Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 2, April 2024, pp. 1138~1146
ISSN: 2302-9285, DOI: 10.11591/eei.v13i2.5910

A machine learning-based computer model for the assessment of tsunami impact on built-up indices using 2A Sentinel imageries

Sri Yulianto Joko Prasetyo[1], Bistok Hasiholan Simanjuntak[2], Yeremia Alfa Susatyo[1], Wiwin Sulistyo[1]
[1]Department of Informatic Engineering, Faculty of Information Technology, Satya Wacana Christian University, Central Java, Indonesia
[2]Bussines and Agriculture Faculty, Satya Wacana Christian University, Central Java, Indonesia

Article Info

Article history:

Received Feb 1, 2023
Revised Jun 11, 2023
Accepted Sep 14, 2023

Keywords:

Build-up indices
Computer model
Machine learning
Remote sensing
Tsunami

ABSTRACT

This study aims to build a computer model to detect built-up land in the identified tsunami hazard zone based on Sentinel 2A imagery using the normalized built up area index (NBI), urban index (UI), normalize difference build-up index (NDBI), a modified built-up index (MBI), index-based built-up index (IBI) algorithms, optimized with machine learning Random Forest (RF) and extreme gradient boosting (XGboost) algorithms and the spatial patterns are predicted using the ordinary kriging (OK) method. Testing of the accuracy of the classification and optimization results was performed using the Kohen Kappa and overall accuracy functions. The results of the study show that a built-up land consisting of open land and water, settlements, industry areas, and agriculture and tourism areas can be identified using the parameters of built-up indices. The accuracy testings that were performed using overall accuracy and Kohen Kappa methods show that classification and prediction are highly accurate using XGboost machine learning, namely >91%. This study produces a novelty of finding, namely a computer model to detect and predict the spatial distribution of built-up land in 4 scales, i.e., very low, low, high, and very high based on NBI, UI, NDBI, MBI, IBI data extracted from Sentinel 2A imagery.

This is an open access article under the CC BY-SA license.

Corresponding Author:

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

1. INTRODUCTION

Currently, the methodology for conducting tsunami vulnerability detect and assessments is very advanced and developed rapidly starting from modeling methods are linear, non linear, numerical, photogrammetry image analysis and remote sensing [1]–[5]. Remote sensing image analysis methods include medium resolution images such as Landsat 8 OLI and Sentinel 2A, or high-resolution images such as SPOT 5 and Quickbird [6]–[8]. A quick calculation of damage to buildings caused by a tsunami can be done because of the existence of various machine learning functions and built-up indices data extracted from remote sensing imageries [9]. Machine learning methods have long been applied to mitigate tsunamis, including predicting inundation, maximum wave height and arrival time of tsunami waves on land, even though the uncertainty of the prediction results is very high [10]. In Indonesia, tsunami is the threat of disaster in the future due to indicators of ancient tsunami silt deposits on the south coast of Java and the Euro-Asian and

Indo-Australian plates which have the potential to cause large earthquakes and trigger tsunami waves from the Sunda Strait to Bali [11]–[18].

In terms of seismotectonic zones in Indonesia, the coastal areas of Central Java and Yogyakarta are included in the zone B with an intensity of tsunami events of more than 2.5 times in a period of 30-50 years. In this zone, tsunamis are generated by two types of earthquakes, namely the subduction of the Indian Ocean Plate under the Eurasian Plate and the pressure of the arc plate which lies east to west in the north of the Islands of Bali, Lombok and Sumbawa [19]. Past data shows that the area of zone B, especially the southern seas of Central Java and Yogyakarta, has been hit by 20 tsunamis with varying strengths throughout human history, which were recorded from before 1600 until the end of 2006 [20]. Currently, there is no computer-based modeling study to find a model that can detect a built-up land quickly and accurately in zone B tsunami disaster risk areas. As a solution to this problem, a study was conducted with the aims: i) building a computer model to detect built-up land from Sentinel 2A satellite imagery data; ii) classifying and optimizing the digital number (DN) data detection process for Sentinel 2A satellite images using machine learning Random Forest (RF) and eXtreme gradient boosting (XGBoost); and iii) predicting the spatial pattern of the distribution of built-up lands using the Ordinary Kriging (OK) method. This study produces a novelty of finding, namely a computer model to detect and predict the spatial distribution of built-up land in 4 scales: very low, low, high, and very high based on normalized built up area index (NBI), urban index (UI), normalize difference build-up index (NDBI), modified built-up index (MBI), index-based built-up index (IBI) data extracted from Sentinel 2A imagery. The cluster K-Mean algorithm (CKA) is an algorithm that works on Euclidian distance data y to the centroid (*c*) [21]–[23]. Euclidian distance is formulated by (1):

$$d\ (y,c) = \sqrt{(\sum\_(i = 1)\^n \blacksquare (y\_i - C\_i\ )\^2\ )} \quad (1)$$

RF algorithm is a combination of non-parametric classification method and the classification and decision tree (CART) $\{\ DT(a, \Theta\_r\ )\}\_(r = 1)\^T$ method, which *a* is the input data observed in vector form, Θ_r is sample data in vector form and is taken randomly from the input data set as a result of $\Theta\_1, \Theta\_2, \Theta\_3 \ldots \Theta\_(r - 1)$ observations. *T* notation is the sample data which is used as training data of bootstrap. If depicted in a forest, sample data are trees that will be grown and selected randomly and classified into certain nodes or classes using the CART method [24]. Number of trees grown is represented as n-tree and the number of classes or classifiers is represented as m-try. To classify a tree and divide it into certain nodes or classes, the Gini index function is used with (2):

$$G\_i = 1 - \sum\_(r = 1)\^n \blacksquare \ [\![S\_n\^2, r]\!] \quad (2)$$

Where $S\_n\^2, r$ is the ratio of training data taken randomly from historical data, and *n* notation is the category of classified nodes [25], [26]. XGBoost is a machine learning algorithm that works based on the concept of a decision tree, where each decision tree node will be connected to one another hierarchically [27]. Each tree will contribute to building a large classifier by forming an ensemble of weak classifiers [28]. The XGBoost as (3)-(5):

$$\hat{y}\_i = \sum\_k\^K \blacksquare \ [\![f\_k\ (x\_i\ )\ ]\!] \quad (3)$$

$$L(\Theta) = \sum\_i \blacksquare l\ (y\_i - \hat{y}\_i\ ) + \sum \blacksquare k\ \Omega(f\_k\ ) \quad (4)$$

$$\Omega(f\_k\ ) = \gamma T + 1/2\ \gamma |(|\varpi|)|\^2 \quad (5)$$

Where f_k is additive function that represents the tree, and K additive function to form new data as a prediction result. The notation $(y\_i - \hat{y}\_i\ )$ is to determine the difference between the prediction value $\hat{y}\_i$ and predictor value y_i. Notation $\Omega(f\_k\ )$ is a complexity model, the notation T represents the number of trees on each node and ϖ represents the value of each tree [29]. OK uses structural analysis and variogram to assess the weight of the location that is not the point of observation in all spatial fields [30]. The OK as (6):

$$\gamma(h) = 1/2N(h)\ \sum\_(i = 1)\^N(h) \blacksquare [Z(x\_i\ ) - Z(x\_i - h)]\^2 \quad (6)$$

Where $\gamma(h)$ is a notation from semi variogram, *h* is a notation from the distance lag, *N (h)* is a notation of the number of observation points with distance h, $Z(x\_i\ )$ and $Z(x\_i - h)$ is a regionalized variable [31]. The accuracy of the classification and optimization of RF and XGboost is carried out using the overall accuracy method and the Kohen Kappa method. The accuracy overall as (7):

$$OA = \sum\_(i = 1)\^B \blacksquare x/n \quad (7)$$

which B is the number of classes used in classification, value x is the amount of data testing and value n is the amount of data analyzed [32]. The Kohen Kappa as (8):

$$K = (x\_0 - x\_e)/(1 - x\_e\ ) \quad (8)$$

where x_0 is its accuracy and x_e is its probability [32]. The proposed computer model algorithms as the novelty of finding of this study is shown are:

```
Begin
Data Sentinel-2 ESA as Numeric = {ML, QL, AL, MρQcal, Aρ}
Aster G-DEM as Numeric = {Elavation}
Process radiometric correction:
        Step 1: The conversion data DN to TOA Radiance: Lλ = MLQcal + AL
        Step 2: The conversion data DN to TOA Reflectance: ρλ'= MρQcal + Aρ
Calculate process atmospheric correction:
        ρ*(λ) =ρr(λ) + ρa(λ) + ρra(λ) + T(λ) ρg(λ) + t(λ) ρwc(λ) + t(λ) ρBOA(λ);
Calculate process geometric correction;
Calculate process Built-Up Indices:
       NBI= (ρ_Red ∗ ρ_SWIR2)/ρ_NIR ; UI= (ρ_SWIR2 − ρ_NIR )/(ρ_SWIR2 + ρ_NIR ) ; NDBI= (ρ_SWIR1 −
       ρ_NIR )/(ρ_SWIR1 + ρ_NIR ); MBI= (ρ_SWIR1 ∗ ρ_Red − (ρ_NIR ∗ ρ_NIR ))/(ρ_Red + ρ_NIR + ρ_SWIR1 ) ;
       IBI=((2xρ_SWIR1)/(ρ_SWIR1 + ρ_NIR ) − (ρ_NIR/(ρ_NIR + ρ_Red )) − (ρ_Green/(ρ_Green + ρ_SWIR1 )))/
       ((2xρ_SWIR1)/(ρ_SWIR1 + ρ_NIR ) + (ρ_NIR/(ρ_NIR + ρ_Red )) − (ρ_Green/(ρ_Green + ρ_SWIR1 ))) ;
Calculate the Random Forest algorithms process:
        G_i = 1 − ∑_(r = 1)^n▒ 〖S_n^2, r〗
Calculate the XGBoost algorithms process:
        Ω(f_k ) = γT + 1/2 γ|(|ϖ|)|^2
Calculate the accuration Prediction Count process:
        Overall Accuracy: OA = ∑_(i = 1)^B▒x/n ; Kohen Kappa: (x_0 − x_e)/(1 − x_e )
Calculate Ordinary Kriging: γ(h) = 1/2N(h) ∑_(i = 1)^N(h)▒[Z(x_i ) − Z(x_i − h)]^2
Display matrix decision assesment scale: x̄ = (∑▒x)/n
Confusion Matrix (CM) values:
Begin
        If CM index < 1,39
                than Very Low;
        else
        If CM index < 1,40 – 2,39
                than Low;
        else
        If CM index < 1,40 – 3,39
                than High;
        else
        If CM index > 3,39
                than Very High;
        EndIf
End
```

The computer model was proposed using a framework consisting of 3 layers, namely: i) pre-processing, ii) analytical data, and iii) interpretation. The pre-processing layer consists of atmospheric, radiometric and geometric correction processes, image extraction using the NBI, UI, NDBI, MBI, IBI algorithms and the Raster Statistics for Polygon function, producing built-up indices numeric data. The data analytical layer consists of data classification process using the CKA method, optimization classification process using machine learning, and an accuracy test process using the Kohen Kappa function and overall accuracy. Layer interpretation is the process of spatial distribution using OK and the output classification process in 4 scales, namely: very low, low, high, and very high (Figure 1). The proposed computer model framework as the novelty of finding of this study is shown in Figure 1.

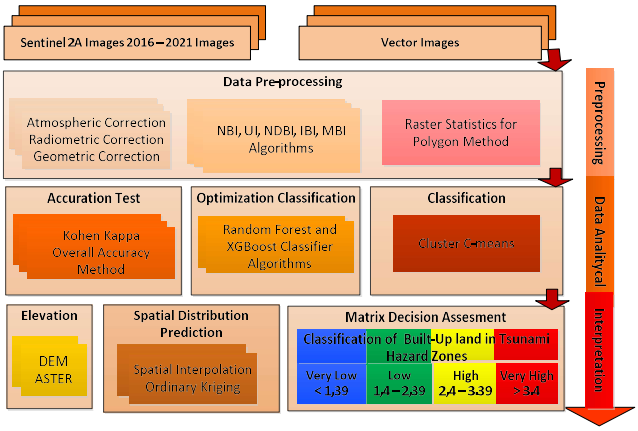


Figure 1. Framework computer model for assessing the impact of the tsunami on built-up indices optimized by machine learning classifier

2. METHOD

This study was conducted on the southern coast of Central Java Province, Indonesia. The coordinates of the study area are: Purworejo Regency: 109 ° 47'-110 ° 08' and 7 ° 32'-7 ° 54', Kebumen Regency: 109 ° 33'-109 ° 50' and 7 ° 27'-7 ° 50', Cilacap Regency: 109°–109°30'and 7°30'–7°50'. The observed areas include 57 villages which consist of 39 villages in Purworejo Regency area, 7 villages in Kebumen Regency, and 11 villages in Cilacap Regency. land use in the study area is classified into 8 types, namely: rivers/ponds fisheries, forests/vegetation, grass, mixed agriculture, built-up lands, bare lands, mix of built-up, and scrub/srub. The data used for this study is remote sensing imagery of the Sentinel 2A satellite for the 2016–2021 observation period. The built-up indices equation can be seen in Table 1.

Table 1. The built-up indices equation used in the study

Ref.	Built-up indices	Formula	
[33]	New built-up index	$NBI = (\rho\_Red * \rho\_SWIR2)/\rho\_NIR$	(9)
[34]	Urban index	$UI = (\rho\_SWIR2 - \rho\_NIR)/(\rho\_SWIR2 + \rho\_NIR)$	(10)
[35]	Normalized difference build-up index	$NDBI = (\rho\_SWIR1 - \rho\_NIR)/(\rho\_SWIR1 + \rho\_NIR)$	(11)
[36]	Modified built-up index	$MBI = (\rho\_SWIR1 * \rho\_Red - (\rho\_NIR * \rho\_NIR))/(\rho\_Red + \rho\_NIR + \rho\_SWIR1)$	(12)
[37]	Index based built-up index	$IBI = ((2x\rho\_SWIR1)/(\rho\_SWIR1 + \rho\_NIR) - (\rho\_NIR/(\rho\_NIR + \rho\_Red)) - (\rho\_Green/(\rho\_Green + \rho\_SWIR1)))/((2x\rho\_SWIR1)/(\rho\_SWIR1 + \rho\_NIR) + (\rho\_NIR/(\rho\_NIR + \rho\_Red)) - (\rho\_Green/(\rho\_Green + \rho\_SWIR1)))$	(13)

Coastal elevation is one of indicators that must be analyzed to determine tsunami vulnerability in addition to land use and land cover characteristics. Elevation is modeled using digital elevation model (DEM) aster imagery and elevation interpretation as shown in Table 2 [38].

Table 2. Elevation and level of vulnerability [38]

Elevation (m)	Vulnerability
<5	Very high
5-10	High
10-15	Medium
15-20	Low
>20	Very low

3. RESULT AND DISCUSSION

The first indicator for assessing building damage as a result of the tsunami in this computer model is the built-up area along the coast. Determining the area of built-up land is carried out by the mechanism of extraction, identification and separation of DN values from Sentinel 2A imagery for the built-up land

category from DN pixels for fisheries, forest and agriculture categories using the supervised classification method. This process will produce a map image with raster data format. The raster data format is converted to a vector data format to make it easier to calculate the land area. The comparison of the land area that is in the category of built-up land within a period of 5 years, namely 2017–2021 is shown in Figure 2.

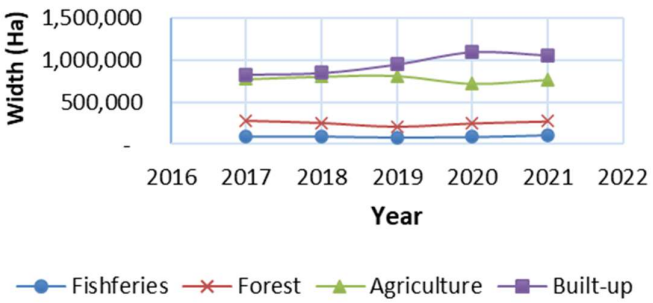


Figure 2. Comparison of built-up land with fisheries, forest and agriculture in 5 years period of 2017-2021

Based on the analysis of Sentinel 2A imageries, there has been an increase in the built-up areas of 226,683 m^2 in 5 years period. The increase of built-up lands includes the construction of settlements, offices, trade buildings, industrial buildings, and social and public facilities in the study area. The indicator of potential damage to buildings can be identified using built-up indices, namely IBI, NBI, MBI, NBAI, DBI and UI. The indicator of built-up indices is the DN value of each pixel as a representation of a built-up area based on the wavelength reflected and received by the satellite sensor. The IBI value which increases every year indicates a change in function and land use from previously a vegetated or vacant land to built-up land both in the form of physical buildings and changes in the land use. These changes can be detected because IBI is actually composed of 3 other indices namely soil-adjusted vegetation index (SAVI), modified normalized difference water index (MNDWI) and NDBI indices [38]. The range of built-up indices is between -1.00 to +1.00, which negative values represent areas dominated by surface water and vegetation while positive values represent areas dominated by physical buildings. In Figure 3(a) the mean value of built-up indices IBI is 0.080 and Figure 3(b) the mean value of built-up indices MBI is 0.077 representing the dominance of built-up lands, a small portion of open lands and surface waters in the study area. High values of built-up indices represent the complexity of settlements, industry or tourism, while low values represent open lands. In Figures 3(c) to (f) the mean value of built-up indices MBI is -0.093, NBAI is -0.241, NDBI is -0.419 and UI is -0.204 which represents the domination of water surface areas such as paddy fields and aquaculture, vegetations such as shrubs, plantations and forests.

Built-up indices data are classified into 4 categories based on built-up land density indicators using the CKA algorithm, namely very low, low, high, and very high. The purpose of this classification process is to provide labels in the form of values for each observation or sampling area for each built index indicator. Testing of the results of the classification is carried out using machine learning RF and XGboost.

The results of testing with machine learning predict the distribution pattern using OK method. The spatial pattern of building density in the study area using the RF algorithm in; 2016 (Figure 4(a)), 2020 (Figure 4(b)), and 2021 (Figure 4(c)). The RF algorithm works by forming vectors from the input data, namely build-up indices which are denoted as $\Theta\_1, \Theta\_2, \Theta\_3 \ldots \Theta\_(r-1)$. Next, the formed vectors are randomly selected as training data which are notated as training data and some of them become testing data before being calculated with the Gini index. The built-up indices in the RF algorithm are nodes that are described as trees which each tree has branches of very low cluster, low cluster, high cluster, and very high cluster. The results of the analysis using the RF algorithm can be seen in Figure 4. Figure 4 shows comparison of the spatial patterns of distribution of building density in the study area using the RF algorithm between 2016 (Figure 4(a)), 2020 (Figure 4(b)), and 2021 (Figure 4(c)) based on built-up indices data. Data of 2016 shows that study area Figure 4(a) has a low to very low building density which is shown in green to blue. In most areas of study area Figure 4(b). the density of built-up land is still high to very high (yellow and red). In 2021, there will be a higher increase compared to 2020, the study area Figure 4(c) has a high to very high building density which is shown in yellow to red colors. In most areas study, they show a high to very high built-up density (yellow and red colors).

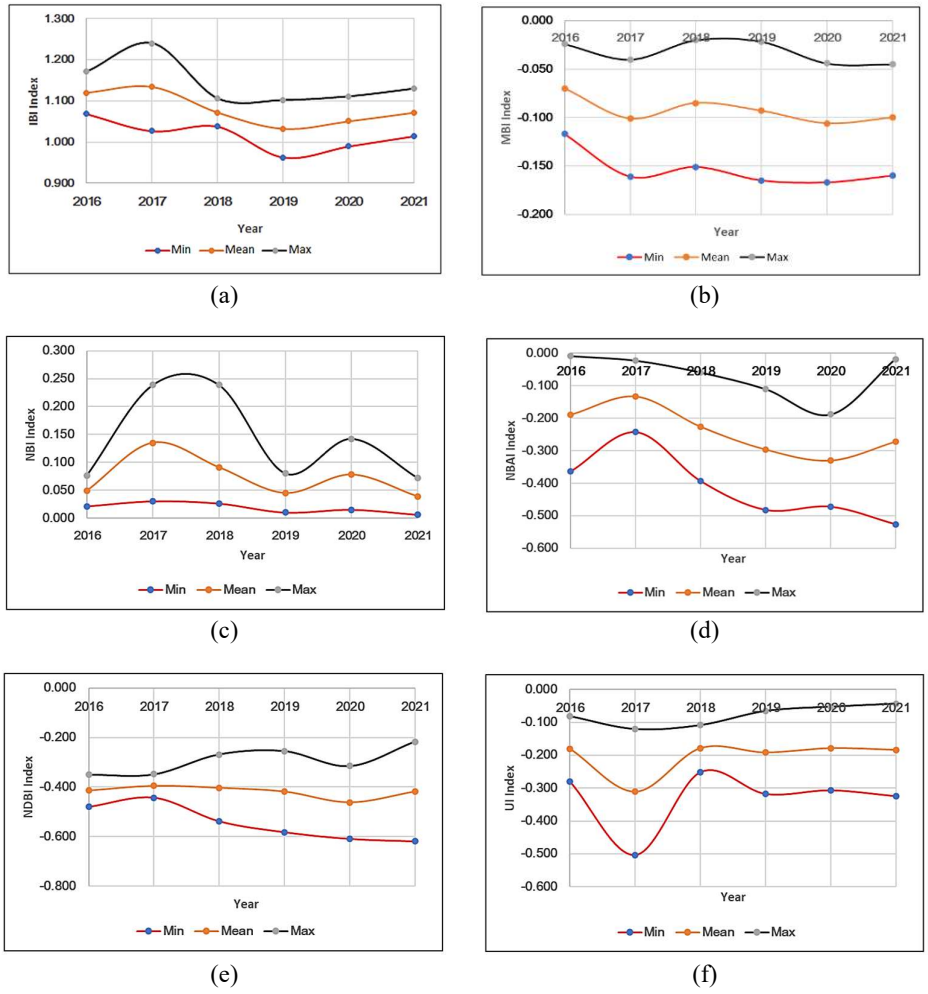


Figure 3. Time series data of built-up indices: (a) IBI, (b) MBI, (c) NBI, (d) NBAI, (e) NDBI, and (f) UI from 2017-2021

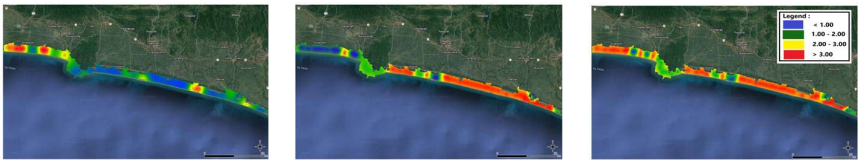
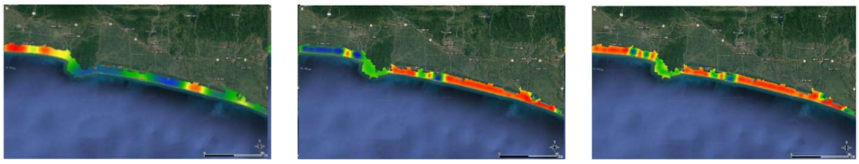


Figure 4. The spatial pattern of building density in the study area using the RF algorithm year: (a) 2016, (b) 2020, and (c) 2021

Figure 5 shows the comparison of the spatial pattern of building density in the study area using the XGBoost algorithm between: 2016 (Figure 5(a)), 2020 (Figure 5(b)), and 2021 (Figure 5(c)) based on built-up indices data. The XGBoost algorithm works through structured data processing using a decision tree, namely testing the data attributes (built up vegetation indices) in each node using the criteria for very low cluster, low cluster, high cluster and very high cluster, and the test results are represented on each branch. The results of the XGBoost analysis are almost the same as that of the RF. In 2016, it can be seen that study area has a low to very low building density which is shown in green to blue colors. In 2021 there will be a higher increase compared to 2016. The areas study has a high to very high building density which is shown in yellow to red colors. Testing the accuracy of the results of RF and XGboost analysis is carried out using 2 methods, namely overall accuracy and Kohen Kappa.

The test results can be seen that the RF and XGBoost tests using the overall accuracy and Kohen Kappa methods have an accuracy above 80% so that classification and prediction are spatially very valid. RF have a test result 0.913 of overall accuracy and 0.866 of Kohen Kappa. XGBoost have a test result 0.947 of overall accuracy and 0.921 of Kohen Kappa. The final decision that will be represented in the computer model is made using the decision matrix method, which this method requires to form a scale for each variable (Table 3). The variable of build-up density is described in a scale as 1 for a very low density, 2 for a low density, 3 for a high density and 4 for a very high density. The elevation variable is described in a scale as 1 for a very low vulnerability, 2 for a low vulnerability, 3 for a medium vulnerability, 4 for a high vulnerability, and 5 for a very high vulnerability. The variable of accuracy is formulated in a scale > 0.9 for 1 and > 0.9 for 2. Based on the decision matrix in Table 3. it can be seen that the most accurate algorithm is XGBoost when it is compared to RF, and in both algorithms. It can be seen that 2020 and 2021 have a very high level of tsunami vulnerability due to their high density of buildings and due to a very low elevation of < 5 m above sea level. An assessment scale is produced by dividing the number of a value with the variable it uses so that the assessment scale is produced from the lowest value of 1 and the highest value of 4. Based on the experiments, it can be seen that the most optimal and accurate algorithm is XGBoost because it produces an assessment scale of a very high vulnerability area with a scale of 3.6, while the RF algorithm shows an assessment scale of a high vulnerability area with a scale of 3.3.



(a) (b) (c)

Figure 5. The spatial pattern of building density in the study area using the XGboost algorithm year: (a) 2016, (b) 2020, and (c) 2021

Table 3. The confusion matrix on computer model for a decision of tsunami high vulnerability assessment

Algorithm	Year of data	Build–up density	Elevation	Accuracy	Assessment scale	Symbol
RF	Data 2016	1	5	1	2.3	Blue
RF	Data 2020	4	5	1	3.3	Yellow
RF	Data 2021	4	5	1	3.3	Yellow
XGBoost	Data 2016	1	5	2	2.6	Green
XGBoost	Data 2020	4	5	2	3.6	Red
XGBoost	Data 2021	4	5	2	3.6	Red

4. CONCLUSION

The results of the study show that the built-up indices IBI value of 0.080 and NBI value of 0.077 as built-up representing the dominance of built-up lands, a small portion of open lands and surface waters in the study area. High values of built-up indices represent the complexity of settlements, industrial areas and tourism areas, while low values represent open lands. The mean value of the built-up indices of MBI (-0.093), NBAI (-0.241), NDBI (-0.419) and UI (-0.204) represents the dominance of surface water areas such as paddy fields and aquacultural areas, vegetations such as shrubs, plantations and forests. Testing the performance accuracy of machine learning RF using the overall accuracy method shows the value of 0.913 with Kohen Kappa of 0.866 which indicates that the classification of built-up indices data is very valid. Testing the performance accuracy of XGBoost machine learning using overall accuracy shows the value of 0.947 and Kohen Kappa of 0.921 also shows that the classification of data built-up indices is very valid.

ACKNOWLEDGEMENTS

The authors would like to thank Education and Culture Ministry Republic Indonesia for Grant Research Schema Fundametal Research Regular (PFR), Contract Number No. 182/E5/PG.02.00.PL/2023, Date June, 19 2023 and No. 001/LL6/PB/AL.04/2023, Date June 20, 2023.

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," *Geoscience Letters*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40562-021-00190-y.

[2] M. Y. Rezaldi *et al.*, "Unmanned aerial vehicle (Uav) and photogrammetric technic for 3d tsunamis safety modeling in cilacap, indonesia," *Applied Sciences (Switzerland)*, vol. 11, no. 23, pp. 1–15, 2021, doi: 10.3390/app112311310.

[3] 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.

[4] J. Behrens and F. Dias, "New computational methods in tsunami science," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. Royal Society of London, Oct. 2015 , doi: 10.1098/rsta.2014.0382.

[5] E. Rupnik, F. Nex, I. Toschi, and F. Remondino, "Contextual classification using photometry and elevation data for damage detection after an earthquake event," *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 543–557, 2018, doi: 10.1080/22797254.2018.1458584.

[6] S. Mateen, N. Nuthammachot, and K. Techato, "Billion Tree Tsunami Forests Classification Using Image Fusion Technique and Random Forest Classifier Applied to Sentinel-2 and Landsat-8 Images : A Case Study of Garhi," *ISPRS International Journal of Geo-Information*, vol. 12, no. 1, p. 9, 2023.

[7] W. L. Farahdita and H. S. R. Siagian, "Analysis of the area affected by the tsunami in Pandeglang, Banten: A case study of the Sunda Strait Tsunami," *IOP Conference Series: Earth and Environmental Science*, vol. 429, no. 1, p. 012052, 2020, doi: 10.1088/1755-1315/429/1/012052.

[8] H. Hébert, P. E. Burg, R. Binet, F. Lavigne, S. Allgeyer, and F. Schindelé, "The 2006 July 17 Java (Indonesia) tsunami from satellite imagery and numerical modelling: A single or complex source?," *Geophysical Journal International*, vol. 191, no. 3, pp. 1255–1271, 2012, doi: 10.1111/j.1365-246X.2012.05666.x.

[9] J. Sublime and E. Kalinicheva, "Automatic post-disaster damage mapping using deep-learning techniques for change detection: Case study of the Tohoku tsunami," *Remote Sensing*, vol. 11, no. 9, p. 1123, 2019, doi: 10.3390/rs11091123.

[10] I. E. Mulia, N. Ueda, T. Miyoshi, A. R. Gusman, and K. Satake, "Machine learning-based tsunami inundation prediction derived from offshore observations," *Nature Communications*, vol. 13, no. 1, 2022, doi: 10.1038/s41467-022-33253-5.

[11] E. A. Okal, "The south of Java earthquake of 1921 September 11: A negative search for a large interplate thrust event at the Java Trench," *Geophysical Journal International*, vol. 190, no. 3, pp. 1657–1672, 2012, doi: 10.1111/j.1365-246X.2012.05570.x.

[12] V. Steinritz, S. Pena-Castellnou, G. I. Marliyani, and K. Reicherter, "GIS-based study of tsunami risk in the Special Region of Yogyakarta (Central Java, Indonesia)," in *IOP Conference Series: Earth and Environmental Science*, Oct. 2021, vol. 851, no. 1, p. 012007 , doi: 10.1088/1755-1315/851/1/012007.

[13] S. Brune, A. Y. Babeyko, S. Ladage, and S. V. Sobolev, "Landslide tsunami hazard in the Indonesian Sunda Arc," *Natural Hazards and Earth System Science*, vol. 10, no. 3, pp. 589–604, 2010, doi: 10.5194/nhess-10-589-2010.

[14] Y. Rizal *et al.*, "Tsunami Evidence in South Coast Java, Case Study: Tsunami Deposit along South Coast of Cilacap," in *IOP Conference Series: Earth and Environmental Science*, Jun. 2017, vol. 71, no. 1, p. 012001 , doi: 10.1088/1755-1315/71/1/012001.

[15] D. M. Salmanidou, A. Ehara, R. Himaz, M. Heidarzadeh, and S. Guillas, "Impact of future tsunamis from the Java trench on household welfare: Merging geophysics and economics through catastrophe modelling," *International Journal of Disaster Risk Reduction*, vol. 61, Jul. 2021, doi: 10.1016/j.ijdrr.2021.102291.

[16] F. A. T. Laksono, A. Widagdo, M. R. Aditama, M. R. Fauzan, and J. Kovács, "Tsunami Hazard Zone and Multiple Scenarios of Tsunami Evacuation Route at Jetis Beach, Cilacap Regency, Indonesia," *Sustainability (Switzerland)*, vol. 14, no. 5, p. 2726, Mar. 2022, doi: 10.3390/su14052726.

[17] K. Prasetyo and Daryono, "Tsunami Disaster Mitigation For Population at South Coatal –Munjungan Distric-Trenggalek-East Java," in *Proceedings of the International Conference on Social Science 2019 (ICSS 2019)*, 2019, pp. 295–300 , doi: 10.2991/icss-19.2019.184.

[18] F. Usman, K. Murakami, A. D. Wicaksono, and E. Setiawan, "Application of Agent-Based Model Simulation for Tsunami Evacuation in Pacitan, Indonesia," in *MATEC Web of Conferences*, 2017, pp. 1–16 , doi: 10.1051/matecconf/20179701064.

[19] D. C. Istiyanto, S. Tanaka, T. Okazumi, and Syamsidik, "Towards Better Mitigation of Tsunami Disaster in Indonesia," in *Proceedings of the International Symposium on Engineering Lessons Learned from the 2011 Great East Japan Earthquake, March 1-4, 2012, Tokyo, Japan*, 2012, pp. 556–567.

[20] BMKG, *Indonesian Tsunami Catalog 416-2018 in bahasa: Katalog Tsunami Indonesia Tahun 416-2018*. Indonesia: Badan Meteorologi Klimatologi dan Geofisika, 2019.

[21] G. Cheng and J. Wei, "Color Quantization Application Based on K-Means in Remote Sensing Image Processing," in *Journal of Physics: Conference Series*, 2019, pp. 1–6 , doi: 10.1088/1742-6596/1213/4/042012.

[22] A. D. P. Pacheco, J. A. D. S. Junior, A. M. Ruiz-Armenteros, and R. F. F. Henriques, "Assessment of k-nearest neighbor and random forest classifiers for mapping forest fire areas in central portugal using landsat-8, sentinel-2, and terra imagery," *Remote Sensing*, vol. 13, no. 7, p. 1345, Apr. 2021, doi: 10.3390/rs13071345.

[23] I. Ali, A. Ur Rehman, D. M. Khan, Z. Khan, M. Shafiq, and J. G. Choi, "Model Selection Using K-Means Clustering Algorithm for the Symmetrical Segmentation of Remote Sensing Datasets," *Symmetry*, vol. 14, no. 6, pp. 1–19, 2022, doi: 10.3390/sym14061149.

[24] H. Guan, J. Yu, J. Li, and L. Luo, "Random Forests-Based Feature Selection for Land-Use Classification Using Lidar Data and Orthoimagery," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXIX-B7, pp. 203–208, 2012, doi: 10.5194/isprsarchives-xxxix-b7-203-2012.

[25] J. Zhang, S. Li, and M. Wang, "Water Depth Inversion based on Landsat-8 Date and Random Forest Algorithm," in *Journal of Physics: Conference Series*, 2020, pp. 1–8 , doi: 10.1088/1742-6596/1437/1/012073.

[26] R. S. Walker and M. J. Hamilton, "Machine learning with remote sensing data to locate uncontacted indigenous villages in Amazonia," *PeerJ Computer Science*, pp. 1–11, 2019, doi: 10.7717/peerj-cs.170.

[27] J. Park, Y. Lee, and J. Lee, "Assessment of machine learning algorithms for land cover classification using remotely sensed data," *Sensors and Materials*, vol. 33, no. 11, pp. 3885–3902, 2021, doi: 10.18494/SAM.2021.3612.

[28] R. Singh, M. Biswas, and M. Pal, "Cloud Detection using Sentinel 2 Imageries: A comparison of XGBoost, RF, SVM and CNN algorithms," *Geocarto International*, pp. 1–32, 2022, doi: 10.1080/10106049.2022.2146211.

[29] Y. Zhang and L. Chen, "A Study on Forecasting the Default Risk of Bond Based on XGboost Algorithm and Over-Sampling Method," *Theoretical Economics Letters*, vol. 11, no. 02, pp. 258–267, 2021, doi: 10.4236/tel.2021.112019.

[30] I. Farooq *et al.*, "Comparison of Random Forest and Kriging Models for Soil Organic Carbon Mapping in the Himalayan Region of Kashmir," *Land*, vol. 11, no. 12, pp. 1–15, 2022, doi: 10.3390/land11122180.

[31] Z. Liu and T. Yan, "Comparison of spatial interpolation methods based on ArcGIS," in *Journal of Physics: Conference Series*, 2021, vol. 1961, doi: 10.1088/1742-6596/1961/1/012050.
[32] M. Werther *et al.*, "Meta-classification of remote sensing reflectance to estimate trophic status of inland and nearshore waters," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 176, pp. 109–126, 2021, doi: 10.1016/j.isprsjprs.2021.04.003.
[33] C. Masinde, D. Rono, and M. Hahn, "Estimation of the degree of surface sealing with Sentinel 2 data using building indices," *Earth Observation and Geomatics Engineering*, vol. 3, no. 1, pp. 112–119, 2019, doi: 10.22059/eoge.2019.290064.1064.
[34] 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," *European Journal of Remote Sensing*, vol. 52, no. 1, pp. 240–252, Jan. 2019, doi: 10.1080/22797254.2019.1584737.
[35] G. Kuc and J. Chormański, "Sentinel-2 imagery for mapping and monitoring imperviousness in urban areas," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. XLII-1/W2, pp. 43–47, 2019, doi: 10.5194/isprs-archives-XLII-1-W2-43-2019.
[36] N. Mostofi and M. Hasanlou, "Feature Selection of Various Land Cover Indices For Monitoring Surface Heat Island in Tehran City Using Landsat 8 Imagery," *Journal of Environmental Engineering and Landscape Management*, vol. 25, no. 3, pp. 241–250, 2017, doi: 10.3846/16486897.2012.721784.
[37] W. Guo, C. Zhao, Y. Zhang, and S. Gao, "Mapping Impervious Surface Distribution and Dynamics in an Arid/Semiarid Area-A Case Study in Ordos, China," *IEEE Access*, vol. 9, pp. 19659–19673, 2021, doi: 10.1109/ACCESS.2021.3054963.
[38] T. A. Kebede, B. T. Hailu, and K. V. Suryabhagavan, "Evaluation of spectral built-up indices for impervious surface extraction using Sentinel-2A MSI imageries: A case of Addis Ababa city, Ethiopia," *Environmental Challenges*, vol. 8, 2022, doi: 10.1016/j.envc.2022.100568.

BIOGRAPHIES OF AUTHORS

Sri Yulianto Joko Prasetyo completed his doctorate degree on the Doctorate Program of Computer Science, Science Faculty of Gadjah Mada University in 2013. He has been active on research since 2008 until now on the Spatial Data Processing and Remote Sensing. He has published his papers on international journals Scopus Indexed 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 management and technology. He can be contacted at email: sri.yulianto@uksw.edu.

Bistok Hasiholan Simanjuntak 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, geographic 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 bistok@uksw.edu.

Yerymia Alfa Susetyo is currently working as Lecturer and Researcher at Department of Informatics, Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia. He has completed his master in Information System from Satya Wacana Christian University, Indonesia. His main area of interest focuses geospatial modeling and distributed system. His area of expertise includes software engineering, web programming, services programming, geospatial modeling, and geospatial programming. He can be contacted at yeremia.alfa@uksw.edu.

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