Vegetation indices' spatial prediction based novel algorithm for determining tsunami risk areas and risk values

Kristoko Dwi Hartomo [Corresp., 1] , **Yessica Nataliani** [1] , **Zainal Arifin Hasibuan** [2]

[1] Department of Information System, Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia

[2] Faculty of Computer Science, University of Dian Nuswantoro, Semarang, Indonesia

Corresponding Author: Kristoko Dwi Hartomo
Email address: kristoko@uksw.edu

This paper aims to propose a new algorithm to detect tsunami risk areas based on spatial modeling of vegetation indices and a prediction model to calculate the tsunami risk value. It employs atmospheric correction using DOS1 algorithm combined with k-NN algorithm to classify and predict tsunami-affected areas from vegetation indices data that have spatial and temporal resolutions. Meanwhile, the model uses the vegetation indices (i.e., NDWI, NDVI, SAVI), slope, and distance. The result of the experiment compared to other classification algorithms demonstrates good results for the proposed model. It has the smallest MSEs of 0.0002 for MNDWI, 0.0002 for SAVI, 0.0006 for NDVI, 0.0003 for NDWI, and 0.0003 for NDBI. The experiment also shows that the accuracy rate for the prediction model is about 93.62%.

1 # Vegetation Indices' Spatial Prediction based Novel
2 # Algorithm for Determining Tsunami Risk Areas and
3 # Risk Values

5 Kristoko Dwi Hartomo[1], Yessica Nataliani[2], Zainal Arifin Hasibuan[3]

7 [1,2] Department of Information System, Faculty of Information Technology, Satya Wacana
8 Christian University, Salatiga, Central of Java, Indonesia
9 [3] Faculty of Computer Science, University of Dian Nuswantoro, Semarang, Central of Java,
10 Indonesia

13 Corresponding Author:
14 Kristoko Dwi Hartomo[1]
15 Diponegoro 52-60, Salatiga, Central of Java, 50711, Indonesia
16 Email address: kristoko@uksw.edu

19 ## Abstract
20 This paper aims to propose a new algorithm to detect tsunami risk areas based on spatial
21 modeling of vegetation indices and a prediction model to calculate the tsunami risk value. It
22 employs atmospheric correction using DOS1 algorithm combined with k-NN algorithm to
23 classify and predict tsunami-affected areas from vegetation indices data that have spatial and
24 temporal resolutions. Meanwhile, the model uses the vegetation indices (i.e., NDWI, NDVI,
25 SAVI), slope, and distance. The result of the experiment compared to other classification
26 algorithms demonstrates good results for the proposed model. It has the smallest MSEs of 0.0002
27 for MNDWI, 0.0002 for SAVI, 0.0006 for NDVI, 0.0003 for NDWI, and 0.0003 for NDBI. The
28 experiment also shows that the accuracy rate for the prediction model is about 93.62%.

30 ## Introduction
31 Tsunami is one of the disaster threats for many coastal areas in Indonesia. This disaster is
32 generally triggered by an earthquake at sea causing a vertical shift in the seabed (Amri et al.,
33 2018). The territories of Indonesia are surrounded by the meeting of the world's main tectonic
34 plates which causes the tsunami. These plates are: (1) the Indian-Australian Ocean plate in the
35 south, moving relatively to the north and presses the Eurasian plate (where most of Indonesia
36 territories are located), and (2) the Pacific plate in the east, which moves relatively westward
37 against the Eurasian plate. This phenomenon causes many sources of earthquakes as well as the
38 growth of active volcanoes in the territories of Indonesia, thus placing Indonesia as one of the
39 areas in the world's most active tectonic zone (Verstappen, 2010).

41 The National Development Planning Agency of Indonesia (Bappenas) stated that the total loss
42 and damage from the tsunami and earthquake in Yogyakarta and Aceh Province was Rp 70.5
43 trillion. The impacts were 80 percent of the infrastructure sector damage (including housing) and
44 11 percent of the productive sector damage (Regulation of the Ministry of Health Republic of
45 Indonesia No. 36 of 2014 about Assessment of Post-Disaster Damage, Loss and Health Resource
46 Needs, 2014). The tsunami itself caused the death toll in the affected area reaching 108,100
47 people and 127,700 people missing (Amri et al., 2018).

48

49 The province of Yogyakarta that has the potential for a tsunami is located in the south coast area.
50 The disaster risk index in Yogyakarta for the tsunami disaster currently reaches 1.74. This value
51 is seen from the vulnerability of the Yogyakarta region from the tsunami disaster by considering
52 geological aspects, tsunami threat maps and demographics in villages that may be exposed to the
53 impact of the tsunami. In general, the Yogyakarta region has a fairly high tsunami disaster index,
54 especially areas directly adjacent to the coastline (Regional Disaster Management Agency,
55 2019). Kulon Progo Regency is the most susceptible to tsunami disaster among the four districts
56 in Yogyakarta. There is a possibility of 60,607 people that will be affected if a tsunami occurs in
57 this district (National Disaster Management Agency, 2012). Areas with a high tsunami risk in
58 Kulon Progo Regency are Temon Sub-District, Wates Sub-District, Panjatan Sub-District,
59 Lendah Sub-District, and Galur Sub-District (Regulation of the Special Region of Yogyakarta
60 No. 5 of 2019 about The Spatial Plan of the Special Region of Yogyakarta for 2019-2039, 2019).
61 Kulon Progo Regency comprises of 47 villages (see Figure 1 for further information on the name
62 of the villages) (Mustaqim, 2019). Figure 1 shows the visualization of the map of Kulon Progo
63 Regency. Figure 1 is obtained from the mapping of Shuttle Radar Topography Mission (SRTM)
64 image from https://earthexplorer.usgs.gov/.

65

66 Figure 1. The Visualization of Kulon Progo Regency Map

67

68 Damage and impacts caused by disasters, especially tsunamis, can be analyzed using the field
69 survey method. The field survey method used during a disaster is difficult and carries a high risk.
70 Disasters can occur over a large area, so surveying the entire affected area takes a long time
71 (Singh et al., 2014). Obtaining fast and accurate information in the event of a disaster can save
72 many human lives and reduce losses. The solution for problems in field survey is the use of
73 satellite technology. With the availability of easily obtained satellite images, remote sensing
74 techniques can be widely used to assess disaster risk areas (Brunner et al., 2010). The use of
75 satellite imagery can overcome the limitations of collecting disaster area image data using
76 traditional methods. Satellite images not only speed up the process of analyzing disaster areas
77 but also provide accurate and timely estimates of the risk of disaster affected areas (Koshimura et
78 al., 2020).

79

80 Landsat 8 satellite captures land cover images in disaster-affected areas that are influenced by
81 soil and vegetation characteristics, in which this process becomes the basis for remote sensing
82 techniques in analyzing land cover (Rendana et al., 2016). In the last five years, the vegetation
83 indices analysis has been applied in the evaluation of land cover. One of them is the analysis of
84 disaster-affected land as a variation of spatial and temporal analysis techniques (Holzman et al.,
85 2014; Mallick et al., 2009). Digital image processing from satellite data enables image analysis
86 through various algorithms and mathematical indices. Meanwhile, feature analysis is based on
87 the reflectance characteristics and the index has been designed to detect prominent features in the
88 image area (Xie et al., 2010). There are several indices that can detect areas containing
89 vegetation in the image obtained from remote sensing, such as Normalized Difference
90 Vegetation Index (NDVI), Normalized Difference Built-Up Index (NDBI), Soil Adjusted
91 Vegetation Index (SAVI), Normalized Difference Water Index (NDWI), and Modified
92 Normalized Difference Water Index (MNDWI). NDVI is a common and widely used index
93 (Karnieli et al., 2010). It is also applied considerably in global environment, climate change, and
94 disasters research (Gao, 1996).

95

96 The response of vegetation to the environment is considered very sensitive. It affects the
97 ecological and climate balances. In addition it becomes an effective barrier against natural
98 disasters. Classification of natural disaster images, especially images before and after the
99 tsunami, can use the average value of various spectrum indices such as NDVI, NDBI, SAVI,
100 MNDWI, and NDWI as training data (Singh et al., 2014). The pre- and post-tsunami images are
101 compared to obtain the inundated area. NDVI, NDBI, SAVI, MNDWI, and NDWI are calculated
102 as a ratio difference between red and near infrared bands respectively in measured canopy
103 reflectance (Hu et al., 2008). The process of analyzing the vegetation index data is carried out
104 using representation, categorization, and classification approaches to produce valid information.
105 The most suitable method to represent, categorize, and classify the vegetation index is a machine
106 learning (Maxwell et al., 2018).

107

108 The most recent development of remote sensing methods for impact analysis and detection of
109 image transformations caused by disasters is the use of machine learning. The machine learning-
110 based transformation detection method consists of two categories, they are supervised and
111 unsupervised detection methods. The supervised method requires training to identify
112 transformations, such as methods of instance support vector machine (SVM) (Bovolo &
113 Bruzzone, 2007; Volpi et al., 2012), post classification comparison (Yang & Wen, 2011),
114 artificial neural network (ANN), and function neural network (FNN) (Mehrotra et al., 2015).
115 Meanwhile, the unsupervised method does not require training as this method analyzes the image
116 to identify any transformations. The unsupervised method is usually used to differentiate images
117 (Wang et al., 2020) and make a principal component analysis (PCA) (Corner et al., 2000).
118

119   In this study, a new algorithm is proposed to predict tsunami risk areas based on spatial
120   modelling prediction of vegetation index. The new algorithm is developed with the addition of
121   atmospheric correction before the data is processed using the k-Nearest Neighbor (k-NN)
122   algorithm. In this algorithm, machine learning is used to classify and predict tsunami-affected
123   areas from vegetation indices data that has spatial and temporal resolutions. To evaluate the
124   performance of this new algorithm, it is compared to Cart, SVM, and ANN algorithms. The
125   experimental results show that this new framework is superior, as indicated by the Mean Squared
126   Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Cohen's
127   Kappa scores that outperform other algorithms. Furthermore, since there is only a category of
128   risk level and no value that indicates the level of tsunami risk, then a new formula using
129   vegetation indices and other parameters is proposed to determine the tsunami risk value.
130
131   This paper is presented as follows. Section 1 describes the background on tsunami problems.
132   Section 2 contains some related works and the reviews of machine learning and vegetation
133   indices. Section 3 presents the research method and flowchart, as well as the proposed algorithm.
134   A new algorithm for detecting tsunami risk areas is proposed, by adding atmospheric correction,
135   before applying prediction and classification algorithms, since some data may contain some
136   noises. This framework uses k-NN algorithm to classify the area, with high, medium, and low
137   risk. Section 4 explains the experiments and comparisons of the proposed framework using k-NN
138   algorithm with other classifying methods using real data. A model to predict the tsunami risk
139   value is also discussed in this section. Finally, the conclusions are stated in Sections 5.
140

Related Works

142   In this section, some research about remote sensing based on prediction and classification are
143   discussed. Prasetyo, et al. implemented spectral vegetation index to the data obtained from the
144   Landsat 8 OLI satellite to provide disaster risk index information (Prasetyo et al., 2020). k-NN
145   with spatial autocorrelation was used to classify the drought risk areas. The spectral vegetation
146   indices used in this study were NDVI, SAVI, Vegetation Condition Index (VCI), Temperature
147   Condition Index (TCI), and Vegetation Health Index (VHI) (Xue & Su, 2017). While, the Kappa
148   accuracy test showed that the SVM and k-NN methods had an accuracy of 88.30.
149

150   Other research were conducted a remote sensing study using machine learning methods for
151   classifying the image data (Ma et al., 2019; Maxwell et al., 2018). One of them wanted to
152   classify the distribution, species, and extent of mangroves using the Akaike Information
153   Criterion (AIC) on Nusa Lembongan Island (Ilham & Marzuki, 2017). This research was
154   important since mangrove forest had important ecological, economic, and social roles. One
155   example was the study on mangrove forest as a green belt for shoreline protection from storms
156   and tsunami waves using Worldview-2 satellite imagery with a data resolution of 0.46 meters.
157   This method automatically identified land classes, sea/water classes, and mangrove classes. The
158   results showed that the classification accuracy was 68.32%.

159

160 Another related study proposed a new method for classifying daily NDVI time series data based
161 on a combination of multi-classifiers (Zhao et al., 2017). In this study, the HJ-CCD satellite was
162 used as data for compiling an NDVI time series model with S-G filtering and spatial
163 interpolation. This study also proposed a dimension reduction method using the statistical
164 features of the daily NDVI time series. The results showed that the accuracy of image
165 classification of disaster-prone areas was 77.45%.

166

167 The next study used the unsupervised classification method to classify coastal forest damage due
168 to tsunamis (Yonezawa, 2015). The extent and distribution of tsunami damage was predicted
169 using NDVI. To assess the classification accuracy, they used the error matrix. Aerial photos were
170 used for reference. From 200 random points, the accuracy was 79.50% with a Kappa Stats of
171 0.6387.

172

173 Similar research proposed NDVI predictions recorded by satellite at Ventspils City in Courland,
174 Latvia and obtained using the Markov chain method (Stepchenko & Chizhov, 2016). In general,
175 Markov chain prediction is a probability forecasting method because the prediction results show
176 the probability of an NDVI value in the future (Liu, 2010). This study demonstrated how Markov
177 chains could predict future values with less memory and random walk capability. Each state was
178 reached directly by other states with a transition matrix to provide a high prediction accuracy of
179 63.93%.

180

181 Subsequent research proposed a method for mapping the value of environmental damage after a
182 disaster (Havivi et al., 2018). The data used are TerraSAR-X (TSX) images with high resolution
183 obtained from before and after the incident and also Landsat 5 images before the incident. The
184 affected areas were analyzed with Synthetic Aperture Radar (SAR) using one SAR
185 interferometric coherence map (InSAR). The accuracy of mapping the environmental damage
186 caused by the tsunami could be improved using the vegetation index (NDVI) (Ghebrezgabher et
187 al., 2020; Koshimura et al., 2020). The affected areas were mapped with the overall accuracy of
188 89% and Kappa coefficient of 82%.

189

190 In summary, it has been shown from this review that spectral vegetation indices indicated the
191 quantitative values for measuring the vegetation canopy in receiving and reflecting the light
192 spectrum. They were interpreted as plant's spectral characteristics, including the infrared
193 spectrum as visible light (IR) and the near infrared spectrum as invisible light (NIR) (Prasetyo et
194 al., 2020).

195

196 **Research Methods**

197 In this section, a research method for detecting the tsunami risk areas is introduced. Atmospheric
198 correction is used in this algorithm since data may contain some noises which will interfere with

199 the prediction and classification performance. The flowchart for this research method is depicted
200 in Figure 2 and can be explained as follows.
201
202 Figure 2. Flowchart of The Research Method
203
204 Figure 2 can be explained as below.
205 1.   Data Extraction
206 Landsat 8 image data that has been downloaded from https://earthexplorer.usgs.gov/ in the form
207 of digital values can be used for land use mapping. The extracted data consists of 11 bands,
208 namely Band 1, Band 2, …, Band 11, each with a different name and wavelength. Each band has
209 a wavelength between 0.43 – 12.51. Table 1 shows the Landsat 8 operational land image (OLI)
210 and thermal infrared sensor (TIRS).
211
212 Table 1. Landsat 8 Operational Land Image (OLI) and Thermal Infrared Sensor (TIRS) (U.S.
213 Geological Survey, 2021)
214
215 2.   Atmospheric Correction
216 The data obtained from data extraction still has a low level of radiometric accuracy because it
217 has errors in the recording process from the sensor in the image. Therefore, if the data will be
218 used for data processing such as biomass, vegetation indices, land-cover classification, and so
219 on, they require an atmospheric correction process. This process is used to improve the accuracy
220 of image classification so that the data obtained can be compared and arranged in a number of
221 solutions and evaluation of tsunami-prone areas. One of the processes is the removal of
222 atmospheric haze and cloud cover. The results of the atmospheric correction process are very
223 important for optimizing foggy satellite imagery regarding the object's surface to detect changes
224 in land cover and land use.
225
226 The method that can be used in the atmospheric correction process is the Dark Object
227 Subtraction (DOS), especially DOS1. This method is chosen because the field data parameters
228 for image correction are not known and also the atmospheric effect model is not known which
229 shows the condition of an image when the image is recorded (Zhang et al., 2010).

230 $$\rho = \frac{\pi(L_{sat} - L_p)d^2}{E_0 \cos(\theta_z)} \quad\quad (1)$$

231 where L_{sat} is the at-satellite radiance, L_p is the path radiance, ρ is the land surface reflectance, d
232 is the Earth–Sun distance in astronomical units, E_0 is the exoatmospheric solar spectral
233 irradiance, and θ_z is the solar zenith angle.
234
235 3.   Data Pre-processing
236 To begin with, the geometric and radiometric corrections need to be known. Geometric
237 correction is used to correct the position of the coordinates of each pixel in the image exactly as

238 displayed on the surface of the earth. It involves satellite movements, earth rotation, and terrain
239 effects. Meanwhile, radiometric correction has a different goal. The stage of data pre-processing
240 includes collecting Sentinel 2 satellite image data obtained from www.earthexplorer.usgs.gov.
241 They are corrected not only geometrically, but also radiometrically and atmospherically.

242

243 After the image is corrected, the clean band is calculated, according to the formula for each
244 index. Sentinel 2 image data extraction uses the NDVI, NDBI, NDWI, MSAVI, and MNDWI
245 formulas. The extraction results are numerical values, which can be used for the classification
246 and prediction using k-NN. NDVI, MNDWI, MSAVI, NDWI, and NDBI formulas are explained
247 as follows.

248 • Normalized Difference Vegetation Index (NDVI)
249   NDVI is a vegetation index that is often used to compare the level of vegetation greenness
250   (chlorophyll level) in plants (Min et al., 2016).

251 $$NDVI = \frac{NIR - Red}{NIR + Red} = \frac{Band\ 5 - Band\ 4}{Band\ 5\ + Band\ 4} \qquad (2)$$

252 • Normalized Difference Water Index (NDWI)
253   NDWI is an index that shows the wetness level of an area. The formula for the NDWI index
254   follows Eq. 3 (Xu, 2006).

255 $$NDWI = \frac{Green - NIR}{Green + NIR} = \frac{Band\ 3 - Band\ 5}{Band\ 3 + Band\ 5} \qquad (3)$$

256 • Modified Normalized Difference Water Index (MNDWI)
257   MNDWI is a modification of the NDWI index. The formula for MNDWI index follows Eq. 4
258   (Acharya et al., 2018).

259 $$MNDWI = \frac{Green - SWIR\ 1}{Green + SWIR\ 1} = \frac{Band\ 3 - Band\ 6}{Band\ 3\ + Band\ 6} \qquad (4)$$

260 • Soil Adjusted Vegetation Index (SAVI)
261   SAVI is used to correct NDVI for minimizing the influence of soil brightness in low
262   vegetative cover areas, using a soil-brightness correction factor (L). SAVI is calculated as a
263   ratio between the Red and NIR values with a soil brightness correction factor, where $L = 0.5$,
264   to accommodate most land cover types. The formula for SAVI index follows Eq. 5 (Huete,
265   1988).

266 $$SAVI = \left(\frac{NIR - Red}{NIR + Red + L}\right) * (1 + L) = \left(\frac{Band\ 5 - Band\ 4}{Band\ 5\ + Band\ 4 + 0.5}\right) * 1.5 \qquad (5)$$

267 • Normalized Difference Built-up Index (NDBI)
268   NDBI is an effective transformation/index used in mapping building lands in an area
269   automatically using Landsat 8 OLI images. The formula for NDBI index follows Eq. 6 (Zha
270   et al., 2003).

271 $$NDBI = \frac{SWIR\ 1 - NIR}{SWIR\ 1\ + NIR} = \frac{Band\ 6 - Band\ 5}{Band\ 6\ + Band\ 5} \qquad (6)$$

272

273 4. Correlation Test
274 Correlation test calculates the correlation between variables and analyzes the level of closeness
275 of the relationship between the independent variable (X) and the dependent variable (Y). Pearson

276 correlation is one of the correlation tests used to measure the strength of the linear relationship
277 between two variables. Two variables are said to be correlated if a change in one variable is
278 accompanied by a change in another variable, both of which change in the same direction or vice
279 versa. The formula for calculating correlation can be seen in Eq. 7.

280 $r_{xy} = \dfrac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{n\Sigma X^2 - (\Sigma X)^2\}\{n\Sigma Y^2 - (\Sigma Y)^2\}}}$ (7)

281 where r_{xy} is the correlation value, X is the variable X, and Y is the variable Y.

282

283 5. Forecast and Classification
284 Supervised classification is used to classify image data and form a thematic map of Land Used
285 Land Cover of the study area. Contour analysis of the study area is the goal behind the
286 classification of the Shuttle Radar Topographic Mission (SRTM) Digital Elevation Model
287 (DEM) image data. k-NN is the classification method for a set of data based on pre-existing
288 learning data. The classification process is done by finding the closest point from the old point of
289 "a" to a new point of "a" (nearest neighbor). The closest point search technique is performed by
290 using the Euclidean distance formula, as shown in Eq. 8. To use the k-NN algorithm, it is
291 necessary to determine the value of k, where k is the number of nearest neighbors used to
292 classify the new data.

293 $d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$ (8)

294 where n is the number of data.

295

296 6. Spatial Interpolation
297 Interpolation is the method used to predict a value at locations where data are not available.
298 Thus, interpolation is used to predict values of the surrounding points outside the sample point.
299 Meanwhile, spatial interpolation shows the process of value estimation in the surrounding areas
300 outside the sample point to determine the distribution of values in the area being mapped.

301

302 Inverse Weighted Distance (IDW) is used to interpolate prediction results. In IDW, the estimated
303 value of a at location of x shows the average of the weighted closest observations.

304 $\hat{a}(x) = \dfrac{\Sigma_i^n h_i j_i}{\Sigma_i^n h_i}$ (9)

305 where $h_i = |x - x_i|^{-\beta}$, $\beta \geq 0$, $|...|$ corresponds to the Euclidean distance and β determines the
306 extent to which a point is preferred over other points. If the point x coincides with the
307 observation location or sample point ($x = x_i$), the value of the sample point x is returned to avoid
308 infinity weight.

309

310 7. Testing the accuracy of prediction results.
311 To test the accuracy of prediction results, the MSE, RMSE, and MAE are used. The analysis of
312 the level of vulnerability will be continued when the result of the prediction testing is accurate.

313   On the contrary, if the result of the analysis is not accurate, the analysis of the level of
314   vulnerability will not be continued. It will be analyzed again using machine learning. The
315   calculations of MSE, RMSE, and MAE, which is close to zero, provide good accuracies.
316

317   Therefore, the proposed algorithm for detecting tsunami risk areas can be described in Algorithm
318   1.
319   **Algorithm 1: Algorithm for detecting tsunami risk areas**
320   Input: Landsat 8 image data (X) and the number of nearest neighbors (k).
321   Step 1: Extract the data into 11 bands.
322   Step 2: Atmospheric correction, using DOS1 (Eq. 1).
323   Step 3: Pre-processing computation of vegetation indices, using NDVI (Eq. 2), NDWI (Eq. 3),
324   MNDWI (Eq. 4), SAVI (Eq. 5), and NDBI (Eq. 6).
325   Step 4: Compute the correlation test, using Pearson correlation (Eq. 7).
326   Step 5: Forecast and classify the data, using k-NN algorithm.
327   Step 7: Spatial interpolation, using IDW (Eq. 9).
328   Step 8: Map the tsunami risk areas.
329   Output: Tsunami risk local map.
330

331   Algorithm 1 is the algorithm to detect and determine area affected by tsunami, based on the
332   vegetation indices, using k-NN algorithm. Each vegetation index produces one tsunami risk local
333   map, so that Algorithm 1 generates five tsunami risk local maps, i.e., NDVI map, NDWI map,
334   MNDWI map, SAVI map, and NDBI map. Therefore, in this paper, an algorithm to obtain a
335   model that combines all vegetation indices is also proposed. The suitable vegetation indices are
336   selected based on the local map results from Algorithm 1. This model calculates the tsunami risk
337   value for each region and their level risks. Multiple linear regression is used to obtain the model.
338   The algorithm for determining the tsunami risk value and level risk is described in Algorithm 2.
339   **Algorithm 2: Algorithm for determining tsunami risk values**
340   Input: independent variable, y, (i.e., level risk value) and dependent variable, x, (i.e., vegetation
341   indices and other parameters)
342   Step 1: Select the dependent variable of vegetation indices based on the results of Algorithm 1.
343   Step 2: Apply a multiple linear regression, $y = a_0 + a_1 x_1 + a_2 x_2 + ... + a_n x_n$ for the training
344   data.
345   Step 3: Calculate the risk value for the testing data.
346   Step 4: Categorize the risk value into level risk.
347   Output: risk value and level risk
348

349   ## Results and Discussions
350   **Data Generation**
351   The data used were Landsat 8 OLI image data, which contained 11 bands as listed in Table 1. All
352   11 bands were processed using QGIS according to the required indices; NDVI, NDBI, MNDWI,

353 NDWI and SAVI. Each index was calculated using Eq. (2)-(6) to see monthly index results. The
354 data taken for processing were the average result of each index. The data used for this study were
355 yearly data for two different months, August and November, from 47 villages in Kulon Progo
356 Regency. For processing, the data for August and November from 2014 to 2020 were combined.
357 The data were arranged sequentially every year starting from August to November, followed by
358 the index values and classification of each index. In the binary field, each index was adjusted to
359 the potential risk posed and divided into two groups, TRUE (affected by tsunami) and FALSE
360 (unaffected by tsunami). The total binary result of each index was calculated as a sum that
361 indicates TRUE. Since there were five vegetation indices, then each region had a minimum of
362 zero TRUE values and a maximum of five TRUE values. A region was classified into three risk
363 levels, i.e., low risk, medium risk, and high risk. Thus, in this research, it is determined and
364 simulated that if the number of TRUE ≤ 1, a region has a low potential/impact (low risk), if the
365 number of TRUE $= 2$, a region has a moderate potential/impact (medium risk), and if the number
366 of TRUE ≥ 3, a region has a high potential/impact of tsunami (high risk). Figure 3 shows the
367 calculation results of each index and their resulting potentials.

368

369 Figure 3. Calculation Results of Each Index

370

371 **Correlation between Variables**
372 Based on the results of the Pearson correlation analysis, it showed that the correlation between
373 the vegetation indices of the three tsunami impact risk classes had a positive or negative
374 correlation value. As presented in Figure 4, it can be seen that the lowest correlation with a
375 negative value was the correlation between the MNDWI and NDVI indices, with a correlation
376 value of -0.761. The scatter diagram showed the increase in the MNDWI index value which was
377 not in line with the increase in the NDVI value for the potential tsunami risk. The highest
378 correlation that had a positive value was the correlation between the SAVI and NDVI indices of
379 0.767. The distribution pattern of the data contained in the scatter diagram was closer to a
380 straight line and the risk of tsunami impact increased in line with the increase in the value of the
381 vegetation indices. The positive correlation between these two indices could be seen as the
382 strongest relationship in the medium risk of tsunami impact. A positive correlation was indicated
383 by the distribution pattern of the data pair points that moved closer to a straight line, which
384 showed a close relationship between the risk of the tsunami impact and the index. This
385 relationship might also be referred to as a unidirectional relationship. The visualization of the
386 data displayed was not only in the form of closeness value between variables but also in the form
387 of a bar plot presenting the distribution of data from 2014 to 2020. Data illustrated that moderate
388 tsunami risks are more dominant than low tsunami risks.

389

390 Figure 4. Correlation between Variables

391

392 **Prediction and Classification (k-NN)**

393   Prediction and classification were done using the k-NN algorithm, where the number of k was
394   determined by classification using k-NN with training data 70% and testing data 30% of the total
395   data. To compare the performance of k, the calculation of the accuracy and value of the Kappa
396   coefficient was carried out starting from $k = 1$ up to $k = 15$. Based on the obtained results in
397   Figure 5, the highest accuracy and value of Kappa was in k-NN with $k = 1$ and 2. Therefore, in
398   this study, predictions were made using $k = 1$.
399
400   Figure 5. Accuracy and Value of Cohen's Kappa
401
402   The testing of performance modeling is carried out by calculating Mean Squared Error (MSE),
403   Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Cohen's Kappa. MSE is
404   the sum of the squared error or the difference between the actual value and the predicted value,
405   where $MSE = \sum \frac{(Y' - Y)^2}{n}$, with Y' is the predicted value, Y is the actual value, and n is the total
406   data. MAE shows the average error value from the actual value compared to the predicted value,
407   where $MAE = \sum \frac{|Y' - Y|}{n}$. Usually, it is used for measuring error prediction in time series analysis.
408   RMSE is the square root of MSE, where $RMSE = \sqrt{\sum \frac{(Y' - Y)^2}{n}}$. Cohen's Kappa is used to measure
409   the degree of agreement of the two raters in classifying objects into groups and in measuring the
410   agreement between the new methods and the existing methods, where $Cohen's\ Kappa =$
411    $\frac{Pr(a) - Pr(e)}{1 - Pr(e)}$, with $Pr(a)$ is the percentage of consistent number of measurements between raters
412   and $Pr(e)$ is the percentage of change in measurement between raters.
413
414   After the predictions are done, the MSE, RMSE, and MAE are calculated to see the prediction
415   performance. Table 2 shows the test of error and the resulting average from data using the
416   prediction results of each index with k-NN, ANN, SVM, and CART algorithms.
417
418   Table 2. Error Value of Each Index with Prediction Results using k-NN, ANN, SVM, and CART
419   algorithms.
420
421   The values of MSE, RMSE, and MAE for k-NN, ANN, SVM, and Cart of each index can be
422   shown in Figures 6-8, respectively. As can be seen from these figures, the k-NN algorithm
423   obtains the best performance for all indices, compared with ANN, SVM, and Cart algorithms.
424
425   Figure 6. MSE values for k-NN, ANN, SVM, and Cart
426
427   Figure 7. RMSE values for k-NN, ANN, SVM, and Cart
428
429   Figure 8. MAE values for k-NN, ANN, SVM, and Cart
430

431   Furthermore, the averages of each index using all algorithms are compared to the actual data.
432   The results can be seen in Table 3. From this table, k-NN obtains the closest average value for
433   each index.
434
435   Table 3. The Average of Each Index using k-NN, ANN, SVM, and Cart
436
437   **The Map of Tsunami Risk Areas**
438   Figure 9 is the map showing the distribution of areas that have the potential for a tsunami using
439   the IDW interpolation results. The higher the level of vulnerability of the area affected by the
440   tsunami, the redder the area will be and the lower the level of vulnerability, the bluer the color
441   will be. An area will be classified as a high vulnerability area when it has a high wettability value
442   and a few vegetation. Figure 9 shows the distribution map of tsunami risk areas based on NDWI
443   index 2021. When an area is dominated by water or has a high level of wetness, it will have a
444   high level of tsunami risk so that the area has a reddish color. On the contrary, an area with a low
445   level of tsunami risk has a bluish color.
446
447   Figure 9. The Distribution Map of Tsunami Risk Areas Based for NDWI Index 2021
448
449   The map of the distribution of potential areas in Figure 10 provides tsunami risk areas based on
450   the NDVI index. The higher the NDVI value, the higher possibility of an area having high
451   vegetation. An area that has a low risk of tsunami, its color will be bluish because it has a lot of
452   vegetation. For an area with a high risk of tsunami, its color will be close to red because the area
453   has a low NDVI value and thus it shows that the area has only a few vegetation.
454
455   Figure 10. The Distribution Map of Tsunami Risk Areas Based for NDVI 2021 Index
456
457   The SAVI index presents areas that are adjusted according to their type of soil or the shape of the
458   areas. Higher SAVI value indicates that the area is in the form of forested vegetation, while
459   lower value means the area has many bodies of water such as rivers. Figure 11 shows that the
460   area that is colored redder, which means that the area has a low SAVI value and has a high level
461   of tsunami vulnerability because it has many areas of water, asphalt, paving, etc., meanwhile the
462   blue area in the figure is an area that has a low level of tsunami vulnerability. Figure 11 also
463   shows the distribution map of tsunami risk areas based on SAVI index 2021.
464
465   Figure 11. The Distribution Map of Tsunami Risk Areas for SAVI Index 2021
466
467   Figure 12 is a slope map that describes the slope angle in the study area of Kulon Progo
468   Regency. The slope map is represented by the percentage of the slope angle, i.e., 0-8%, 8-15%,
469   15-25%, 25-40%, and >40%. This research analyzes the slope as one of the reference data for
470   making maps of tsunami-prone areas. The results of the spatial analysis suggest that most of the

471   study areas have undulating slopes, some are moderately sloping and some are hilly. The
472   quantitative value of the slope map is used as an indicator to determine areas having a high risk
473   of being prone to tsunami.
474
475   Figure 12. The Slope Map
476
477   Furthermore, since there are only three categories of risk level, they are high, medium, and low
478   risk and no value that indicates the level of tsunami risk, then a new formula is also proposed and
479   discussed to determine the tsunami risk value. In this formula, vegetation indices, slope, and
480   distance are used to obtain the risk value, where distance is the shortest distance of a region with
481   the shoreline. The value is on interval [0,1]. The higher the risk value, RV, the higher the area is
482   at risk of a tsunami.
483
484   In this experiment, the data of vegetation indices, slope, and distance from 47 villages of 2014
485   until 2019 were used as the training data. NDBI and MNDWI were not used in the formula since
486   it had no effect in determining the risk area, based on the experiment using k-NN. Since a region
487   with low slope value, low NDVI value, or low distance had a high tsunami risk level, then they
488   were applied as denominators in the model. Therefore, according to slope, NDWI, NDVI, and
489   SAVI values of the training data as the independent variables and level risk values of the training
490   data as the dependent variables, Algorithm 2 can be used to obtain the risk values of the testing
491   data. With multiple linear regression model using Excel, the tsunami risk level, RV, was modeled
492   as in Eq. (10) as below,

493   $$RV = \frac{0.95}{slope} + 0.044 * NDWI - \frac{0.007}{NDVI} + 0.3 * SAVI + \frac{0.59}{distance} \qquad (10)$$

494   with R^2 of 0.921, meaning 92.1% of the variation in the dependent variable (i.e., risk value)
495   could be explained by the independent variables (i.e., slope, NDWI, NDVI, SAVI, and distance).
496
497   To test the model, data from 2020 were used as the testing data. The MSE of the testing data was
498   0.047. Besides that, the risk level was categorized into low, medium, and high risk. Low risk was
499   in interval [0, 0.4) of risk value, medium risk was in interval [0.4, 0.8) of risk value, and high
500   risk was in interval [0.8, 1] of risk value. The model obtained an accuracy rate of 93.62%. Using
501   the model from Eq. (10), the prediction of 2021 tsunami risk values for 47 villages is listed in
502   Table 4.
503
504   Table 4. Prediction of 2021 Tsunami Risk Values
505

506   ## Conclusions
507   To sum up, the paper proposed a new algorithm to predict tsunami risk areas based on spatial
508   prediction of vegetation indices. Atmospheric correction using the DOS1 algorithm is used for
509   image correction since the field data parameters are not known. k-NN is used to classify and
510   predict tsunami-affected areas from vegetation indices data that has spatial and temporal

511 resolutions. Experimental results and comparisons demonstrate the effectiveness of the proposed
512 algorithm to detect the tsunami risk area. The result with *k*-NN algorithm gives MSEs of 0.0002
513 for MNDWI, 0.0002 for SAVI, 0.0006 for NDVI, 0.0003 for NDWI, and 0.0003 for NDBI.
514 Moreover, based on the tsunami risk area of each vegetation index, a prediction model also has
515 been proposed using NDVI, NDWI, SAVI, slope, and distance to obtain the tsunami risk value
516 of an area, by using multiple linear regression with R^2 of 0.921. The accuracy rate from
517 categorizing the risk value into level risk is about 93.62%. For future work, besides vegetation
518 indices, slope, and distance, other parameters can also be used to find the tsunami risk area with
519 a more accurate risk value, such as elevation, sea level, etc. Parameters that are not related to the
520 determination of a potentially tsunami-prone area may also be found by certain methods.
521 Another future work is to compare the Euclidean distance used in *k*-NN algorithm with other
522 numerical similarity metric, such as cosine distance, Jaccard distance, etc.
523

524 References

525 Acharya, T. D., Subedi, A., & Lee, D. H. (2018). Evaluation of water indices for surface water
526   extraction in a landsat 8 scene of Nepal. *Sensors (Switzerland)*, *18*(8), 1–15.
527   https://doi.org/10.3390/s18082580
528 Amri, M. R., Yulianti, G., Yunus, R., Wiguna, S., W. Adi, A., Ichwana, A. N., & Randongkir,
529   Roling Evans Septian, R. T. (2018). *Indonesia's disaster risk (RBI)* (Vol. 9, Issue 3).
530   National Development Planning Agency (Badan Nasional Penanggulangan Bencana).
531 Bovolo, F., & Bruzzone, L. (2007). A split-based approach to unsupervised change detection in
532   large-size SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(6),
533   1658–1670. https://doi.org/10.1117/12.690710
534 Brunner, D., Lemoine, G., & Bruzzone, L. (2010). Earthquake damage assessment of buildings
535   using VHR optical and SAR imagery. *IEEE Transactions on Geoscience and Remote*
536   *Sensing*, *48*(5), 2403–2420. https://doi.org/10.1109/TGRS.2009.2038274
537 Corner, B. R., Narayanan, R. M., & Reichenbach, S. E. (2000). Noise reduction in remote
538   sensing imagery using data masking and principal component analysis. *Applications of*
539   *Digital Image Processing XXIII*, *4115*(4), 1–11. https://doi.org/10.1117/12.411533
540 Gao, B.-C. (1996). NDWI - A normalized difference water index for remote sensing of
541   vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266.
542   https://doi.org/https://doi.org/10.1016/S0034-4257(96)00067-3
543 Ghebrezgabher, M. G., Yang, T., Yang, X., & Eyassu Sereke, T. (2020). Assessment of NDVI
544   variations in responses to climate change in the Horn of Africa. *Egyptian Journal of Remote*
545   *Sensing and Space Science*, *23*(3), 249–261. https://doi.org/10.1016/j.ejrs.2020.08.003
546 Regulation of the Special Region of Yogyakarta no. 5 of 2019 about The spatial plan of the
547   Special Region of Yogyakarta for 2019-2039, Pub. L. No. 5, 1 (2019).
548 Havivi, S., Schvartzman, I., Maman, S., Rotman, S. R., & Blumberg, D. G. (2018). Combining
549   TerraSAR-X and Landsat images for emergency response in urban environments. *Remote*
550   *Sensing*, *10*(5), 1–13. https://doi.org/10.3390/rs10050802
551 Holzman, M. E., Rivas, R., & Bayala, M. (2014). Subsurface soil moisture estimation by VI-LST
552   method. *IEEE Geoscience and Remote Sensing Letters*, *11*(11), 1951–1955.
553   https://doi.org/10.1109/LGRS.2014.2314617
554 Hu, Y., Ban, Y., Zhang, Q., Zhang, X., Liu, J., & Zhuang, D. (2008). Spatial-temporal pattern of

555   GIMMS NDVI and its dynamics in Mongolian Plateau. *2008 International Workshop on*
556   *Earth Observation and Remote Sensing Applications, EORSA*, 1–6.
557   https://doi.org/10.1109/EORSA.2008.4620310

558 Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*,
559   *25*, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

560 Ilham, A., & Marzuki, M. I. (2017). Machine learning-based mangrove land classification on
561   Worldview-2 satellite image in Nusa Lembongan island. *International Journal of Remote*
562   *Sensing and Earth Sciences*, *14*(2), 159–166.
563   https://doi.org/10.30536/j.ijreses.2017.v14.a2820

564 Karnieli, A., Agam, N., Pinker, R. T., Anderson, M., Imhoff, M. L., Gutman, G. G., Panov, N.,
565   & Goldberg, A. (2010). Use of NDVI and land surface temperature for drought assessment:
566   Merits and limitations. *Journal of Climate*, *23*(3), 618–633.
567   https://doi.org/10.1175/2009JCLI2900.1

568 Koshimura, S., Moya, L., Mas, E., & Bai, Y. (2020). Tsunami damage detection with remote
569   sensing: A review. *Geosciences*, *10*(5), 1–28. https://doi.org/10.3390/geosciences10050177

570 Liu, T. (2010). Application of Markov chains to analyze and predict the time series. *Modern*
571   *Applied Science*, *4*(5), 508–511. https://doi.org/10.5539/mas.v4n5p162

572 Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote
573   sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and*
574   *Remote Sensing*, *152*, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

575 Mallick, K., Bhattacharya, B. K., & Patel, N. K. (2009). Estimating volumetric surface moisture
576   content for cropped soils using a soil wetness index based on surface temperature and
577   NDVI. *Agricultural and Forest Meteorology*, *149*(8), 1327–1342.
578   https://doi.org/10.1016/j.agrformet.2009.03.004

579 Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning
580   classification in remote sensing: An applied review. *International Journal of Remote*
581   *Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

582 Mehrotra, A., Singh, K. K., Nigam, M. J., & Pal, K. (2015). Detection of tsunami-induced
583   changes using generalized improved fuzzy radial basis function neural network. *Natural*
584   *Hazards*, *77*(1), 367–381. https://doi.org/10.1007/s11069-015-1595-z

585 Min, C. K., Bahar, A. M. A., & Udin, W. S. (2016). Landslide assessment using Normalized
586   Difference Vegetation Index (NDVI). *J. Trop. Resour. Sustain. Sci*, *4*, 98–104.

587 Regulation of the Ministry of Health Republic of Indonesia no. 36 of 2014 about Assessment of
588   post-disaster damage, loss and health resource needs, Pub. L. No. 36 (2014). internal-
589   pdf://0183006508/files21279PMK No 36 Th 2014 ttg Penilaian Keru.pdf

590 Mustaqim, A. (2019). *40 villages in DIY are prone to tsunami disasters*.
591   https://www.medcom.id/nasional/daerah/nN9w3PEk-40-desa-di-diy-rawan-bencana-
592   tsunami

593 National Disaster Management Agency, B. (2012). Towards a resilient Indonesia to face the
594   tsunami. In *Masterplan Pengurangan Risiko Bencana Tsunami*.
595   https://bnpb.go.id/uploads/migration/pubs/578.pdf

596 Prasetyo, S. Y. J., Hartomo, K. D., Paseleng, M. C., Chandra, D. W., & Winarko, E. (2020).
597   Satellite imagery and machine learning for aridity disaster classification using vegetation
598   indices. *Bulletin of Electrical Engineering and Informatics*, *9*(3), 1149–1158.
599   https://doi.org/10.11591/eei.v9i3.1916

600 Regional Disaster Management Agency, B. (2019). Implementation of local government affairs.

601   In *Accountability report*.

602 Rendana, M., Rahim, S. A., Idris, W. M. R., Lihan, T., & Rahman, Z. A. (2016). Mapping
603   nutrient status in oil palm plantation using geographic information system. *Asian Journal of*
604   *Agricultural Research*, *10*(3–4), 144–153. https://doi.org/10.3923/ajar.2016.144.153

605 Singh, K. K., Nigam, M. J., & Pal, K. (2014). Detection of 2011 Tohoku tsunami inundated areas
606   in Ishinomaki city using generalized improved fuzzy Kohonen clustering network.
607   *European Journal of Remote Sensing*, *47*(1), 461–475.
608   https://doi.org/10.5721/EuJRS20144726

609 Stepchenko, A., & Chizhov, J. (2016). Applying Markov Chains for NDVI time series
610   forecasting of Latvian Regions. *Information Technology and Management Science*, *18*(1),
611   57–61. https://doi.org/10.1515/itms-2015-0009

612 U.S. Geological Survey. (2021). *What are the best Landsat spectral bands for use in my*
613   *research?* https://www.usgs.gov/faqs/what-are-best-landsat-spectral-bands-use-my-
614   research?qt-news_science_products=0#qt-news_science_products

615 Verstappen, H. T. (2010). Indonesian landforms and plate tectonics. *Indonesian Journal on*
616   *Geoscience*, *5*(3), 197–207. https://doi.org/10.17014/ijog.v5i3.103

617 Volpi, M., Tuia, D., Bovolo, F., Kanevski, M., & Bruzzone, L. (2012). Supervised change
618   detection in VHR images using contextual information and support vector machines.
619   *International Journal of Applied Earth Observation and Geoinformation*, *20*(1), 77–85.
620   https://doi.org/10.1016/j.jag.2011.10.013

621 Wang, J., Yang, X., Yang, X., Jia, L., & Fang, S. (2020). Unsupervised change detection
622   between SAR images based on hypergraphs. *ISPRS Journal of Photogrammetry and*
623   *Remote Sensing*, *164*, 61–72. https://doi.org/10.1016/j.isprsjprs.2020.04.007

624 Xie, Y., Zhao, X., Li, L., & Wang, H. (2010). Calculating NDVI for landsat7-ETM data after
625   atmospheric correction using 6S model: A case study in Zhangye City, China. *2010 18th*
626   *International Conference on Geoinformatics, Geoinformatics*, 1–4.
627   https://doi.org/10.1109/GEOINFORMATICS.2010.5567553

628 Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open
629   water features in remotely sensed imagery. *International Journal of Remote Sensing*,
630   *27*(14), 3025–3033. https://doi.org/10.1080/01431160600589179

631 Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of
632   developments and applications. *Journal of Sensors*, *2017*, 1–18.
633   https://doi.org/10.1155/2017/1353691

634 Yang, X., & Wen, X. (2011). Post classification comparison change detection of Guangzhou
635   Metropolis, China. *Key Engineering Materials*, *467–469*, 19–22.
636   https://doi.org/10.4028/www.scientific.net/KEM.467-469.19

637 Yonezawa, S. I. and C. (2015). Observation of tsunami damage to coastal forest using middle
638   spatial resolution satellite data. *Journal of Integrated Field Science*, *12*, 1–7.

639 Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically
640   mapping urban areas from TM imagery. *International Journal of Remote Sensing*, *24*(3),
641   583–594. https://doi.org/10.1080/01431160304987

642 Zhang, Z., He, G., & Wang, X. (2010). A practical DOS model-based atmospheric correction
643   algorithm. *International Journal of Remote Sensing*, *31*(11), 2837–2852.
644   https://doi.org/10.1080/01431160903124682

645 Zhao, L., Zhang, P., Ma, X., & Pan, Z. (2017). Land cover information extraction based on daily
646   NDVI time series and multiclassifier combination. *Mathematical Problems in Engineering*,

647   *2017*, 1–13. https://doi.org/10.1155/2017/6824051
648

Landsat 8 Operational Land Image (OLI) and Thermal Infrared Sensor (TIRS) (U.S. Geological Survey, 2021)

1 Table 1. Landsat 8 Operational Land Image (OLI) and Thermal Infrared Sensor (TIRS) (U.S.
2 Geological Survey, 2021)

Band Name	Band	Bandwidth	Resolution
Band 1	Coastal aerosol	0.43 – 0.45	30
Band 2	Blue	0.45 – 0.51	30
Band 3	Green	0.53 – 0.59	30
Band 4	Red	0.64 – 0.67	30
Band 5	NIR (Near Infrared)	0.85 – 0.88	30
Band 6	SWIR (Short-wave Infrared) 1	1.57 – 1.65	30
Band 7	SWIR (Short-wave Infrared) 2	2.11 – 2.29	30
Band 8	Panchromatic	0.50 – 0.68	15
Band 9	Cirrus	1.36 – 1.68	30
Band 10	TIRS 1	10.60 – 10.19	100
Band 11	TIRS 2	11.50 – 12.51	100

MSE, RMSE, and MAE Values of Each Index with Prediction Results using k-NN, ANN, SVM, and Cart Algorithms

1   Table 2. MSE, RMSE, and MAE Values of Each Index with Prediction Results using *k*-NN, ANN, SVM, and Cart Algorithms

Algorithm / Index	MSE				RMSE				MAE			
	k-NN	ANN	SVM	Cart	*k*-NN	ANN	SVM	Cart	*k*-NN	ANN	SVM	Cart
MNDWI	0.0015	0.0015	0.0013	0.0015	0.0147	0.0387	0.0363	0.0388	0.0070	0.0308	0.0256	0.0309
SAVI	0.0002	0.0006	0.0007	0.0007	0.0130	0.0244	0.0266	0.0266	0.0063	0.0199	0.0191	0.0212
NDVI	0.0006	0.0015	0.0015	0.0018	0.0244	0.0385	0.0384	0.0423	0.0106	0.0302	0.0260	0.0335
NDWI	0.0003	0.0007	0.0006	0.0008	0.0180	0.0262	0.0253	0.0286	0.0086	0.0191	0.0166	0.0218
NDBI	0.0003	0.0008	0.0007	0.0008	0.0184	0.0281	0.0258	0.0286	0.0086	0.0214	0.0174	0.0218

The Average of Each Index using k-NN, ANN, SVM, and Cart

1   Table 3. The Average of Each Index using *k*-NN, ANN, SVM, and Cart

	MNDWI	SAVI	NDVI	NDWI	NDBI
k-NN	-0.3411	0.2928	0.4873	0.2058	-0.2055
ANN	-0.3308	0.2909	0.4841	0.2024	-0.2018
SVM	-0.3305	0.2867	0.4809	0.2000	-0.1991
Cart	-0.3305	0.2907	0.4838	0.2056	-0.2056
Actual	-0.3406	0.2962	0.4941	0.2055	-0.2055

Prediction of 2021 Tsunami Risk Values

1   Table 4. Prediction of 2021 Tsunami Risk Values

Village	Risk Level	Category
Banaran	0.943	High risk
Bendungan	0.486	Medium risk
Bojong	0.626	Medium risk
Brosot	0.401	Medium risk
Bugel	0.959	High risk
Bumirejo	0.388	Low risk
Cerme	0.382	Low risk
Demen	0.541	Medium risk
Depok	0.576	Medium risk
Garongan	0.953	High risk
Giripeni	0.297	Low risk
Glagah	0.965	High risk
Gotakan	0.350	Low risk
Gulurejo	0.282	Low risk
Jangkaran	0.897	High risk
Janten	0.546	Medium risk
Jatirejo	0.341	Low risk
Kalidengen	0.671	Medium risk
Kaligintung	0.353	Low risk
Kanoman	0.583	Medium risk
Karangsewu	0.937	High risk
Karangwuluh	0.625	Medium risk
Karangwuni	0.967	High risk
Kebonrejo	0.675	Medium risk
Kedundang	0.590	Medium risk
Kranggan	0.784	Medium risk
Krembangan	0.367	Low risk
Kulur	0.354	Low risk
Kulwaru	0.640	Medium risk
Ngentakrejo	0.555	Medium risk
Ngestiharjo	0.558	Medium risk
Nomporejo	0.663	Medium risk
Palihan	0.965	High risk
Pandowan	0.460	Medium risk
Panjatan	0.927	High risk
Pleret	0.918	High risk
Plumbon	0.768	Medium risk
Sidorejo	0.309	Low risk

Sindutan	0.955	High risk
Sogan	0.752	Medium risk
Tayuban	0.531	Medium risk
Temon Kulon	0.664	Medium risk
Temon Wetan	0.410	Medium risk
Tirtorahayu	0.607	Medium risk
Triharjo	0.495	Medium risk
Wahyuharjo	0.484	Medium risk
Wates	0.860	High risk

Figure 1

The Visualization of Kulon Progo Regency Map
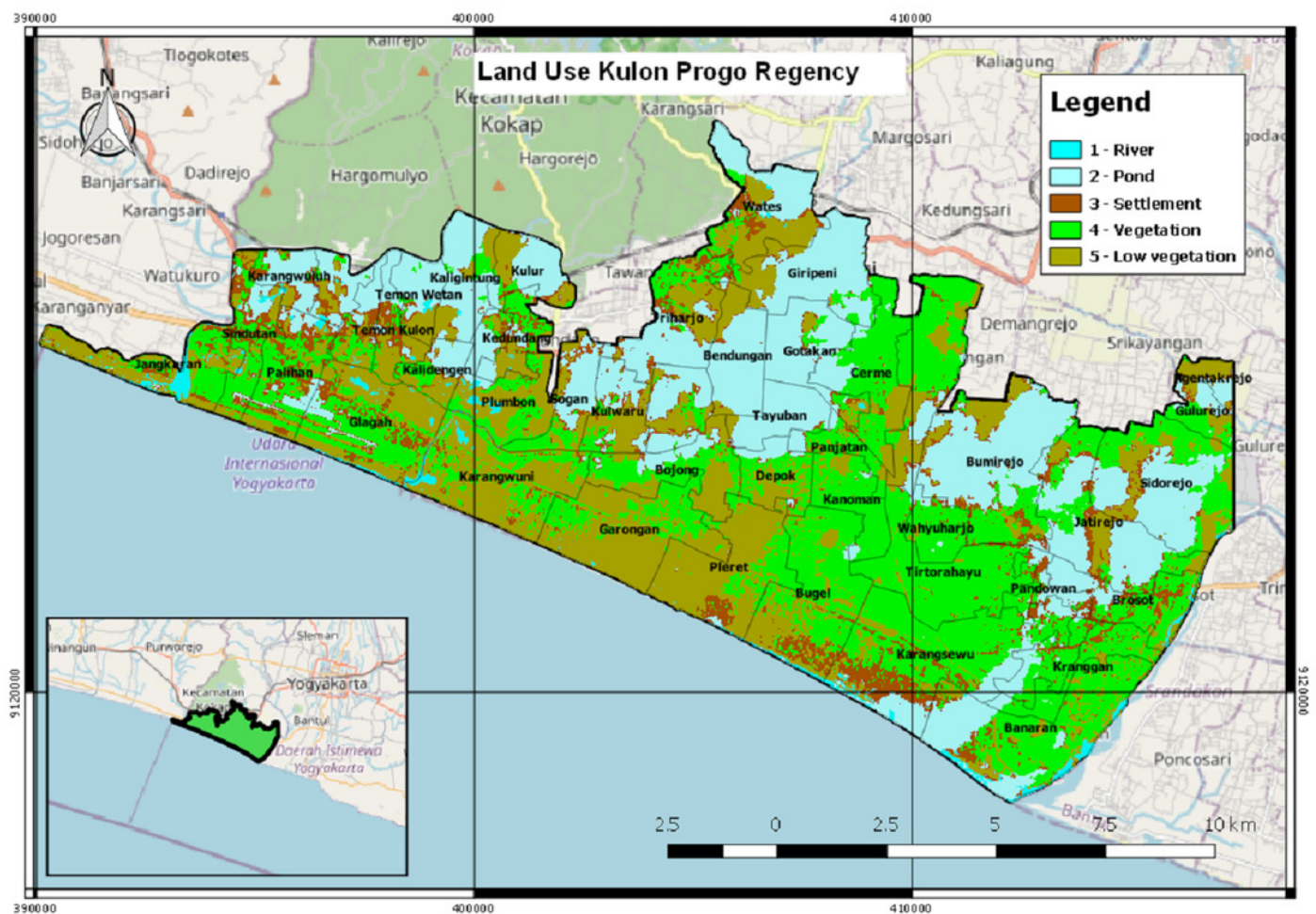
Figure 2

Flowchart of The Research Method
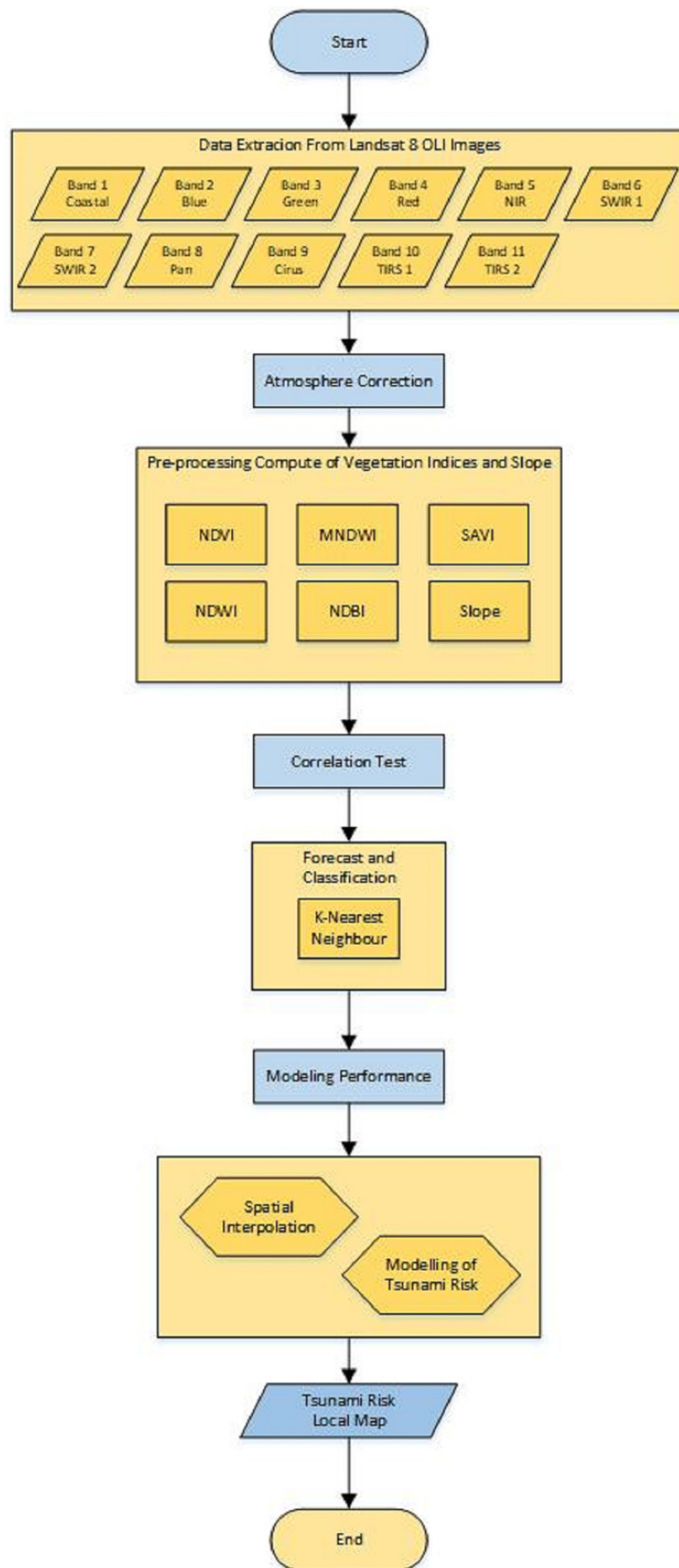
Figure 3

Calculation Results of Each Index

Code	Village	Subdistrict	Month	Year	MNDWI	Class_MNDWI	Biner_NDWI	NDBI	Class_NDBI	Biner_NDBI	NDVI	Class_NDVI	Biner_NDVI	SAVI	Class_SAVI	Biner_SAVI
1	Tirtorahayu	Galur	Ags	14	-0.113683636	Negative	FALSE	-0.267873886	Low Building	FALSE	0.423695406	High Vegetation	FALSE	0.250334571	Water Body	TRUE
2	Pandowan	Galur	Ags	14	-0.133624775	Negative	FALSE	-0.232603365	Low Building	FALSE	0.39977869	High Vegetation	FALSE	0.235702273	Water Body	TRUE
3	Brosot	Galur	Ags	14	-0.11739655	Negative	FALSE	-0.234611121	Low Building	FALSE	0.394055183	High Vegetation	FALSE	0.235803603	Water Body	TRUE
4	Karangsewu	Galur	Ags	14	-0.180442941	Negative	FALSE	-0.189978264	Low Building	FALSE	0.403557498	High Vegetation	FALSE	0.242450261	Water Body	TRUE
5	Nomporejo	Galur	Ags	14	-0.189308851	Negative	FALSE	-0.225053675	Low Building	FALSE	0.445765956	High Vegetation	FALSE	0.270885935	Water Body	TRUE
6	Kranggan	Galur	Ags	14	-0.127726195	Negative	FALSE	-0.238212577	Low Building	FALSE	0.397963424	High Vegetation	FALSE	0.239913419	Water Body	TRUE
7	Banaran	Galur	Ags	14	-0.158888931	Negative	FALSE	-0.173422056	Low Building	FALSE	0.355732602	High Vegetation	FALSE	0.22040389	Water Body	TRUE
8	Ngentakrejo	Lendah	Ags	14	-0.228313264	Negative	FALSE	-0.104133792	Low Building	FALSE	0.352866227	High Vegetation	FALSE	0.221476572	Water Body	TRUE
9	Bumirejo	Lendah	Ags	14	-0.181545655	Negative	FALSE	-0.245630879	Low Building	FALSE	0.477728224	High Vegetation	FALSE	0.292518802	Water Body	TRUE
10	Gulurejo	Lendah	Ags	14	-0.22599853	Negative	FALSE	-0.175048159	Low Building	FALSE	0.436135425	High Vegetation	FALSE	0.278032241	Water Body	TRUE
11	Sidorejo	Lendah	Ags	14	-0.210477713	Negative	FALSE	-0.251611881	Low Building	FALSE	0.512092307	High Vegetation	FALSE	0.324956368	Water Body	TRUE
12	Jatirejo	Lendah	Ags	14	-0.177740322	Negative	FALSE	-0.242566069	Low Building	FALSE	0.46990353	High Vegetation	FALSE	0.285023188	Water Body	TRUE
13	Wahyuharjo	Lendah	Ags	14	-0.128595963	Negative	FALSE	-0.236725	Low Building	FALSE	0.405894962	High Vegetation	FALSE	0.237501492	Water Body	TRUE
14	Gulurejo	Lendah	Ags	14	-0.210490726	Negative	FALSE	-0.23578095	Low Building	FALSE	0.490729429	High Vegetation	FALSE	0.31586667	Water Body	TRUE
15	Sidorejo	Lendah	Ags	14	-0.182961158	Negative	FALSE	-0.257163309	Low Building	FALSE	0.492976367	High Vegetation	FALSE	0.313620342	Water Body	TRUE
16	Sendadangsari	Lendah	Ags	14	-0.261867434	Negative	FALSE	-0.177387938	Low Building	FALSE	0.456650198	High Vegetation	FALSE	0.301415563	Water Body	TRUE
17	Ngentakrejo	Lendah	Ags	14	-0.222437853	Negative	FALSE	-0.230771433	Low Building	FALSE	0.498312137	High Vegetation	FALSE	0.317370885	Water Body	TRUE
18	Krembangan	Panjatan	Ags	14	-0.215802013	Negative	FALSE	-0.252987413	Low Building	FALSE	0.519000752	High Vegetation	FALSE	0.325243217	Water Body	TRUE
19	Cerme	Panjatan	Ags	14	-0.224602919	Negative	FALSE	-0.272899267	Low Building	FALSE	0.549492544	High Vegetation	FALSE	0.341841331	Water Body	TRUE
20	Gotakan	Panjatan	Ags	14	-0.210954366	Negative	FALSE	-0.25279081	Low Building	FALSE	0.518485258	High Vegetation	FALSE	0.318002723	Water Body	TRUE

Figure 4

Correlation between Variables
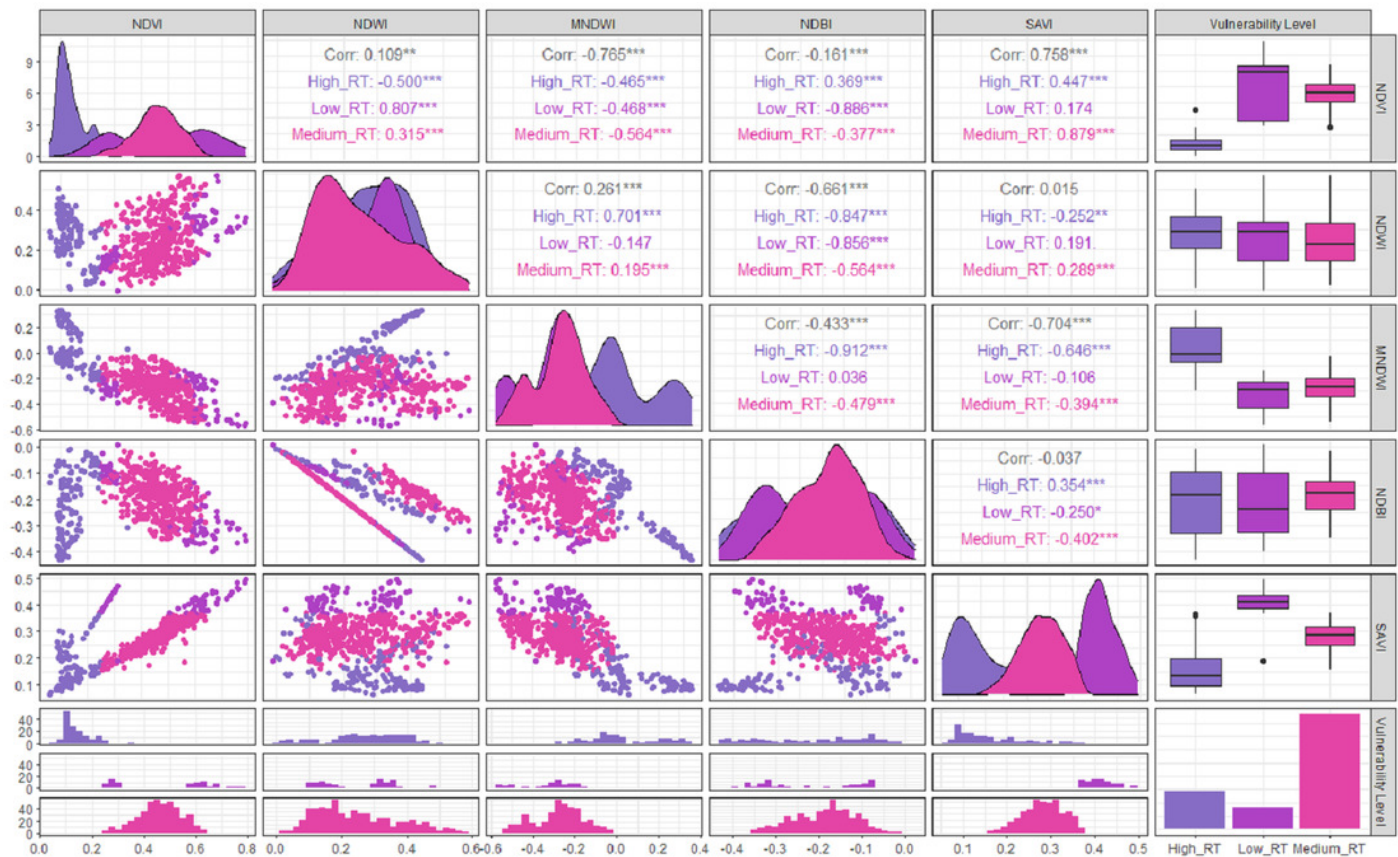
Figure 5

Accuracy and Value of Cohen's Kappa
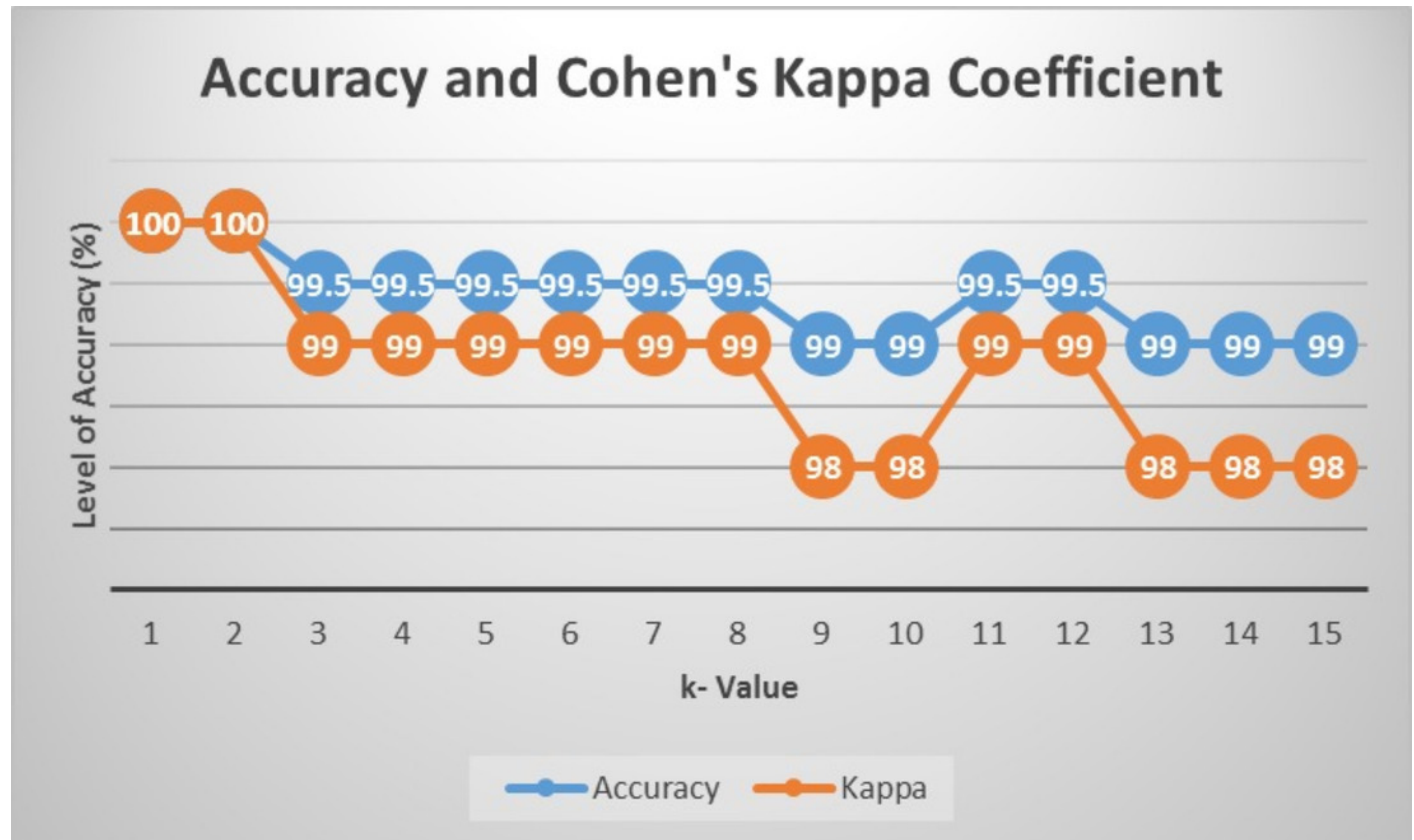
Figure 6

MSE values for *k*-NN, ANN, SVM, and Cart
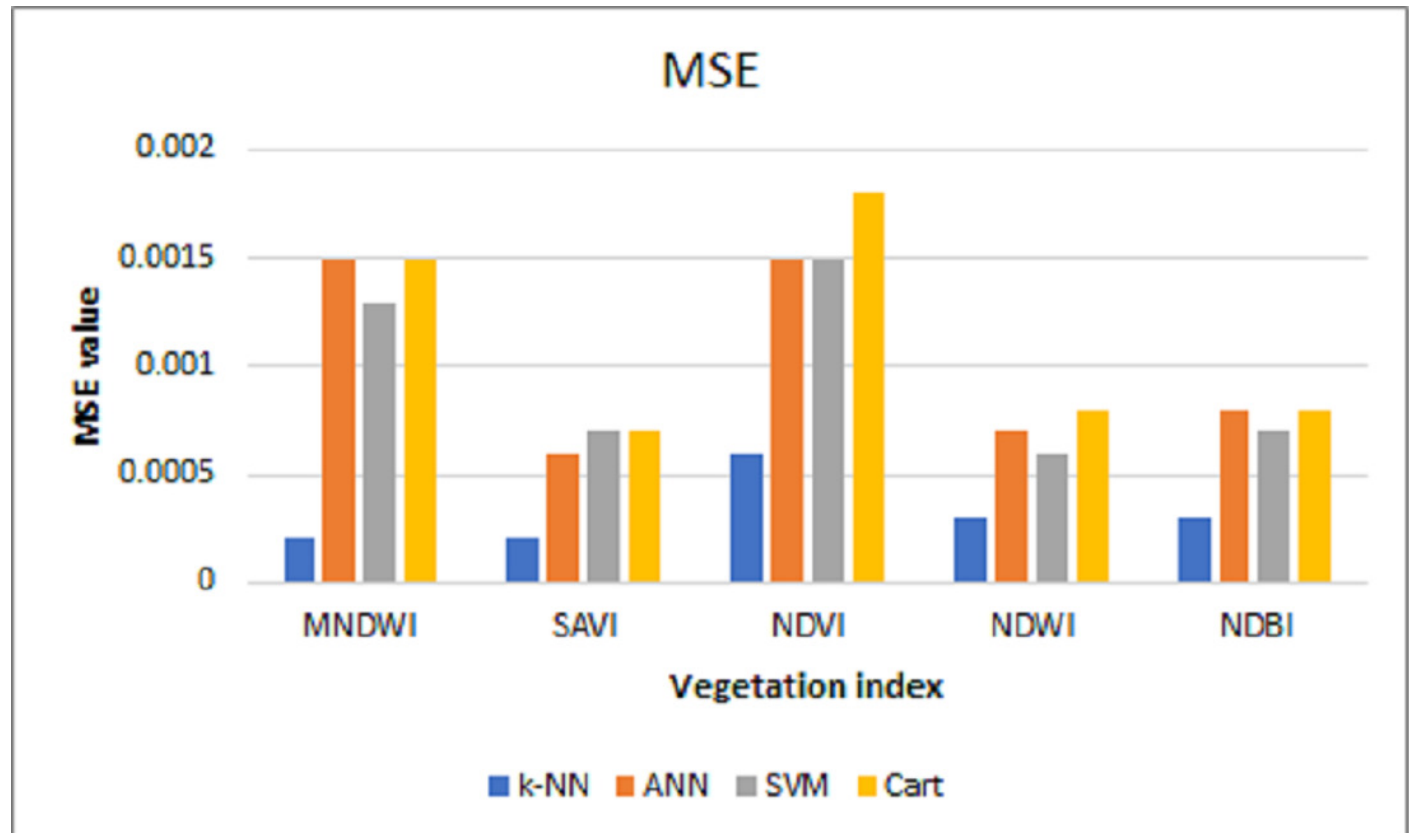
Figure 7

RMSE values for *k*-NN, ANN, SVM, and Cart
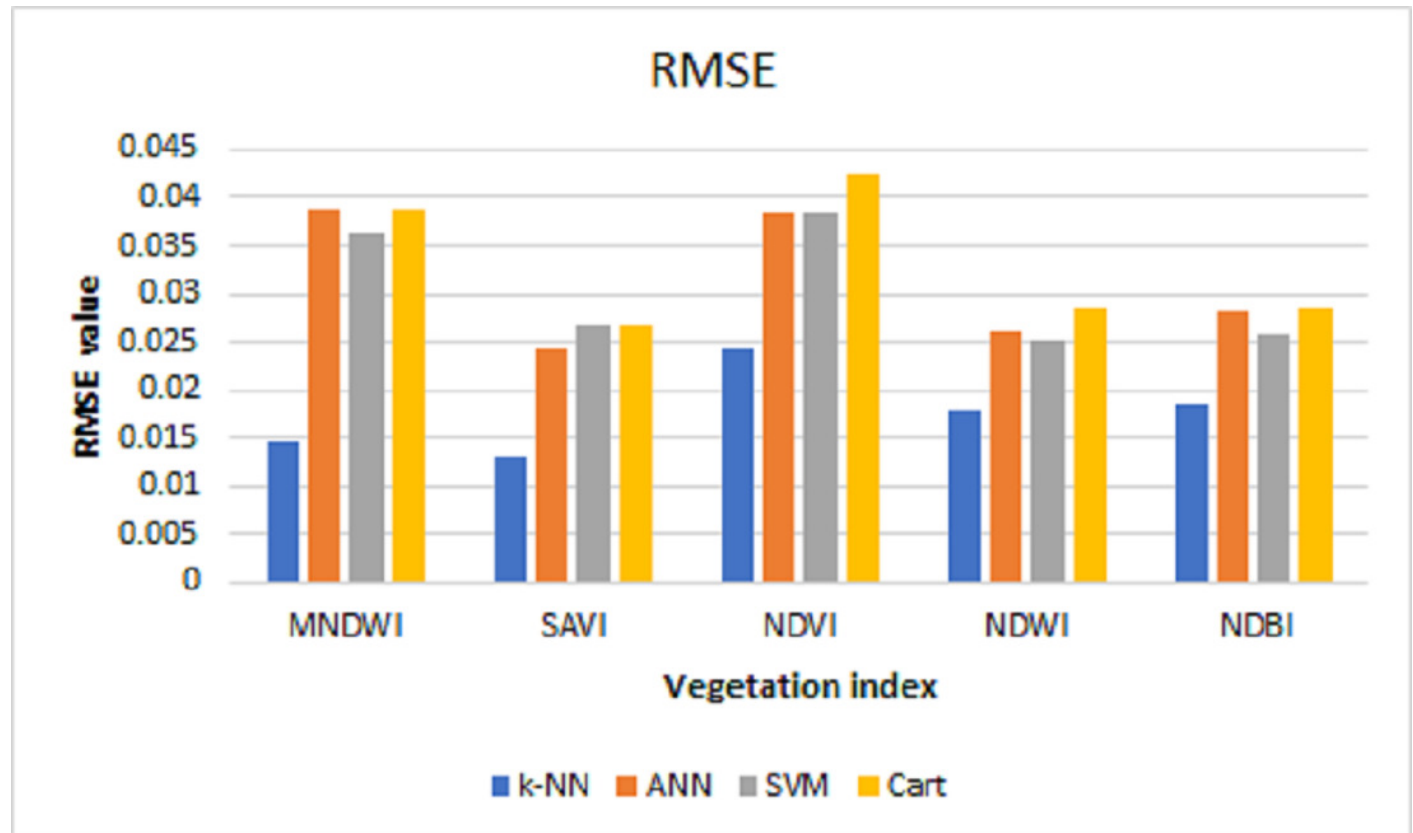
Figure 8

MAE values for *k*-NN, ANN, SVM, and Cart
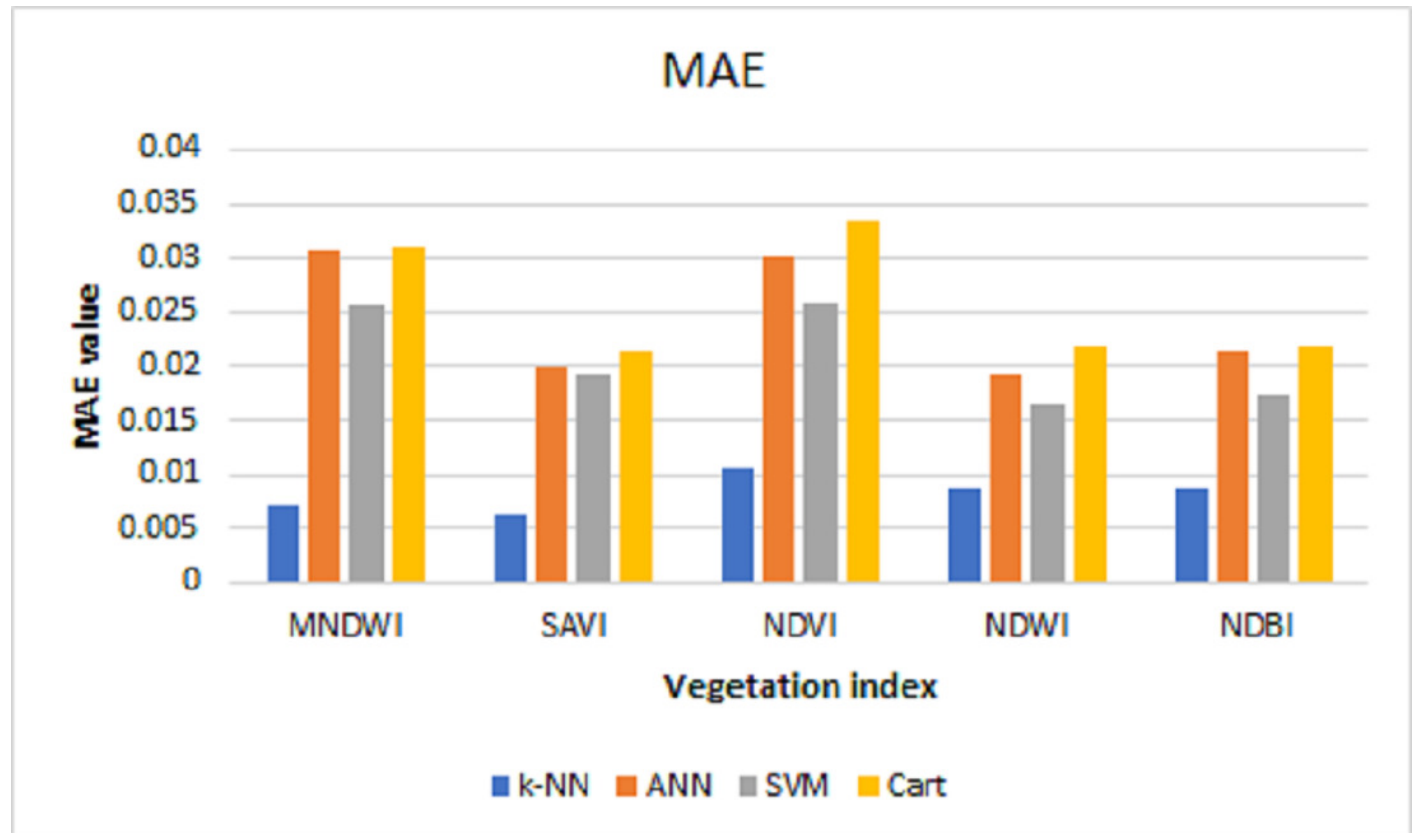
Figure 9

The Distribution Map of Tsunami Risk Areas Based for NDWI Index 2021
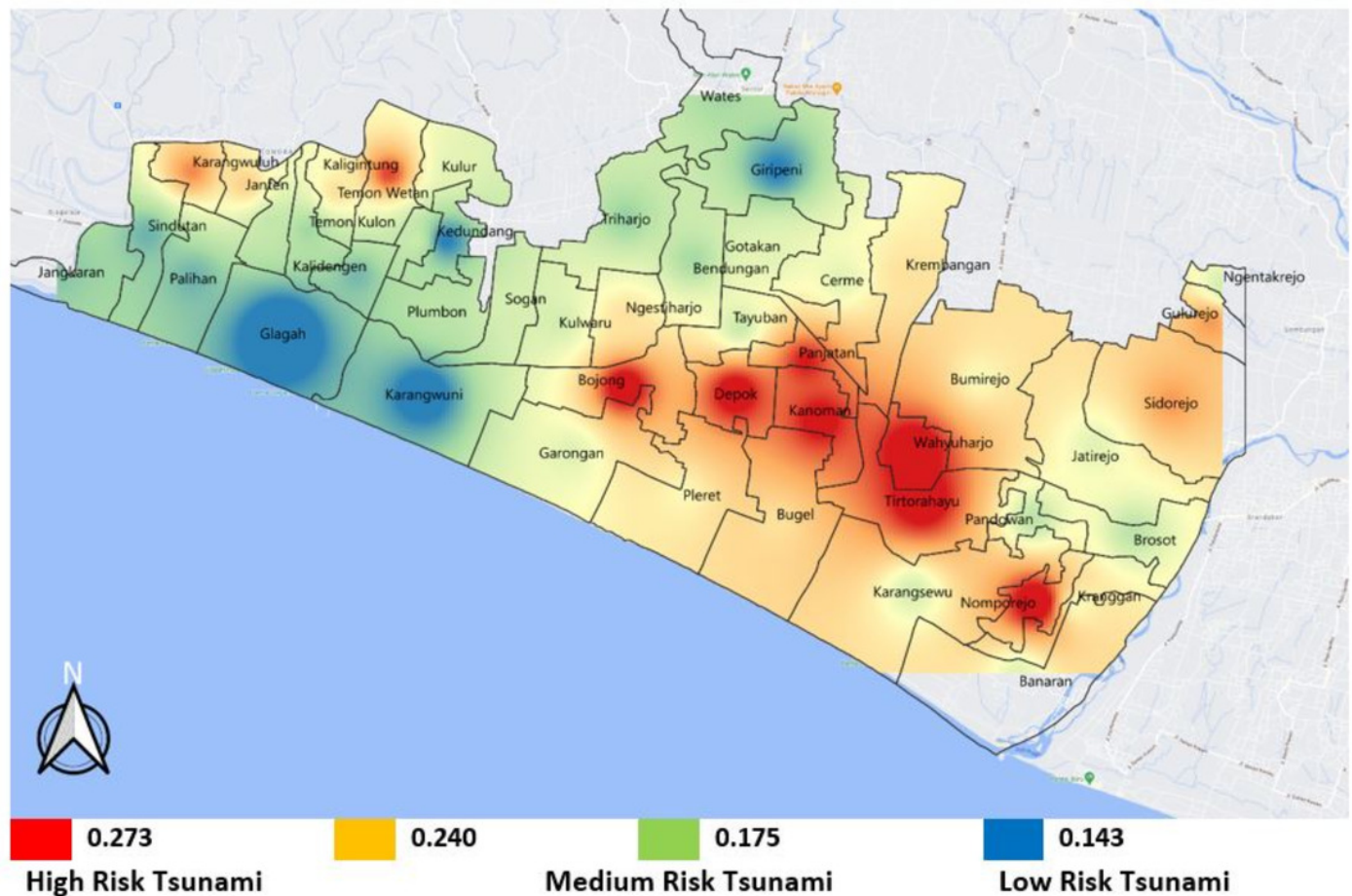
Figure 10

The Distribution Map of Tsunami Risk Areas Based for NDVI 2021 Index
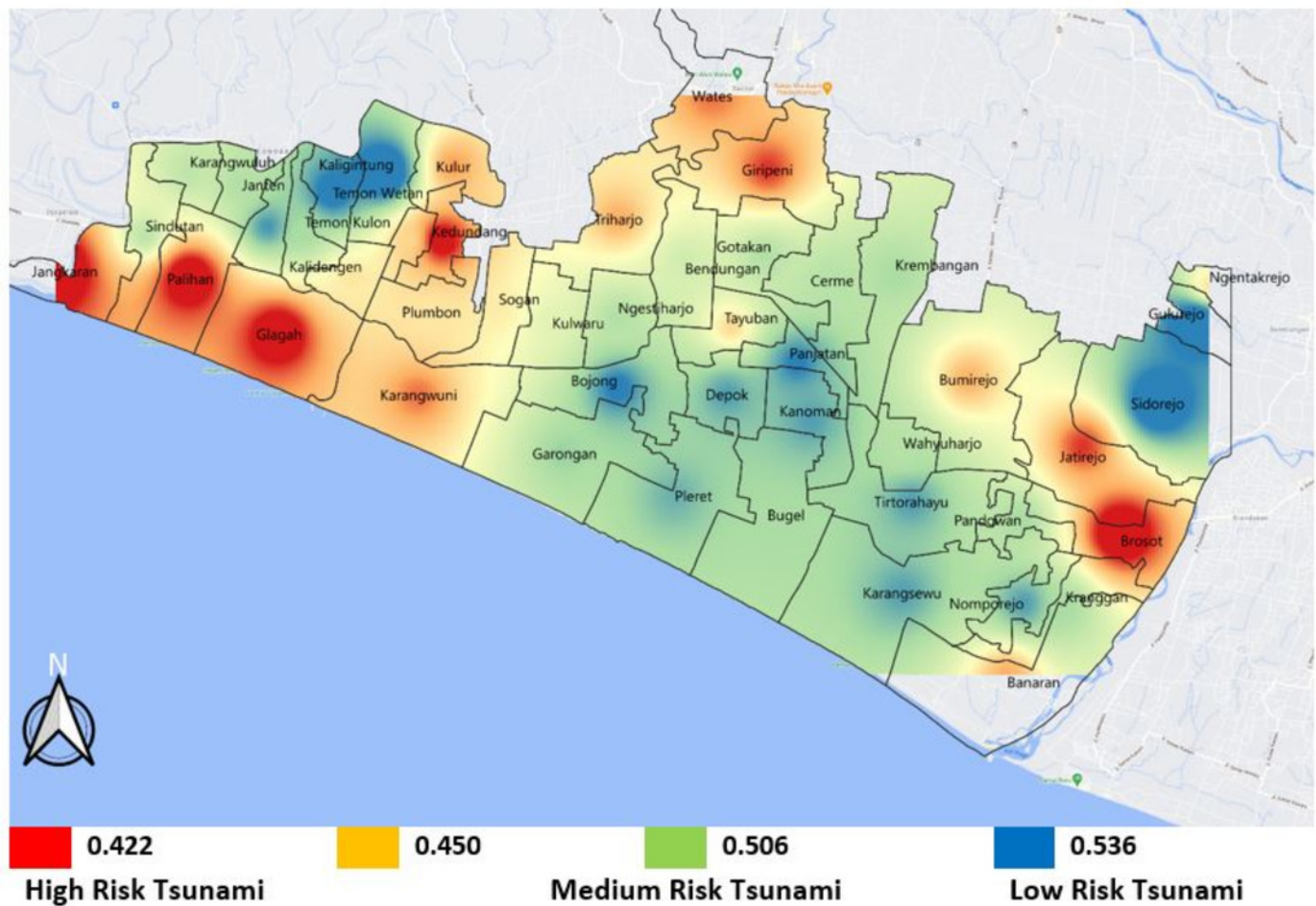
Figure 11

The Distribution Map of Tsunami Risk Areas for SAVI Index 2021

Figure 12

The Slope Map