Low
8	Sumberasri	-0.11	-0.11	-0.14	Low
9	Purwoasri	-0.12	-0.12	-0.14	Low
10	Kendalrejo	-0.09	-0.09	-0.13	Low

Table 5. NDWI Results for 10 Regions

No.	Region	NDWI Results			Vulnerability
		KNN	RF	SVM	
1	Sarongan	0.01	0.05	0.02	High
2	Kandangan	0.01	0.07	0.03	High
3	Sumberagung	0.08	0.01	0.02	High
4	Pesanggaran	-0.09	-0.08	-0.11	Low
5	Buluagung	-0.12	-0.11	-0.13	Low
6	Tumurejo	-0.18	-0.18	-0.19	Low
7	Grajagan	-0.22	-0.22	-0.23	Low
8	Sumberasri	-0.24	-0.25	-0.25	Low
9	Purwoasri	-0.24	-0.25	-0.25	Low
10	Kendalrejo	-0.23	-0.23	-0.24	Low

In Tables 4 and 5, the vulnerability level of an area is taken based on three measurement indicators, namely KNN, Random Forest, and SVM. The level of vulnerability is calculated based on the results of 3 indicator values for each region, and the average value is taken to determine the level of tsunami vulnerability based on table 3. The average value of NDWI or MNDWI, which is in the range of more than 0.02, has a high level of vulnerability. Values 0.02 to -0.05 have moderate susceptibility, and values greater than (0.05) have low susceptibility.

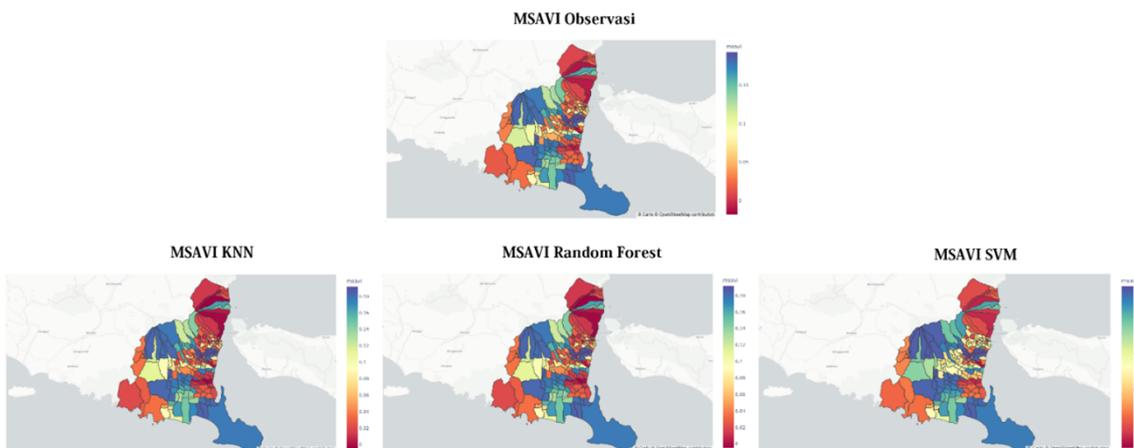


Figure 6. MSAVI results of KNN, Random Forest, and SVM Kriging Interpolation

Figure 6 is a spatial prediction model of the potential impact of the tsunami in each study area using machine learning combined with Kriging interpolation. In areas with high MSAVI, it represents high vegetation density so that, in theory, it reduces the potential damage impact in the event of a tsunami. The results showed that the MSAVI value was in the range of 0 -0.18, which means that the research has LULC vegetation density and low to the low land slope. The green color represents the density of vegetation and the high slope of the land, which has a low level of tsunami vulnerability. The yellow color indicates the level of density of vegetation and the medium slope of the land, which has a moderate level of vulnerability to a tsunami. The red color indicates the density of vegetation, indicating that the area has a low slope due to the development of some of the land and high vulnerability to the tsunami. In Table 6, the MSAVI values are classified into three categories based on the level of vegetation density and land slope. The level of tsunami vulnerability can be seen from the MSAVI value obtained. MSAVI is an indicator to determine the level of vegetation density in an area and the area's slope level. Vegetation density significantly affects tsunami hazards because high vegetation density, such as in forests, can dampen the energy of tsunami waves [18]. The slope of an area affects the height or run-up of the tsunami. The steeper the place, the lower the sea waves. Tsunamis will not travel too far inland on coastal slopes as it is received and reflected by the coastal cliffs. Whereas on a sloping coast, a tsunami can reach several hundred meters inland [33].

Table 6. Classification MSAVI Banyuwangi Regency [10]

Interpretation	Value Range	Vulnerability level	Color
LULC Vegetation Density and Low Land Slope	0 0.05	High	Red
LULC Vegetation Density and Medium Slope	0.05 0.12	Medium	Yellow
LULC Vegetation Density and High Slope	0.12 0.18	Low	Green

In table 7, the vulnerability level of an area is taken based on three measurement indicators, namely KNN, Random Forest, and SVM. The level of vulnerability is calculated based on the results of the three indicator values for each region, and the average value is taken to determine the level of tsunami vulnerability based on table 6s. MSAVI values that are in the range of 0 to 0.05 have a high level of vulnerability. Values 0.05 to 0.12 have a moderate vulnerability, and values 0.12 to 0.18 have low vulnerability.

Table 7. MSAVI Results from 10 Regions

No	Region	MSAVI Results			Color
		KNN	RF	SVM	
1	Sarongan	0.02	0.01	0.01	High
2	Kandangan	0.02	0.02	0.01	High
3	Sumberagung	0.03	0.02	0.01	High
4	Pesanggaran	0.09	0.08	0.07	Medium
5	Buluagung	0.1	0.1	0.09	Medium
6	Tumurejo	0.14	0.14	0.14	Low
7	Grajagan	0.17	0.17	0.16	Low
8	Sumberasri	0.18	0.18	0.18	Low
9	Purwoasri	0.18	0.18	0.18	Low
10	Kendalrejo	0.17	0.17	0.17	Low

Figure 7 is a spatial prediction model of the potential impact of the tsunami in each study area using machine learning combined with Kriging interpolation. Areas with a high NDBI represent a high density of built-up land, so in theory, it increases the potential for damage if a tsunami occurs. Changes in short-term NDBI values indicate massive changes in land use carried out in coastal areas as physical development activities in coastal areas, as shown in Figure 7. Physical development is a change in land use from open land to buildings, mesh networks, and the fishing industry, such as ponds shrimp. In Figure 7, the green color indicates the low density of built-up land with a low level of tsunami vulnerability, the yellow color indicates the medium density of built-up land with a moderate level of tsunami vulnerability, and the red color indicates the high density of built-up land with a high level of vulnerability. NDBI is the VI used to map built-up land.

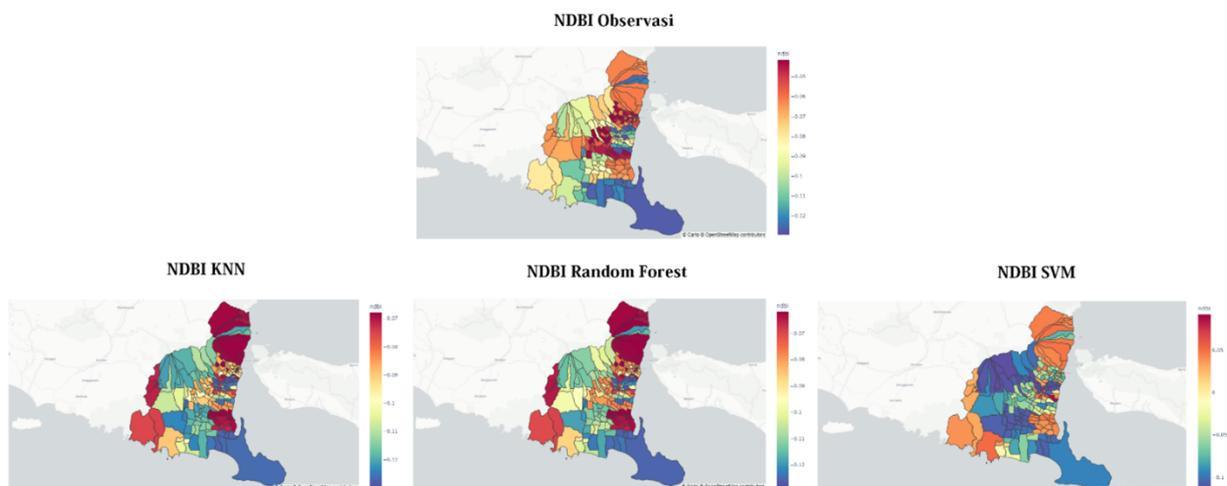


Figure 7. NDBI interpolation KNN, random forest, and SVM kriging interpolation results

In Table 8, the NDBI values are classified into three categories based on the level of density of built-up land. The level of tsunami vulnerability can be seen from the NDBI value obtained. NDBI is an indicator to determine the level of density of built-up land in an area. The NDBI is referred to as an urban index or an index that represents an area that functions as a place of settlement, concentration, and distribution of government services, social services, and economic activity [17]. The NDBI index shows that the larger the area with built-up land, the higher the level of vulnerability to tsunami risk [34]. Vegetation density significantly affects tsunami hazards because high vegetation density, such as in forests, can dampen the energy of tsunami waves [24]. The slope of an area affects the height or run-up of the tsunami. The steeper the place, the lower the sea waves. The tsunami will not go too far inland on the steep slopes of the coast because it is caught and reflected by the coastal cliffs. Meanwhile, on a sloping coast, a tsunami can reach several hundred meters on land [35].

Table 8. NDBI Results for 10 Regions [9]

Interpretation	Range of Values	Vulnerability level	Color
Built-up Land Density High	-0.05 (-0.08)	High	Red
Built-up Land Density Medium	(-0.08) (-0.1)	Medium	Yellow
Built-up Land Density Low	(-0.1) (-0.12)	Low	Green

In Table 9, the vulnerability level of an area is taken based on three measurement indicators, namely KNN, Random Forest, and SVM. The level of vulnerability is calculated based on the results of the three indicator values for each region, and the average value is taken to determine the level of tsunami vulnerability based on Table 9. For example, NDBI values in the range of -0.05 to -0.08 have a high level of tsunami vulnerability and high density. Conversely, values -0.08 to -0.1 have a moderate vulnerability, and values -0.1 to -0.12 have low vulnerability. Based on the results of research on the level of vulnerability using machine learning algorithms on the NDVI, NDWI, MNDWI, MSAVI, and NDBI, their indexes are combined in the table below to find overall results using the indicators in the previous table.

Table 9. NDBI Results of 10 Regions

No.	Region	NDBI Results			Vulnerability
		KNN	RF	SVM	
1	Sarongan	-0.07	-0.07	-0.04	High
2	Kandangan	-0.07	-0.06	-0.05	High
3	Sumberagung	-0.08	-0.08	-0.08	High
4	Pesanggaran	-0.1	-0.1	-0.08	Medium
5	Buluagung	-0.1	-0.1	-0.09	Medium
6	Tumurejo	-0.1	-0.1	-0.09	Low
7	Grajagan	-0.12	-0.12	-0.09	Low
8	Sumberasri	-0.12	-0.12	-0.09	Low
9	Purwoasri	-0.12	-0.12	-0.09	Low
10	Kendalrejo	-0.12	-0.12	-0.09	Low

In Table 10, the level of vulnerability is divided into three parts, namely low (L), medium (M), and high (H).

Table 10. Vulnerability Results Overall Vegetation Index

No.	Region	Vulnerability Result of Vegetation index					Result
		NDVI	NDWI	MNDWI	MSAVI	NDBI	
1	Sarongan	H	H	H	H	H	High
2	Kandangan	H	H	H	H	H	High
3	Sumberagung	H	H	H	H	H	High
4	Pesanggaran	M	L	M	M	M	Medium
5	Buluagung	M	L	M	M	M	Medium
6	Tumurejo	L	L	L	L	M	Low
7	Grajagan	L	L	L	L	L	Low
8	Sumberasri	L	L	L	L	L	Low
9	Purwoasri	L	L	L	L	L	Low
10	Kendalrejo	L	L	L	L	L	Low

Machine learning results are the initial process of obtaining VI data that will be used in the kriging process. The use of Kriging

aims to find out the pattern of the spatial distribution of each VI. Kriging will determine an unknown VI value because there are no sample points, so areas where there are no sample points, will get a VI value. This VI is important because it will be used to obtain regional vulnerability data. The resulting machine learning data are NDVI, MNDWI, NDWI, NDBI, and MSAVI. The results are separated based on the machine learning algorithm used. MSE, MAE, and RMSE test results that are close to zero have high precision, and a result not close to zero has low precision.

The coastal topography of the study area can be visualized using the hillshade method, as shown in Figure 8. Hillshade visualizes surface relief using a DEM raster data source in 2D format, adding light to make it look like a 3D object. The purpose of using the hillshade method is to sharpen the visualization of the relief of the ground. A visualization of shadows and hill heights is shown in Figure 8. This represents a model of the relief and elevation characteristics of the earth's surface in the 3D format as a background where the NDVI vegetation, MNDWI water bodies, MSAVI open land, and MNDWI built-up land are located. Figure 8 (1, 2, 3, and 6) shows the sub-districts of Sarongan, Kandangan, Sumberagung, and Tumurejo, with terrain on the coast dominated by hills that could theoretically protect the island from high tsunami waves. The elevations at 1, 2, 3, and 6 indicate that these points are at an altitude of about > 500 meters above sea level. This area is more protected from tsunamis because it is hilly and heavily forested. Regions 4 and 5, namely Pesanggaran and Buluagung, show open coastal terrain without natural protection, such as hills. Moreover, there are only buildings and city parks to withstand the tsunami. In areas 7, 8, and 9, namely Grajagan, Sumpersari, and Purwoasri, there are rivers where if a tsunami occurs, the water will overflow and hit the settlements of local residents, and there are no hills in those areas. Furthermore, Region 10, Kendallejo, is a wooded area with more trees and taller bases than the surrounding areas. There are also not many residential areas in this area, so if a tsunami occurs, it will not significantly impact the surrounding population.

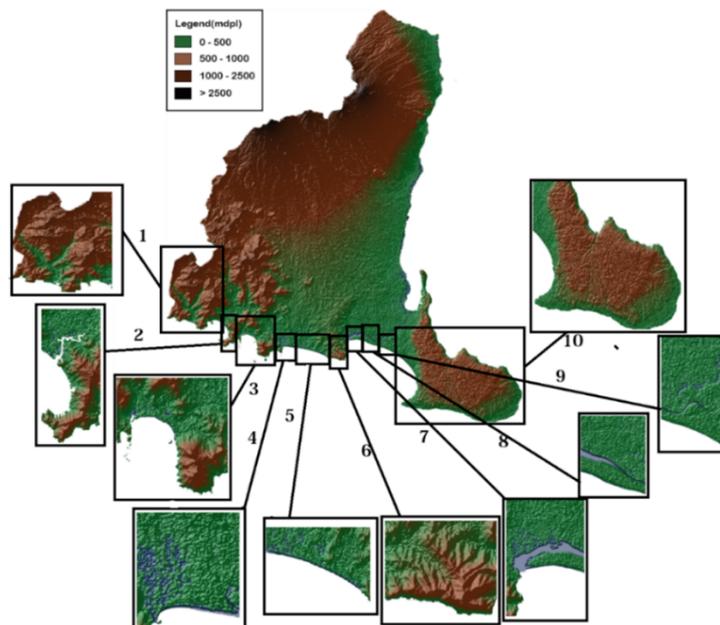


Figure 8. Model of the Relief and Elevation Characteristics of the Earth's Surface in Banyuwangi District

The tsunami vulnerability assessment model for land use and land cover using NDVI information, NDWI, MDWI, MSAVI, and NDBI vegetation index extracted from Landsat 8 satellite imagery shows effective performance. Based on the test results of the machine learning algorithm used in this study, namely KNN, Random Forest, SVM, and spatial interpolation tests, namely Ordinary Kriging using MSE, MAE, and RMSE, show high accuracy compared to previous studies with an accuracy value of MSE of 0.063, MAE of 0.04, and RMSE of 0.0742. This study identified an area for tsunami vulnerability using five indicators of vegetation index, NDVI, NDWI, MNDWI, MSAVI, and NDBI. After using five indicators found vulnerability. From the results of research on the ten areas studied, there were three areas (Sarongan, Kandangan, and Sumberagung) with conditions that have a high level of vulnerability to tsunamis because they have low vegetation levels, with water levels, open spaces, and tall buildings. 2 regions (Pesanggaran and Buluagung) with conditions that have a moderate level of vulnerability to tsunamis because they have low ground elevation and

no hills, relatively low vegetation levels, and only a few water levels and buildings, and finally five regions (Tumurejo, Grajagan, Sumberasri, Purwoasri, and Kendalrejo) with conditions that have a low level of vulnerability to tsunamis because they have tall vegetation, low air surface area, and buildings. The results of this study have a higher level of accuracy compared to studies without using machine learning. The comparison of the proposed method is better than previous studies, which can be shown in Table 11.

Table 11. Comparison of The Proposed Model Performance with Previous Studies

No.	Author (Year)	Dataset	Method	Accuracy (MAE)
1	[36]	NDVI	ANN	0.055
2	[37]	NDVI	Linear Regression	0.06
3	[38]	NDVI	Support Vector Regression (SVR)	0.25
4	The Proposed Method	NDVI	KNN, SVM, Random Forest+Ordinary Krigging	0.04

4. CONCLUSION

The tsunami-vulnerability-assessment-model results using MSAVI, NDBI, NDWI, and MNDWI extracted from Landsat 8 satellite imagery and optimized with machine learning show effective and accurate results. In general, the tsunami risk assessment using the vegetation index is mostly carried out without the use of machine learning and uses only a few vegetation index indicators. This study used three machine learning algorithms to optimize the vegetation index to produce more accurate data. Compared to only processing the vegetation index without using machine learning and kriging interpolation. From the results of research on the ten areas studied, there are three areas (Sarongan, Kandangan, and Sumberagung) with conditions that have a high level of vulnerability to tsunamis because they have low vegetation levels, with water levels, open spaces, and tall buildings. 2 regions (Pesanggaran and Buluagung) with conditions that have a moderate level of vulnerability to tsunamis because they have low ground elevation and no hills, relatively low vegetation levels, and only a few water levels and buildings, and finally five regions (Tumurejo, Grajagan, Sumberasri, Purwoasri, and Kendalrejo). Further research can be carried out in other areas with a high tsunami risk and add vegetation index indicators or new indicators so that results with a higher level of accuracy are obtained and can add to the algorithm used.

REFERENCES

- [1] L. Y. Irawan, Sumarmi, S. Bachri, A. W. Sholeha, S. A. Arysandi, S. D. Dirgantara, M. M. R. Devy, E. B. Prasetyo, A. D. Febrianto, and M. G. Rosyendra, "Assessing Community Coping Capacity in Face of Tsunami Disaster Risk (Case Study: Sumberagung Coastal Area, Banyuwangi, East Java)," in *IOP Conference Series: Earth and Environmental Science*, mar 2021, pp. 1–12.
- [2] S. P. D. Sriyanto, P. A. Angmalisang, and L. Manu, "Optimal Tide Gauge Location for Tsunami Validation in the Lembeh Island, North Sulawesi," *Indonesian Journal on Geoscience*, vol. 9, no. 3, pp. 315–327, 2022.
- [3] A. D. Permatasari and S. Y. Joko Prasetyo, "Identifikasi Wilayah Resiko Kerusakan Lahan Terbangun Sebagai Dampak Tsunami Berdasarkan Analisis Building Indices," *Jurnal Transformatika*, vol. 20, no. 1, p. 13, jul 2022.
- [4] A. Y. Isnaeni and S. Y. J. Prasetyo, "Klasifikasi Wilayah Potensi Risiko Kerusakan Lahan Akibat Bencana Tsunami Menggunakan Machine Learning," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 1, pp. 33–42, 2022.
- [5] J. Keller, *Tsunami Modeling and Hazard Assessment*. New York: States Academic Pr, 2022.
- [6] 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," *Geoscience Letters*, vol. 8, no. 1, pp. 1–16, dec 2021.
- [7] 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," *Quaestiones Geographicae*, vol. 39, no. 3, pp. 87–109, 2020.
- [8] M. Dangulla, L. A. Munaf, and F. R. Mohammad, "Spatio-Temporal Analysis of Land Use/Land Cover Dynamics in Sokoto Metropolis Using Multi-Temporal Satellite Data and Land Change Modeller," *Indonesian Journal of Geography*, vol. 52, no. 3, pp. 306–316, 2021.

- [9] T. H. Rehman, M. E. Lundy, and B. A. Linquist, "Comparative Sensitivity of Vegetation Indices Measured via Proximal and Aerial Sensors for Assessing N Status and Predicting Grain Yield in Rice Cropping Systems," *Remote Sensing*, vol. 14, no. 12, pp. 1–18, 2022.
- [10] S. Y. J. Prasetyo, B. H. Simanjuntak, K. D. Hartomo, and W. Sulisty, "Computer Model for Tsunami Vulnerability Using Sentinel 2A and SRTM Images Optimized by Machine Learning," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2821–2835, 2021.
- [11] T. R. Fariz, F. Daeni, and H. Sultan, "Pemetaan Perubahan Penutup Lahan Di Sub-DAS Kreo Menggunakan Machine Learning Pada Google Earth Engine," *Jurnal Sumberdaya Alam dan Lingkungan*, vol. 8, no. 2, pp. 85–92, 2021.
- [12] T. Wahyono, *Fundamental of Python for Machine Learning*, cetakan 1 ed. Yogyakarta: Penerbit Gava Media, 2018.
- [13] M. Aldi, I. R. Siregar, and A. Bilqis, "Pemetaan Daerah Rawan Longsor Menggunakan Machine Learning di Kecamatan Muara Tami, Kota Jayapura, Papua," *Jurnal Geofisika*, vol. 19, no. 1, pp. 24–30, 2021.
- [14] D. Indrajaya, A. Setiawan, and B. Susanto, "Comparison of K-Nearest Neighbor and Naive Bayes Methods for SNP Data Classification," *Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, vol. 22, no. 1, pp. 149–164, 2022.
- [15] D. Y. Heryadi and T. Wahyono, *Machine Learning Konsep dan Implementasi*, cetakan 1 ed. Yogyakarta: Penerbit Gava Media, 2020.
- [16] S. Efendi and P. Sihombing, "Sentiment Analysis of Food Order Tweets to Find Out Demographic Customer Profile Using SVM," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 583–594, jul 2022.
- [17] M. Trishiani, S. Sugianto, T. Arabia, and M. Rusdi, "Vegetation Density Analysis in Banda Aceh City Before and After the Tsunami Using Satellite Imagery Data," in *IOP Conference Series: Earth and Environmental Science*, vol. 951, no. 1, jan 2022, pp. 1–8.
- [18] S. Koshimura, L. Moya, E. Mas, and Y. Bai, "Tsunami Damage Detection with Remote Sensing: A Review," *Geosciences (Switzerland)*, vol. 10, no. 5, pp. 1–28, 2020.
- [19] A.-L. Balogun, S. T. Yekeen, B. Pradhan, and O. F. Althuwaynee, "Spatio-Temporal Analysis of Oil Spill Impact and Recovery Pattern of Coastal Vegetation and Wetland Using Multispectral Satellite Landsat 8-OLI Imagery and Machine Learning Models," *Remote Sensing*, vol. 12, no. 7, pp. 1–25, apr 2020.
- [20] A. Andriani, G. D. Putra, S. Ramadhani, Ismael, and H. G. Putra, "Analysis of Landslide Potential Due to Changes of Land Use/Land Cover at the Kuranji Watershed, Padang Using Normalized Difference Built-Up Index (NDBI)," in *International Conference on Disaster Mitigation and Management (ICDMM 2021)*, L. Comfort, S. Saravanan, I. Sengara, and Fauzan, Eds., dec 2021, pp. 1–6.
- [21] T. N. Nuklianggraita, A. Adiwijaya, and A. Aditsania, "On the Feature Selection of Microarray Data for Cancer Detection based on Random Forest Classifier," *Jurnal Infotel*, vol. 12, no. 3, pp. 89–96, 2020.
- [22] D. Kurniadi, A. Mulyani, and I. Muliana, "Prediction System for Problem Students Using K-Nearest Neighbor and Strength and Difficulties Questionnaire," *Jurnal Online Informatika*, vol. 6, no. 1, pp. 53–62, jun 2021.
- [23] N. Hasdyna, B. Sianipar, and E. M. Zamzami, "Improving The Performance of K-Nearest Neighbor Algorithm by Reducing The Attributes of Dataset Using Gain Ratio," in *Journal of Physics: Conference Series*, jun 2020, pp. 1–6.
- [24] N. R. Wardani, S. Saepudin, and C. Warman, "Sentimen Analisis Kegiatan Trading Pada Aplikasi Twitter dengan Algoritma SVM, KNN dan Random Forrest," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 6, no. 2, pp. 863–870, 2022.
- [25] 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.
- [26] N. Elizabeth Michael, M. Mishra, S. Hasan, and A. Al-Durra, "Short-Term Solar Power Predicting Model Based on Multi-Step CNN Stacked LSTM Technique," *Energies*, vol. 15, no. 6, pp. 1–20, mar 2022.

- [27] T. O. Hodson, “Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not,” *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, 2022.
- [28] S. M. Robeson and C. J. Willmott, “Decomposition of the Mean Absolute Error (MAE) Into Systematic and Unsystematic Components,” *PLOS ONE*, vol. 18, no. 2, pp. 1–8, feb 2023.
- [29] Y. Yamanaka and T. Shimozone, “Tsunami Inundation Characteristics Along the Japan Sea Coastline: Effect of Dunes, Breakwaters, and Rivers,” *Earth, Planets and Space*, vol. 74, no. 1, pp. 1–16, dec 2022.
- [30] V. Zuzulová, J. Vido, and B. Šiška, *Agricultural Drought in Slovakia: An Impact Assessment NDVI and Satellite Based Data*. Switzerland: Springer Cham, 2020.
- [31] G. Cárdenas and P. A. Catalán, “Accelerating Tsunami Modeling for Evacuation Studies through Modification of the Manning Roughness Values,” *GeoHazards*, vol. 3, no. 4, pp. 492–508, 2022.
- [32] S. Sharifzadeh and S. Adhikari, “A Support Vector Machine-Based Water Detection Analysis in a Heterogeneous Landscape Using Landsat TM Imagery,” *The California Geographer*, vol. 59, pp. 1–22, 2020.
- [33] F. A. Safira, C. Muryani, and G. A. Tjahjono, “Tsunami Susceptibility Analysis in Coastal Area Petanahan District, Kebumen Regency,” *Jambura Geoscience Review*, vol. 4, no. 2, pp. 110–122, 2022.
- [34] R. Paulik, H. Craig, and B. Popovich, “A National-Scale Assessment of Population and Built-Environment Exposure in Tsunami Evacuation Zones,” *Geosciences (Switzerland)*, vol. 10, no. 8, pp. 1–15, 2020.
- [35] S. M. RAHAYU and A. S. ANDINI, “Study of Tsunami Mitigation Based on Vegetation in Serenting Beach, Mandalika Special Economic Zone, Lombok Island,” *International Journal of Research -GRANTHAALAYAH*, vol. 8, no. 12, pp. 60–68, 2020.
- [36] Y. Zhao, P. Hou, J. Jiang, J. Zhao, Y. Chen, and J. Zhai, “High-Spatial-Resolution NDVI Reconstruction with GA-ANN,” *Sensors*, vol. 23, no. 4, p. 2040, feb 2023.
- [37] M. Sarafanov, E. Kazakov, N. O. Nikitin, and A. V. Kalyuzhnaya, “A Machine Learning Approach for Remote Sensing Data Gap-Filling With Open-Source Implementation: An Example Regarding Land Surface Temperature, Surface Albedo and NDVI,” *Remote Sensing*, vol. 12, no. 23, pp. 1–21, 2020.
- [38] F. H. Evans and J. Shen, “Long-Term Hindcasts of Wheat Yield in Fields Using Remotely Sensed Phenology, Climate Data and Machine Learning,” *Remote Sensing*, vol. 13, no. 13, pp. 1–20, jun 2021.

[This page intentionally left blank.]