

Satellite imagery and machine learning for aridity disaster classification using vegetation indices

Sri Yulianto Joko Prasetyo¹, Kristoko Dwi Hartomo², Mila Chrismawati Paseleng³, Dian Widiyanto Chandra⁴, Edi Winarko⁵

^{1,2,3,4}Department of Informatic Engineering, Satya Wacana Christian University, Indonesia

⁵Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia

Article Info

Article history:

Received Nov 18, 2019

Revised Jan 15, 2020

Accepted Feb 20, 2020

Keywords:

Machine learning

Remote sensing

Spatial autocorrelation

Vegetation indices

ABSTRACT

Central Java Province is one of provinces in Indonesia that has a high aridity risk index. Aridity disaster risk monitoring and detection can be done more accurately in larger areas and with lower costs if the vegetation index is extracted from the remote sensing imagery. This study aims to provide accurate aridity risk index information using spectral vegetation index data obtained from LANDSAT 8 OLI satellite. The classification of drought risk areas was carried out using k-nn with the Spatial Autocorrelation method. The spectral vegetation indices used in the study are NDVI, SAVI, VHI, TCI and VCI. The results show a positive correlation and trend between the spectral vegetation index influenced by seasonal dynamics and the characteristics of the High R.A. and Middle R.A. drought risk areas. The highest correlation coefficient is SAVI with a High R.A. amounted to 0.967 and Middle R.A. amounted to 0.951. The results of the Kappa accuracy test comparison show that SVM and k-nn have the same accuracy of 88.30. The result of spatial prediction using the IDW method shows that spectral vegetation index data that initially as an outlier, using the k-nn method, the spectral vegetation index data can be identified as data in the aridity classification. The spatial connectivity test among sub-districts that experience drought was done using Moran's I Analysis.

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



Corresponding Author:

Sri Yulianto Joko Prasetyo

Department of Informatic Engineering,

Satya Wacana Christian University,

Jln Diponegoro 52–60, Salatiga, Central java, Indonesia.

Email: sri.yulianto@uksw.edu

1. INTRODUCTION

In Indonesia, disasters that often occur are classified into two types of disaster, namely, hydro meteorological disaster (78% or 11,650 cases) and geological disaster (22% or 3,800 cases) [1, 2]. The negative impacts that arise from a disaster are determined using a disaster risk index, which is calculated based on the factors of threat or danger, vulnerability, and disaster capacity. Central Java Province is one of the regions in Indonesia with a high disaster risk index (based on the level of vulnerability and disaster capacity) and has become a priority area for hydro meteorological disaster management. Aridity threat indices in Central Java have been classified in the high category covering an area of 3.2 million hectares. The assessment of the aridity disaster index in Central Java was carried out using meteorological parameters namely standardized precipitation index (SPI) [3]. The SPI method produces accurate aridity information, though it is considered as a conventional method because of its dependence on the availability of rainfall data; and is only effective for a long-term aridity assessment. In areas that have a limited number

of meteorological stations, the aridity information produces by this method is inaccurate and invalid [4, 5]. Based on the data obtained from the Indonesian geophysics, climatology and meteorology agency (BMKG), Central Java Province only has six meteorology stations i.e. stations in Cilacap, Tegal, Semarang, Tanjung Mas, Ahmad Yani, and Banjarnegara. The limited number of meteorology stations in Central Java has caused inaccurate and invalid aridity indices information which does not represent the real situation that also affects the SPI analysis although the analysis is combined with spatial interpolation.

Based on literature studies, it can be seen that the technology currently available makes it possible to detect and monitor the risk of aridity in a wider area with lower costs and high accuracy by using vegetation index extracted from remote sensing imagery [6-8]. Remote sensing has been used to monitoring and modeling the phenomena of hydrometeorology, soil movement, disaster and landuse. However, remote sensing data analysis requires the development of techniques, models, and methods to transform spectral information into easily interpretable forms [9-13]. Currently, more than 150 spectral vegetation indices have been discovered and published in various scientific literatures throughout the world. The spectral vegetation indices can be used as indicators of aridity disaster assessment because they provide information in the form of spectral characteristics in the process of aridity that runs very slowly and lasts for a long time, and which the onset of aridity cannot be ascertained. The spectral vegetation indices are represented as a single number of intensity, duration, and spatial extent of aridity risk [14].

In this paper, we propose to develop a framework to transform the spectral vegetation index into aridity risk index information. The proposed framework is a combination of the k-nearest neighbor (k-nn) and the Spatial Autocorrelation (SA) methods [15, 16]. We name the framework for assessing the aridity index as Satellite Imagery and Machine Learning Autocorrelation. The choice of k-nn method as the algorithm on the framework of aridity risk indices prediction is due to its ability to identify and classify the nearest neighbors of each spectral vegetation indices data extracted from the pixels of satellite imageries. The results of identification and classification of the spectral vegetation indices data group will be classified in different classes according to the distance of neighbors [17]. The SA method will calculate the strength of spatial connectivity of the class group data that have been classified by the k-nn. The results of spatial connectivity are visualized in the form of regional aridity pattern maps. Interpolation using the inverse distance weight (IDW) method is conducted in order to determine the aridity value not as a sample. The SA is the statistical method used to visualize the patterns of spatial distribution and exploration of vegetation, aridity levels and daily rainfall [17-19]. In the next sections of this paper, section 2 discusses the theoretical background proposed in the framework, Section 3 discusses the research methods, Section 4 covers the results and discussions, Section 5 discusses the conclusions, and the last section is the references.

2. BACKGROUND

Spectral vegetation indices is a quantitative value of the measurement of vegetation canopy in receiving and reflecting the spectrum of light and interpreted as the spectral characteristic of vegetation, including the infrared spectrum as visible light (IR) and near infrared spectrum as invisible light (NIR). The reflectance of IR and NIR light from a vegetation canopy is determined by the structure of the canopy (number and orientation of the leaf) and the biochemical properties of the canopy (chlorophyll and carotenoid). VI is related to the characteristics of vegetation which includes: (1) types of vegetation (trees, shrubs, grasslands, etc.), (2) cropping patterns, (3) phases of plant growth, (4) leaf pigments, (5) water on plants, and (6) land. The aridity indicator vegetation index are: NDVI, SAVI, VCI, TCI and VHI was made using formula (1-5) [20-27]:

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} = \frac{band\ 5 - band\ 4}{band\ 5 + band\ 4} \quad (1)$$

$$VCI = \frac{NDVI_a - NDVI_{min}}{NDVI_{max} + NDVI_{min}} \times 100 \quad (2)$$

$$VHI = \alpha VCI + (1 - \alpha) TCI \quad (3)$$

$$SAVI = \frac{(\rho_{NIR} - \rho_{Red})}{(\rho_{NIR} + \rho_{Red} + L)} (1 + L) \quad (4)$$

$$TCI = \frac{LST_{max} - LST_a}{LST_{max} + LST_{min}} \times 100 \quad (5)$$

The classification and prediction stages were carried out by calculating each vegetation index value of NDVI, SAVI, VCI, TCI and VHI using formulas are: (1) k-nn, (2) RF and (3) SVM. The k-nn technique is an accurate method for pattern recognition and classification methods based on spatial objects that have been previously identified from remote sensing data particularly of pectral vegetation indices [15, 28]. In remote sensing analysis, the k-nn algorithm works in two stages, i.e. (1) learning, aims to create a class of data classification through neighboring relation among data, and (2) prediction to determine the classification of new data that has not yet have a neighboring relationship in a class that has been formed in the previous stage [15]. The main principle of k-nn is to classify data points that are distributed spatially based on the category of nearest neighbors. A set of training data of which $M = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\}$ contains n data entities, which $x_i \in R^d$ and $y \in Y = \{e_1, e_2, e_3, \dots, e_n\}$. If there is new data with a value, then the size of neighborliness is $N_k(a)$ [29]. The distance function between data points is calculated to compare the similarity of one sample point with another sample point. The distance function is determined using the Euclidean distance and Manhattan distance equations as follows (6) [30]:

$$(d(x_m, x_n) = \sqrt{\sum_i (x_{m,i} - x_{n,i})^2}) = (d(x_m, x_n) = \sum_i |x_{m,i} - x_{n,i}|) \quad (6)$$

The testing the performance classification of the area of aridity using the k-nn algorithm on spectral vegetation indices particularly of NDVI and SAVI data shows the value of accuracy of Cohen's Kappa Statistics is more stable with a confidence level between 95-100 percent [31].

$$Z = \frac{\sum_{i=1}^n \frac{z}{(h_{ij} + \delta)^\beta}}{\sum_{i=1}^n \frac{1}{(h_{ij} + \delta)^\beta}} \quad (7)$$

Determination of aridity risk was done by determining the distribution of aridity index throughout all regions using IDW interpolation method. By using the IDW method, regions that were not used as samples and whose aridity indices were unknown, the assessment was determined based on the mean of surrounding sample regions. Determination of IDW was made using formula (8) [32]. Where Z is the non-sampled location, Z_i is the known value, β is the weight and δ is the smoothing factor.

$$h_{ij} = \sqrt{(\Delta x)^2 + (\Delta y)^2} \quad (8)$$

The value of h_{ij} is the distance between the known and unknown sample points and is determined by the Euclidean equation as in equation (8). Where Δx and Δy are the distance between unknown points on j and sample one i . The purpose of Moran's I calculations on vegetation index data is to determine the spatial connectivity of vegetation characteristics between sampling points. The Moran's I equation is shown in (9):

$$I = \frac{N \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^n \sum_{j=1}^n w_{ij}) \sum_{i=1}^n (x_i - \bar{x})^2} \quad (9)$$

where N is the number of observations made, x is the vegetation index used as an observation indicator. While, the variable w_{ij} is the weight between the observation areas to i and j at t time [16].

3. RESEARCH METHOD

The research area consists of 346 subdistricts in Central Java Province, Indonesia. The data used for this research are remote sensing data obtained from the monthly data of Landsat 8 OLI (operational land imager) satellite image of 2018–2019 periods and data from the United States geological survey (USGS) <https://earthexplorer.usgs.gov/> with path/row 120/65. The framework of aridity risk assessment that is proposed in this research is presented in Figure 1. which the stages of research implementation are as follows:

- Data preprocessing, LANDSAT 8 OLI images consisting of 11 bands in which each band has a different wavelength. The image wavelengths between 0.43 – 12.51. Vegetation index data is numeric types consisting of NDVI, SAVI, VHI, TCI and VCI data. The image band will be calculated to produce vegetation index as in the examples shown in (2).
- Data classification, the vegetation indices of NDVI, SAVI, VHI, TCI and VCI were sought for their correlation values to see the relationship between indices, distribution patterns and trends over a certain period. Each data of vegetation indices where divided into two categories, namely testing data at 70%

- and learning data at 30% of the total data. Vegetation indices data were classified into two groups, namely, high risk aridity and middle risk aridity using the k-nn method.
- Performance test, conducting of data analysis using SVM and RF as a comparison of classification performance. Classification and prediction of vegetation indices in High R.A aridity and Middle R.A risk areas groups were done using the R programming.
 - Predicting the spatial distribution of k-nn classification results using the Inverse Distance Weight method. Determining the spatial connectivity between regions to determine the spatial pattern risk aridity distributions in the research areas using Moran's I. Compiling the local aridity indices using the data of k-nn spatial prediction.

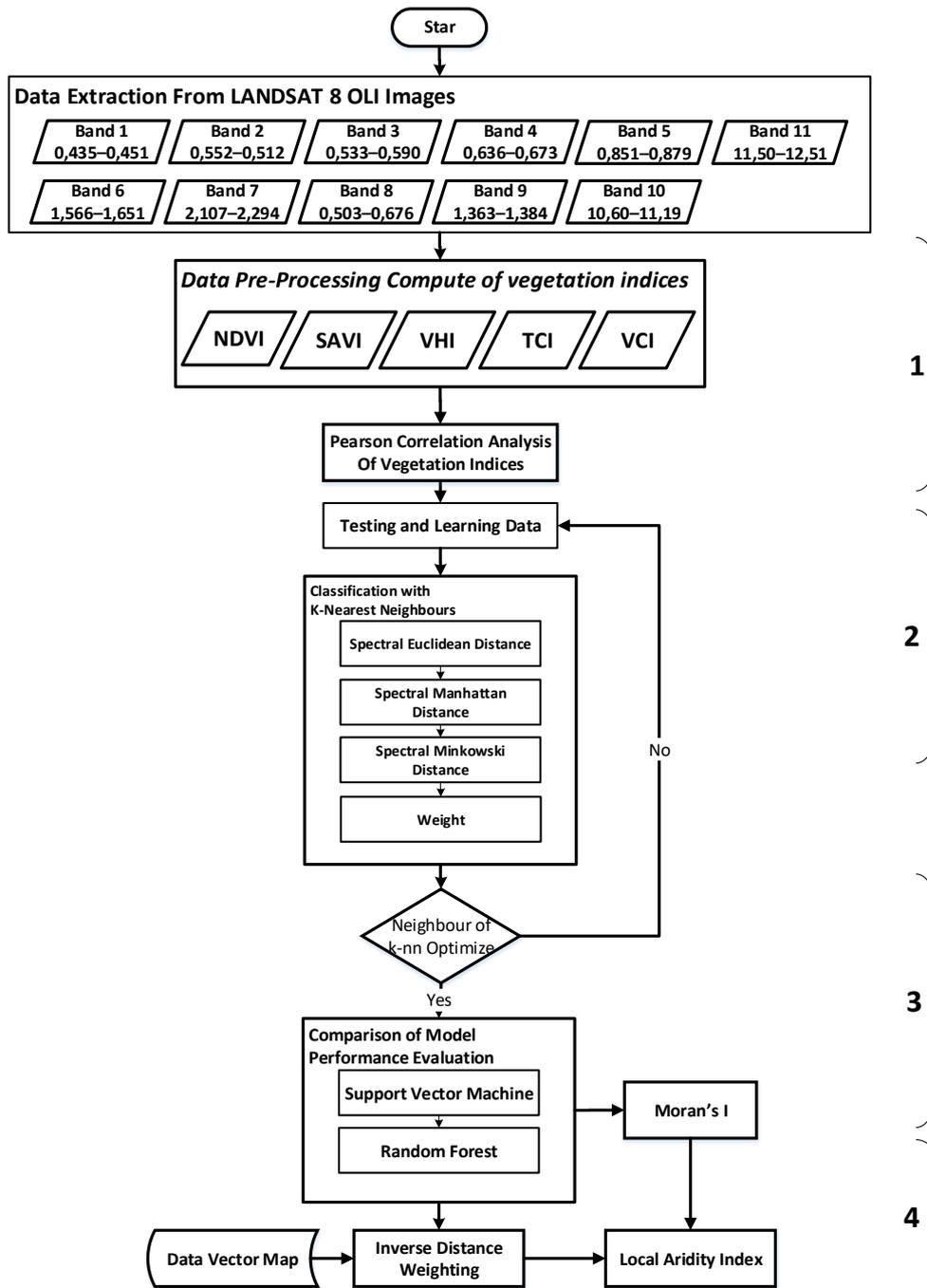


Figure 1. The proposed regional aridity risk assessment framework

The procedures of classification, prediction and data visualization algorithm were made as shown in the following pseudocode:

```

Input:data=(value_vi),(value_pred_vi),mi=raining data, mj=lable data, ml=sample data,
location_data=coordinates(data),location_sample=coordinates(point_sample),data_idw,power=2,
estimated_value,weighted_value ,estimated_value,distance
Classification (mi , mj , ml) ,
  for l to x do calculation of node distance at n(mi , ml)
end for
  calculation of n distance > ni = classification n(mi , ml)
  calculation of n distance < ni back to mj in which mi,l ∈ I
end for
morans'i
mor={}
  for i=1 to n do value_calculation={}, for j=1 to n do value_calculation={}
  a=a + weighti,j * (xi - averagex) * (xj - averagex)
  b2=(b2 + xi - averagex), b1=(b1 * weighti,j * b2)
  I = N * a/b1
  if I > -1/N-1 then positive aurocorrelation
  else
  if I = 0 then positive aurocorrelation
  else
  if I < -1/N-1 then negative aurocorrelation
end if
end for
Inverse distance weight
idw={}
  for i=1 to n do value_calculation={}
  value_weight={} for j=1 to m do distance = location_sample - location_observation;
  value_weightj=1/distancepower value)calculationj=value_weightj* dataj
end for
  data_idwi= (∑value_calculation)/(∑value_weight)

```

4. RESULTS AND DISCUSSION

Vegetation index data follows the seasonal pattern of wet months (occurring from November to March), dry months (occurring from May to September), and transitional months (occurring from April and October) [33]. In addition to the seasonal cycles of wet, dry and transitional months, there is a classification of aridity risk areas consisting high risk aridity (High R.A.) and middle risk aridity (Middle R.A.). In the preprocessing step of the experiment, is to analyze correlations and trends among seasonal patterns and the classification of risk areas. The results of this analysis will help understanding the temporal connectivity of vegetation index, seasonal changes and the risk of local aridity. Pearson correlation analysis results show a positive correlation and trend between vegetation index influenced by seasonal dynamics and the characteristics of the High R.A aridity and Middle R.A risk are as shown Figure 2.

The analysis shows that all vegetation indices have a positive correlation and trend with High R.A and Middle R.A. The highest correlation coefficient is SAVI with a High R.A amounted to 0.967 and Middle R.A. amounted to 0.951. SAVI is a vegetation index that provides indicators of photosynthesis, biomasses, and biogeochemical processes at the local scale even though the area has seasonal dynamics. The lowest correlation coefficient is VCI with High R.A. amounted to 0.068 and Middle R.A. amounted to 0.382. VCI is a vegetation index that provides indicators of the diversity of local vegetation in relation to seasonal factors as shown in Figure 2. The next step is to perform the k-nn analysis and test the accuracy with confusion matrix method on k-1 up to k-10. The results of the accuracy testing from k-1 to k-10 are presented as the curve of Figure 3.

The curve in Figure 3 shows that the lowest value is 80.55 at k-2 and the average is 86.99 and the highest value is 92.03 at k-9. The next step is to compare the accuracy test between the confusion matrix method and the Kappa method from the data from the k-nearest neighbors (k-nn) classification, support vector machine (SVM) and random forest (RF) as shown in Table 1. The results of confusion matrix accuracy test comparison show that SVM is 96.00, which is more accurate when compared to RF and k-nn. The results of the Kappa accuracy test comparison show that between SVM and k-nn have the same accuracy of 88.30.

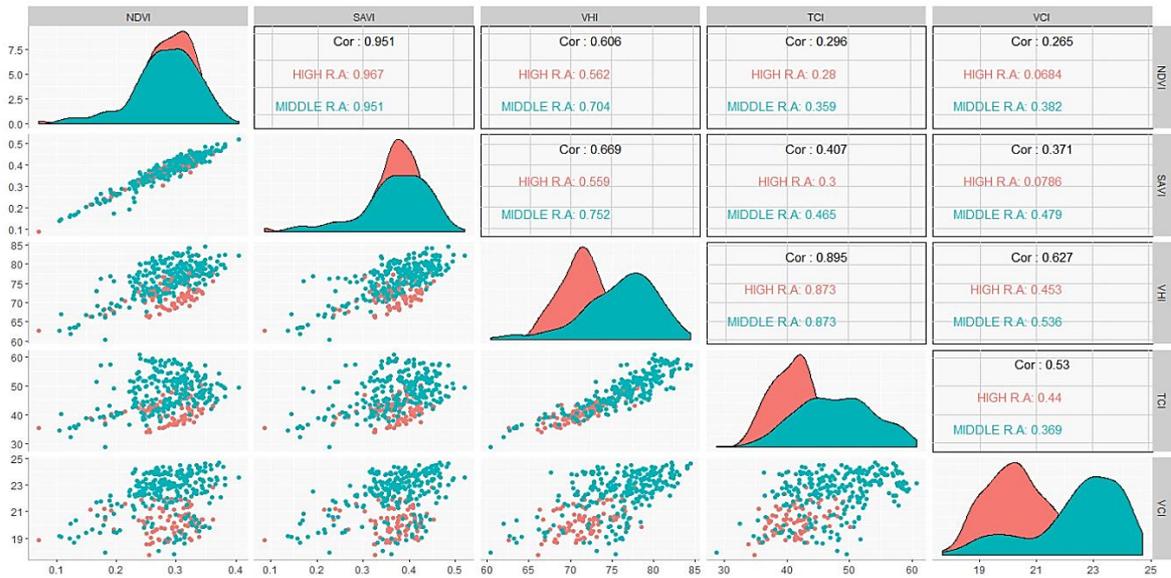


Figure 2. The scatterplot analysis using Pearson correlation shows positive correlations and positive trends



Figure 3. Trend analysis of k-nn and accuracy testing using the confusion matrix method

Table 1. Comparison of data classification using other ML methods

Classification method	Confusion matrix	Kappa
Support Vector Machine	96.00	88.30
Random Forest	91.00	75.00
k-nearest neighbours	92.03	88.30

Classification data are analyzed using a boxplot diagram, with the aim of determining: (1) the lowest observation value, (2) the lowest quartile value (Q1), (3) the median value (Q2), (4) the highest quartile value (Q3), and (5) outlier value. The boxplot analysis results show that the classification data have a wide range that exceeds the range of testing data from the minimum limit (Q1) or the maximum limit (Q3) as shown in Figure 4. The width of the Q1 and Q2 data ranges, and the shift in the center of the boxplot are caused by the inclusion of data that is previously included as outlier into the classification of High R.A or Middle R.A. The visualized outlier data points are at the High R.A and Middle R.A boxplot classifications both in predictive data and LANSAT imagery data. An example is the outlier at the knn prediction data which is almost as the same as the data from the LANDSAT imagery. Shifts in the values of Q1, Q2, and Q3 are clearly identified in TCI, VCI and VHI. The vegetation index of TCI and VCI are indicators of the occurrence of aridity caused by seasonal cycles and their effect on canopy temperature, leaf surface stomata movement and humidity level of the canopy. The dynamics of TCI and VCI are caused by seasonal fluctuation factors that will have an impact on the health of vegetation which is shown to increase the VHI value. Figure 4 shows that most of the classification data are in the Middle R.A classification,

which is indicated by the length of the inner quartile range (IQR). The previous outlier shifting data that has become data in the High R.A group or in the middle R.A group shows that there are changes in the curve peaks and line trends on the line graph. The red line represents High R.A and the blue line represents Middle R.A. The line curve between the NDVI and SAVI vegetation indices indicates a similar trend which the position of the second peak point of the vegetation indices is between 0.2-0.4, which indicates that in some parts of the study areas have vegetation and have high photosynthetic activities as shown in Figure 4. The line curve shows that knn.NDVI and knn.SAVI prediction lines show patterns that almost identical to the NDVI and SAVI line patterns from the LANDSAT imagery data. On the line curve as shown in Figure 4, it can be seen that the Middle R.A line has a higher and wider curve compared to the High R.A line.

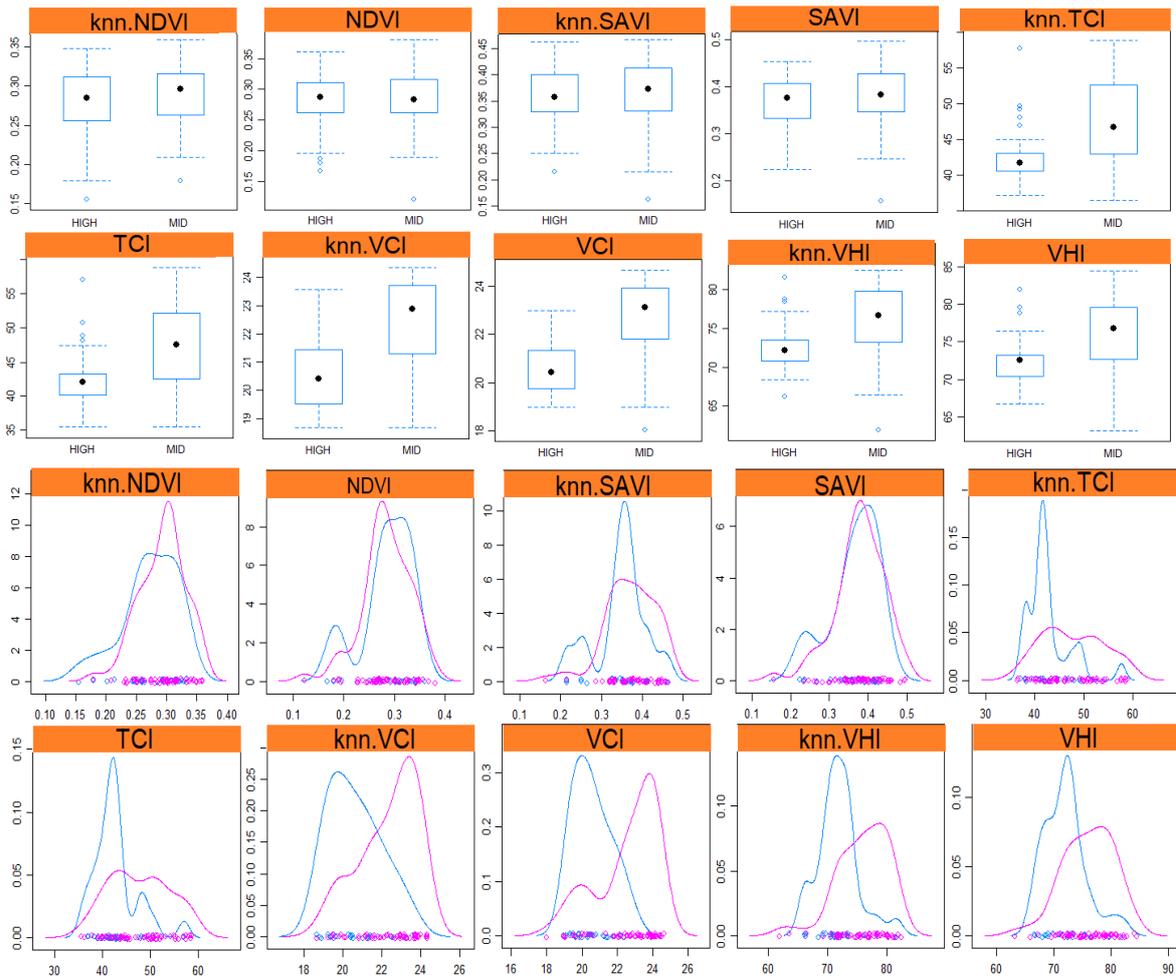


Figure 4. Data are in the Middle R.A classification, which is shown by the length of the inner quartile range (IQR). The Middle R.A curve line has a higher and wider curve compared to High R.A curve line

The next step is to make a spatial prediction on the value of the classification results using inverse distance weight (IDW) to visualize of spatial pattern Vegetation Indices and value of k-nn vegetation indices at areas that are not used as observation points. Spatial prediction results are in the form of an aridity risk map for sub-district level or lower as shown in Figure 5. In Figure 5 several sub-districts are shown as brighter blue nodes and are shown in darker color. This phenomenon shows that the vegetation index data that were previously in the outlier classification, through the k-nn prediction process, the data are then included in the Middle R.A (gradation of blue colors) and High R.A classifications (gradation red colors).

Natural phenomena indicate that various spatial elements interact with each other especially with neighboring spatial objects [30]. The results of the Moran's I test shown in Table 2. Table 2 shows that all vegetation indices are positive autocorrelation, which is interpreted to have Middle R.A. and High R.A. in the entire observation areas.

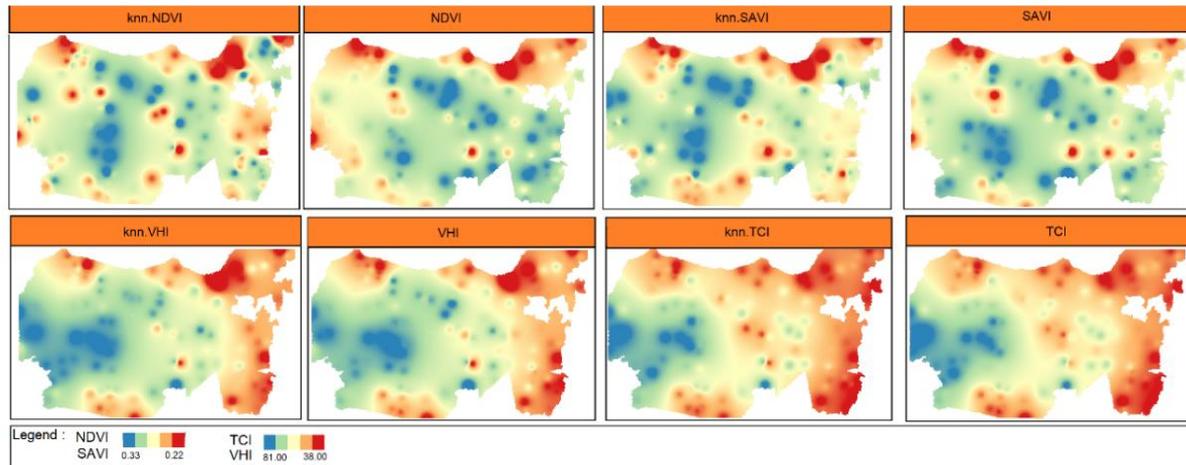


Figure 5. The figure shows the comparison of predicted results of High R.A. and Middle R.A. k-NN algorithm works by inserting outlier data into the classification of vegetation index data, and data classification

Table 2. Analysis of Moran's I on testing data and prediction vegetation indices data

Vegetation index	Moran's I	Interpretation
NDVI	0.996	Positive Autocorrelation
Knn-NDVI	0.996	Positive Autocorrelation
SAVI	0.996	Positive Autocorrelation
Knn-SAVI	0.996	Positive Autocorrelation
VHI	0.997	Positive Autocorrelation
Knn-VHI	0.995	Positive Autocorrelation
TCI	0.995	Positive Autocorrelation
Knn-TCI	0.995	Positive Autocorrelation
VCI	0.995	Positive Autocorrelation
Knn-VCI	0.995	Positive Autocorrelation

5. CONCLUSION

The analysis results of Pearson correlation show that there is a positive correlation and trend among vegetation indices influenced by seasonal dynamics and the characteristics of High R.A. and Middle R.A. drought risk areas. All vegetation indices have a positive correlation and trend with High R.A. and Middle R.A. The correlation coefficient of spectral vegetation index shows that SAVI has the highest correlation with High R.A. amounted to 0.967 and with Middle R.A. amounted to 0.951. SAVI provides indicators of photosynthetic activities, biomasses, and biogeochemical processes on a local scale even though the region has seasonal dynamics. The lowest correlation coefficient is VCI with High R.A. amounted to 0.068 and with Middle R.A. amounted to 0.382. VCI provides indicators of the diversity of local vegetation in relation to seasonal factors. Testing of the classification accuracy between the confusion matrix method and the kappa method was made from the data of k-nearest neighbors (k-nn) classification, which support vector machine (SVM) and random forest (RF) show that the SVM is 96.00, which is more accurate when compared to RF or k-nn. The results of the Kappa accuracy test comparison show that SVM and k-nn have the same accuracy of 88.30. The spatial prediction of drought risk maps at the sub-district level shows that the spectral vegetation index data that was previously in the outlier classification, through the process of k-nn prediction into data on Middle R.A. and High R.A. Moran's I test results show that all vegetation indices are positive autocorrelation, which is interpreted to be the Middle R.A. and High R.A. in the entire observation area.

ACKNOWLEDGEMENTS

This research was conducted using Featured Grant PTUPT from Directorate General of Higher Education, Ministry of National Education Republic Indonesia year 2018 - 2021. This research was conducted with the support of data from the Climatology Meteorology and Geophysics Agency of Central Java.

REFERENCES

- [1] M. R. Amri et al., "Risiko bencana indonesia," *Badan Nasional Penanggulangan Bencana*, Jakarta, 2016.
- [2] A. W. Jayawardena, "Hydrometeorological disasters: Causes, effects and mitigation measures with special reference to early warning with data driven approaches of forecasting" *Procedia IUTAM*, vol. 17, no. 2013, pp. 3-12, 2015.
- [3] BNPB, "Kajian risiko bencana jawa tengah 2016-2020," *Deputi Bidang Pencegahan dan Kesiapsiagaan Badan Nasional Penanggulangan Bencana*, 2016.
- [4] A. Shahabfar and J. Eitzinger, "Spatio-temporal analysis of droughts in semi-arid regions by using meteorological drought indices," *Atmosphere (Basel)*, vol. 4, no. 2, pp. 94-112, 2013.
- [5] N. Sanchez et al., "A new soil moisture agricultural drought index (SMADI) integrating MODIS and SMOS products: A case of study over the Iberian Peninsula," *Remote Sens.*, vol. 8, no. 4, pp. 1-25, 2016.
- [6] S. S. Patil, U. B. Angadi, and D. K. Prabhuraj, "Machine learning technique approaches versus statistical methods in classification of multispectral remote sensing data using maximum likelihood classification: Koluru Hobli, Bellary Taluk, District , Karnataka, India," *International Journal of Advanced Remote Sensing and GIS*, vol. 3, no. 1, pp. 525-531, 2014.
- [7] M. Kanevski, A. Pozdnoukhov, and V. Timonin, "Machine learning algorithms for geospatial data. Applications and software tools," *9th Int. Congr. Environ. Model. Softw.*, pp. 320-327, 2008.
- [8] J. A. Quinn, M. M. Nyhan, C. Navarro, D. Coluccia, L. Bromley, and M. Luengo-Oroz, "Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 376, no. 2128, pp. 1-16, 2018.
- [9] A. Melesse, Q. Weng, S. Prasad, and G. Senay, "Remote sensing sensors and applications in environmental resources mapping and modelling" *Sensors*, vol. 7, no. 21, pp. 3209-3241, 2007.
- [10] M. L. Villarreal, L. M. Norman, S. Buckley, C. S. A. Wallace, and M. A. Coe, "Multi-index time series monitoring of drought and fire effects on desert grasslands," *Remote Sens. Environ.*, vol. 183, pp. 186-197, 2016.
- [11] A. Bannari, D. Morin, F. Bonn, and A. R. Huete, "A review of vegetation indices," *Remote Sens. Rev.*, vol. 13, no. 1-2, pp. 95-120, 1995.
- [12] J. Xue and B. Su, "Significant remote sensing vegetation indices: a review of developments and applications," *J. sensors*, vol. 2017, pp. 1-17, 2017.
- [13] F. Caparrini and F. Manzella, "Hydrometeorological and vegetation indices for the drought monitoring system in Tuscany region, Italy," *Adv. Geosci.*, vol. 17, pp. 105-110, 2008.
- [14] L. Ji and A. J. Peters, "Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices," *Remote Sens. Environ.*, vol. 87, no. 1, pp. 85-98, 2003.
- [15] L. Samaniego, A. Bardossy, and K. Schulz, "Supervised classification of remotely sensed imagery using a modified k-NN technique," *IEEE T Geosci Remote*, vol. 46, no. 7, pp. 2112-2125, 2008.
- [16] S. Yulianto Joko Praetyo, K. Dwi Hartomo, B. Hasiholan Simanjuntak, and D. Widiyanto Candra, "Mitigation & identification for local aridity, based of vegetation indices combined with spatial statistics & clustering K Means," *OP Conf. Series: Journal of Physics: Conf. Series*, vol. 1235, pp. 1-8, 2019.
- [17] S. Y. J. Prasetyo, Subanar, E. Winarko, and B. S. Daryono, "The prediction of population dynamics based on the spatial distribution pattern of brown planthopper (*Nilaparvata lugen* Stal.) using exponential smoothing-local spatial statistics," *J. Agric. Sci.*, vol. 5, no. 5, pp. 1-17, 2013.
- [18] M. Javari, "Geostatistical modeling to simulate daily rainfall variability in Iran," *Cogent Geosci.*, vol. 3, no. 1, pp. 1-29, 2017.
- [19] Y. Shu-jie, L. Xiao-dong, and Y. Hou-gang, "The study for the spatial distribution pattern of NDVI in the western of Jilin Province the study for the spatial distribution pattern of NDVI in the western of Jilin Province," *IOP Conference Series: Earth and Environmental Science*, vol. 113, pp. 1-7, 2018.
- [20] N. P. Robinson et al., "A dynamic landsat derived normalized difference vegetation index (NDVI) product for the conterminous United States," *Remote Sens.*, vol. 9, no. 8, pp. 1-14, 2017.
- [21] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3025-3033, 2006.
- [22] M. Svoboda and B. A. Fuchs, "Handbook of drought indicators and indices," *World Meteorological Organization and Global Water Partnership, Integrated Drought Management Programme (IDMP), Integrated Drought Management Tools and Guidelines Series 2*, Geneva, 2016.
- [23] R. Bajgain, X. Xiao, P. Wagle, J. Basara, and Y. Zhou, "Sensitivity analysis of vegetation indices to drought over two tallgrass prairie sites," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 151-160, 2015.
- [24] B. Matsushita, W. Yang, J. Chen, Y. Onda, and G. Qiu, "Sensitivity of the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) to topographic effects: A case study in high-density cypress forest," *Sensors*, vol. 7, no. 11, pp. 2636-2651, 2007.
- [25] A. Klisch and C. Atzberger, "Operational drought monitoring in Kenya using MODIS NDVI time series," *Remote Sens.*, vol. 8, no. 4, pp. 267-1-22, 2016.
- [26] C. F. C. R. Chen, N. T. Son, C. F. C. R. Chen, S. H. Chiang, L. Y. Chang, and M. Valdez, "Drought monitoring in cultivated areas of Central America using multi-temporal MODIS data," *Geomatics, Nat. Hazards Risk*, vol. 5705, pp. 1-16, 2016.
- [27] R. I. Sholihah et al., "Identification of agricultural drought extent based on vegetation health indices of landsat data: Case of subang and Karawang, Indonesia," *Procedia Environ. Sci.*, vol. 33, pp. 14-20, 2016.
- [28] Y. Guo, S. Han, Y. Li, C. Zhang, and Y. Bai, "K-Nearest neighbor combined with guided filter for hyperspectral image classification," *Procedia Comput. Sci.*, vol. 129, pp. 159-165, 2018.

- [29] B. Javier and B. Javier, "K-nn Algorithm K-nn Regression," [Available at]: http://www.ccs.neu.edu/home/vip/teach/DMcourse/1_intro_sim_knn/notes_slides/03d-algind-knn-eng.pdf, (February, 2020)
- [30] T. L. T. Du, D. Du Bui, M. D. Nguyen, and H. Lee, "Satellite-based, multi-indices for evaluation of agricultural droughts in a highly dynamic tropical catchment, Central Vietnam," *Water (Switzerland)*, vol. 10, no. 5, pp. 1-24, 2018.
- [31] S. Y. Joko Prasetyo, K. D. Hartomo, M. C. Paseleng, D. C. Widiyanto, and B. H. Simanjuntak, "The machine learning to detect drought risk in Central Java using Landsat 8 OLI remote sensing images," *The 5th International Conference on Science and Technology*, 2019.
- [32] G. Garnero and D. Godone, "Comparisons between different interpolation techniques," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 40, no. XL-5/W3, pp. 139-144, 2013.
- [33] F. H. Widodo, E. Aldrian, and L. Du, "Seasonal variability of Indonesian rainfall in ECHAM4 simulations and in the reanalyses : The role of ENSO," *Theoretical and Applied Climatology*, vol. 87, no. 1, pp. 41-59, 2007.

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. Further, he's had 5 copy rights, and written some reference books on remote sensing data analysis and modelling in Indonesia and English. His another achievement was the runner up of outstanding lecturer at LLDikti VI.



Kristoko Dwi Hartomo, completed his study on the Doctorate Program of Computer Science, Science Faculty of Gadjah Mada University Yogyakarta in 2019. He has been active on research since 2008 until now on the Geography Information System. He has published his papers on international journals. Moreover, he's had 5 copy rights and written some reference books on computer science.



Mila Chrismawati Paseleng, has completed her study on the Education Technology, Pelita Harapan University in 2009. She has been active on research since 2009 until now on Education Technology. She has published her papers on international journals.



Dian Widiyanto Chandra, has completed her study on the Magister Information Science, Satya Wacana Christian University. He has been active on research since 2018 until now on Computer Science and Modeling. He has published her papers on international journals.



Edi Winarko, Received the B.Sc at Statistics from Faculty of Mathematics and Natural Sciences Gadjah Mada University Yogyakarta Indonesia in 2001, M.Sc. degree in Computer Science from the Queen's University Canada, United States in 2002. He received his Ph.D. degree in Computer Science from the Flinders University of South Australia in 2007 He is currently a lecturer in the Faculty of Mathematics and Natural Science from the Gadjah Mada University Yogyakarta Indonesia. His current research interests include data warehousing, data mining, and information retrieval