Satellite imagery and machine learning for identification of aridity risk in central Java Indonesia

Sri Yulianto Joko Prasetyo, Kristoko Dwi Hartomo and
Mila Chrismawati Paseleng

Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Central Java, Indonesia

ABSTRACT

This study aims to develop a software framework for predicting aridity using vegetation indices (*VI*) from LANDSAT 8 OLI images. *VI* data are predicted using machine learning (*ml*): Random Forest (RF) and Correlation and Regression Trees (CART). Comparison of prediction using Artificial Neural Network (ANN), Support Vector Machine (SVM), k-nearest neighbors (k-nn) and Multivariate Adaptive Regression Spline (MARS). Prediction results are interpolated using Inverse Distance Weight (IDW). This study was conducted in stages: (1) Image preprocessing; (2) calculating numerical data extracted from the LANDSAT band imagery using vegetation indices; (3) analyzing correlation coefficients between *VI*; (4) prediction using RF and CART; (5) comparing performances between RF and CART using ANN, SVM, k-nn, and MARS; (6) testing the accuracy of prediction using Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE); (7) interpolating with IDW. Correlation coefficient of *VI* data shows a positive correlation, the lowest r (0.07) and the highest r (0.98). The experiments show that the RF and CART algorithms have efficiency and effectivity in determining the aridity areas better than the ANN, SVM, k-nn, and MARS algorithm. RF has a difference between the predicted results and 1.04% survey data MAPE and the smallest value close to zero is 0.05 MSE. CART has a difference between the predicted results and 1.05% survey data MAPE and the smallest value approaching to zero which is 0.05 MSE. The prediction results of *VI* show that in 2020 most of the study areas were low vegetation areas with the Normalized Difference Vegetation Index (NDVI) < 0.21, had an indication of drought with the Vegetation Health Index (VHI) < 31.10, had a Vegetation Condition Index (VCI) in some areas between 35%–50% (moderate drought) and < 35% (high drought). The Burn Area Index (dBAI) values are between $-3,971$ and $-2,376$ that show the areas have a low fire risk, and index values are between $-0,208$ and $-0,412$ that show the areas are starting vegetation growth. The result of this study shows that the machine learning algorithms is an accurate and stable algorithm in predicting the risks of drought and land fire based on the VI data extracted from the LANDSAT 8 OLL imagery. The VI data contain the record of vegetation condition and its environment, including humidity, temperatures, and the environmental vegetation health.

Subjects Artificial Intelligence, Data Mining and Machine Learning, Spatial and Geographic Information Systems, Software Engineering
Keywords Vegetation indices, Machine learning, Remote sensing, Aridity

INTRODUCTION

Based on the study of disaster risk in Central Java Province, Indonesia, the number of people potentially exposed to drought disasters was estimated at 33.7 million people. Economic losses as a result of exposure to the drought were estimated at 65.2 billion rupiahs, and damage to the land area of 61 thousand hectares (*Deputi Bidang Pencegahan dan Kesiapsiagaan, 2016*). Drought is a disaster that occurs with a slow onset and can spread in wide areas, so that it has an impact on food security, the economy and social security (*Çelik & Gülersoy, 2018*). In most countries in the world, including Indonesia, drought is determined based on the rainfall data collected from meteorological stations (*Nam et al., 2018*). The method commonly used to calculate the amount of drought index is the Standardized Precipitation Index (SPI) method. The drought index is calculated by comparing the magnitude of deviations in the amount of rainfall from normal conditions or the long-term average in a particular region (*Bayissa et al., 2017*). Drought assessment using the SPI method and meteorological data has high accuracy and precision, especially in areas with uniform density and distribution of meteorological stations. However, in developing countries, the density and distribution of meteorological stations and limited operational costs make it difficult to obtain rainfall measurement data. The limitation of rainfall data becomes an obstacle in identifying the intensity, frequency and spatial distribution of drought (*Dutta et al., 2015a*). The SPI method is one of the 150 methods developed as drought indices. Apart from the SPI, the other common methods used in drought studies are the Standardized Precipitation Evapotranspiration Index (SPEI), the Palmer Drought Severity Index (PDSI), and the Precipitation Anomaly Percentage (*Wang et al., 2020*). The SPEI is the index used to take into account the drought damage based on the climate balance between precipitation and evapotranspiration. The timescale of SPEI ranges from 1-24 months and is interpolated using high resolutions in order to obtain vegetation responses toward potential water deficit over different timescales (*Gouveia et al., 2017*). The Palmer Drought Severity Index (PDSI) is an estimate of relative dryness in an area using temperature, precipitation, and soil characteristics. In general, the PDSI used to take into account of long-term drought which has affected an area for more several months (*Mika et al., 2005*). The PDSI works based on the water supply and water demand system while considering: (1) potential evapotranspiration calculated using Thornwaite method, (2) the amount of measured soil moisture, (3) the reduced amount of soil moisture due to evapotranspiration, (4) runoff. The Precipitation Anomaly Percentage (PAP) is precipitation anomaly percentage index which reflects hydrology factors deviating from the same average periods over years. The PAP index shows drought degrees on the same periods (*Zhengmin et al., 2014*).

In the late 20th or early 21st century, satellite imagery was widely used to monitor the risk of drought, both spatially and temporally. Satellite imagery data was chosen because it was available with high spatial and temporal resolution and can be obtained free of charge. The satellite imagery data will be extracted to get numerical data of vegetation indices, which are the main data to classify environmental conditions such as drought, moisture stress and temperature on the earth's surface. Vegetation indices are numerical values of

the results of spectral transformation of image-forming bands that are formulated to assess the photosynthetic activity of the vegetation canopy layer which correlates to plant types, water content in tissues, and photosynthetic activities. Vegetation Indices are mathematical combination of different spectral bands, and are designed to visualize various features of different images (*Karakacan Kuzucu & Bektas Balcik, 2017*). The vegetation index is dimensionless radiometric measurements acquired from linear combination of the red spectrum and NIR, which is calculated based on the light reflectance from satellites, or light reflectance from various objects on the earth surface (*Gherghina, Maftei & Filip, 2011*). Various aridity indices have been formulated by experts for classifying drought, including: (1) index for monitoring vegetation conditions as a result of drought (Vegetation Condition Index or VCI), (2) index for monitoring the surface temperature of the earth (Temperature Condition Index or TCI), (3) index for monitoring soil moisture (Microwave Integrated Drought Index or MIDI), (4) index for monitoring the combination of temperature and humidity on the earth's surface and their impact on vegetation health (Vegetation Health Index or VHI) and (5) index for monitoring the level of vegetation greenness (Normalized Difference Vegetation Index or NDVI) (*Elhag & Zhang, 2018*). The most suitable method used for the purpose of representation, categorization and classification of vegetation indices is machine learning (*Maxwell, Warner & Fang, 2018*). In essence, the machine learning algorithm shows how a machine compiles and learns to create valid information from observational and historical data. Machine learning is composed of a group of algorithms including neural networks (NN), support vector machines (SVM), self-organizing maps (SOM), decision trees (Dtree), random forests (RF), case-based reasoning (CBR), k-nearest neighbours (k-nn) and genetic programming (GA) (*Lary et al., 2015*).

In this study, we develop a software framework for predicting drought and land fires risk areas based on spatial prediction of vegetation indices. The developed framework is a combination of machine learning algorithms that include ANN, SVM, RF, k-nn, CART and MARS with an algorithm compiled from the interpolation of Inverse Distance Weight. The use of machine learning in this framework aims to classify and predict drought indices from vegetation indices data that have spatial and temporal resolutions. The use spatial interpolation is an approach to determine numerical values in non-sampling areas. The software framework names are Satellite Imagery based Machine learning for the Index of Aridity (SIMIA). 'Background' in this paper discusses the proposed theoretical background in the framework, 'Materials & Methods' discusses the research methods, 'Results' discusses the results and discussions, 'Conclusions' discusses the conclusions, and the last section is the references.

BACKGROUND

ML (Machine Learning) is a subfield of artificial intelligence in which computers learn the relationship of a data set through a training process. The ML explores new relations of the location, date and other qualities data in a large data set based on the experiences from the historical data acquired (*Huntingford et al., 2019*; *Fang et al, 2020*). The ML can

classify high dimensional data effectively and efficiently, map class of data which have very complex characteristics as in remote sensing (*Zaraza Aguilera, 2020*). In the field of remote sensing, the machine learning technique has been used to calculate identification and pattern recognition from a set of historical climate data and topography to produce information on land coverage changes. CART is an algorithm in machine learning that works based on the Decision Tree (DT). The core principle of this algorithm classifies a number of object data by divide them into both discrete binary data and continuous binary data as the predictable variables. The feature-selection of the classified objects used Gini Index (1), in which TS is the training data, $f(K_i, TS)$ is the probability of data corresponds *to the K class* (*Fang et al, 2020*).

$$G = \sum \sum_{j \neq i} \left(\frac{f(K_i, TS)}{|TS|} \right) \left(\frac{f(K_j, TS)}{|TS|} \right) \tag{1}$$

$$E = \sum_{i=1}^{c} -p_i * log_2(p_i) \tag{2}$$

E is an entropy calculated using the Gini Index using the Gini Index logarithmic. The concept of Entropy in the CART algorithm classifies a set of data using the characteristics that are in the data class (training data) (*Kamal et al., 2000*). The characteristic of this algorithm is when the Gini Index increases, the objects heterogeneity in the class is going to rise. Otherwise, when the Gini Index is decreasing, the object homogeneity in the class is increasing. CART is one of machine learning algorithms used for prediction using multidimensional Cartesian space with a function: $X = f(Y) + \epsilon$ where $Y \epsilon \Re^d$, $X \epsilon \Re^d$. CART prediction is calculated by the function $\widehat{X} = \hat{f}(\hat{Y})$. The notation ϵ represents various intrinsic data condition or diversity of data variables that are unpredictable or not observed. The disturbance in the prediction results is determined by a factor (Y) (*Wang et al., 2004*).

RF is a non-parametric ensemble classification and machine learning regression that uses DT as the basis for classification of data sets. It means that the mean prediction of multiple decision trees is trained using data obtained from different subsets but from the same data (*Li et al., 2017*). RF is an object classification method consisted of a group of tree-structured classifier, or it is denoted as $\{h(y, \Theta_s) k = 1, 2, \ldots\}$ in which $\{\Theta_s\}$ is the random vector distributed in each classifier node that is denoted as the y-class (*Kullarni & Sinha, 2013*). The RF equation is described as follow (3):

$$T_s = T_r(s|p_1, p_2, \ldots, p_n) = \frac{\Pi_i^T T_r(p_1|s) T_r(s)}{T_r(p_1, p_2, \ldots, p_n)} \tag{3}$$

notation of $T_r(p_i|s)$ is the probability of input data on the s class. ANN is composed of a group of interconnected nodes and is called a neuron or neuronode. Interconnection between neuronodes can be described as the interconnection of a large number of simple processors arranged in layers (*Raczko & Zagajewski, 2017*; *Akhand, Nizamuddin*

& Roytman, 2018). The equation used is (4):

$$\gamma^k(t_s) = f_0 \left[\sum_{j=1}^{m} w_{kj} \cdot f_n \left(\sum_{i=1}^{n} w_{ji} x_i(t_s) + (wj_0) \right) + wk_0 \right] \qquad (4)$$

Where n is the number of data input variables, m is the number of neurons, $x_i(t)$ is the ith of input variable at time ts, w_{ji} is the weight of the ith of variable neuron which is interconnected with the jth neuron, wj_0 is biased for neuron at j, fn is a function of neuron activation, wkj is the variable weight of neuron at j and neuron at k, wk_0 is biased for external neuron k, f_0 is an activation function for external neuron, and $\gamma^k(t_s)$ is the external neuron k which is predicted at time t_s (*Belayneh & Adamowski, 2013*).

SVM is a method that has been implemented to classify tree types of hyperspectral data. SVM works by using training data to find the most optimal hyperplane of attribute classes. It can be simply described that the optimal hyperplane is determined from the margin value between the two training data classes. If a hyperplane cannot be determined, a kernel function is used to convert the original data into a higher dimension.

The k-nn algorithm is a non-parametric method that is widely used for classification in pattern recognition. The main principle of k-nn is that the category of data point is determined according to the classification of neighbour k closest to that point. As an illustration, an arbitrary point which is a subset in the training set $S = \{(p_1, q_1), (p_2, q_2), \dots (p_N, q_N)\}$ with the number of attributes or entities N. The value of $p_i \in R^d$ is a vector feature and $q_i \in y = a_1, a_2 \dots a_M$ as the classification label, $i = 1, 2, \dots, N$. If there is input data p then the value of k-nn or $N_k(p)$ is the distance of the training set entities (*Guo et al., 2018*). To calculate k-nn, Euclidean Distance should be determined first using (5):

$$d(p, q) = \sqrt{\sum_{i=1}^{n} (p_1 - q_1)^2} \qquad (5)$$

The notation d is used for the Euclidean distance, the notation p is used for the training data set, and the notation for the testing data set is q. The distance between the training data was compared in order to determine the closest distance.

MARS is a non-parametric multivariate regression analysis that uses a linear regression approach one by one for each variable, whereas if the variable is non-linear then it is approximated by using the slope interval of the variable. MARS analyses sequentially and looks for interactions between variables (*Yang et al., 2016*). The MARS equation is (6):

$$f(x) = \sum_{i=1}^{k} c_i b_i(x) \qquad (6)$$

Which b_i is the basis of the linear function, and c_i is the constant coefficient.

VI is a quantitative value of the measurement of vegetation canopy in receiving and reflecting the spectrum of light and interpreted as a spectral characteristic of vegetation (*Paseleng, Joko Prasetyo & Hartomo, 2018*). Formulation VI that is used for drought monitoring is classified into four groups name (*Du et al., 2018*; *Wang et al., 2018*): (1) Greenness Related Vegetation Indices that includes NDVI and VCI ; (2) Temperature

Prasetyo et al. (2021), *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.415

5/21

PeerJ Computer Science

Table 1 The VI in the proposed framework.

VI	Equation	Description	Ref.
NDVI	$\frac{\rho_{NIR}-\rho_{red}}{\rho_{NIR}+\rho_{red}}$	NDVI is a vegetation index that uses a combination of NIR and RED spectrum reflectance to measure vegetation greenness. Chlorophyll and carotenoid elements of a plant absorb more red spectrum with λ of 0.66 μm, and ρ_{NIR} with λ of 0.86 μm. NDVI represents photosynthesis, biomass, dominant species and biogeochemical processes.	*Hashim, Abd Latif & Adnan (2019)*
VCI	$\frac{NDVI_a-NDVI_{min}}{NDVI_{max}+NDVI_{min}} \times 100$	VCI is used as an indicator to identify changes in vegetation from bad condition to good condition. NDVImin and NDVImax are the mean values of seasonal NDVI. NDVI represents the dynamics of seasonal vegetation, while VCI reflects relative changes in humidity conditions from very bad to optimal condition.	*Garnero & Godone (2013)*
TCI	$\frac{LST_{max}-LST_a}{LST_{max}+LST_{min}} \times 100$	TCI is an indicator of drought that gives an indication of stress on vegetation caused by high temperature and excessive humidity. LST_{max} or Land Surface Temperature maximum is the highest average temperature of the current month and LST_{min} is the lowest average temperature of the current month.	*Angka (2019)*
VHI	$\alpha VCI + (1-\alpha) TCI$	VHI is a vegetation index which is a combination of VCI and TCI. VCI represents humidity in the vegetation canopy which absorbs visible light and NIR spectra. TCI represents the air temperature in the vegetation canopy which absorbs the IR spectrum. VHI is a vegetation index that represents the level of air temperature (TCI) and soil moisture (VCI). A lower VHI value indicates a high drought while a higher VHI value indicates a wet or no drought condition.	*Shi et al. (2020)*
NBR	$\frac{\rho_{NIR}-\rho_{SWIR}}{\rho_{NIR}+\rho_{SWIR}}$	Normalized Burn Ratio (NBR) is the ratio between NIR and SWIR with both λ at 2.08–2.35 μm. This index is used to identify a burnt area after a fire incident and provide a quantitative measure of the severity of the burned area.	*Verdin et al. (1999)*
BAI	$\frac{1}{(\rho_{cr}-\rho_r)^2(\rho_{cnir}-\rho_{nir})^2}$	BAI is a special index to distinguish areas affected by fire. This index is calculated from the spectral distance from each pixel to the reference spectral point, where the newly burned area tends to coalesce.	*Tripathi et al. (2013)*

Related Vegetation Indices such as the TCI ; (3) Combined Vegetation Indices, such as the VHI ; (4) Burn Severity Indices, such as NBR, and BAI, CSI. The VI in the proposed framework are as presented in Table 1.

The class of vegetation status of NDVI in LANDSAT 8 OLI are as presented in Table 2A (*Hashim, Abd Latif & Adnan, 2019*).

The IDW method is used to determine the non-observation point value, which is calculated based on the weight and distance of the closest observation point. Determination of IDW is done using (7) (*Garnero & Godone, 2013*).

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

The notation Z is the value of observation point that will be calculated or sought, the notation Z_i is the value of observation point that is known based on the calculation, and

Table 2 The class vegetation status of NDVI (A) and class of indicator VHI, TCI, VCI (B) (*Hashim, Abd Latif & Adnan, 2019*).

(A)

Vegetation Class	Description	NDVI Value
Non Vegetation	Barren areas, build up area, road network	−1 to 0.199
Low Vegetation	Shrub and grassland	0.2 to 0.5
High Vegetation	Temperate and Tropical urban forest	0.501 to 1.0

(B)

Vegetation Status	VHI
Extreme Drought	<10
Severe Drought	10–20
Moderate Drought	20–30
Mild Drought	30–40
No Drought	>40

Table 3 Band name, bandwidth and resolution of LANDSAT 8 OLI.

Band	Bandwidth	Resolution
Band 1 Coastal	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	0.85–0.88	30
Band 6 SWIR 1	1.57–1. 65	30
Band 7 SWIR 2	2.11–2.29	30
Band 8 Pan	0.50–0.68	15
Band 9 Cirus	1.36–1.68	30
Band 10 TIRS 1	10.6–10. 19	100
Band 11 TIRS 2	11.5–12.51	100

the notation β is the weight factor and δ is the smoothing factor. The value of h_{ij} is the distance between the known and unknown sample points and is determined using the Euclidean equation as shown below (8).

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

where Δx and Δy are the distance between unknown point at j and sample one at i.

MATERIALS & METHODS

This study uses monthly LANDSAT 8 OLI (Operational Land Imager) imagery data from 2018 to 2019. The data source is the United States Geological Survey (USGS) https://earthexplorer.usgs.gov/ on path/row 120/65. LANDSAT 8 OLI imagery data consists of 11 bands with different bandwidth and resolution (Table 3).

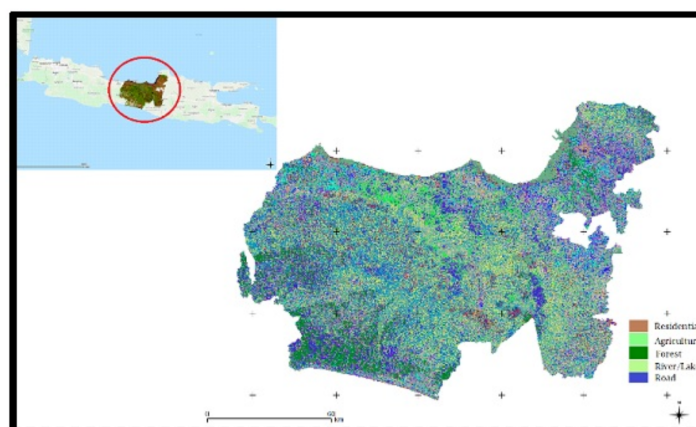**Figure 1** **The land cover study area of Central Java Province of Indonesia.**

Full-size ⬚ DOI: 10.7717/peerjcs.415/fig-1

This study was conducted in Central Java Province, Indonesia covering 346 sub-districts. Geographically the study area is in the coordinates of 5°40′ and 8°30′ South Latitude and 108°30′ and 111°30′. The air temperature at the study area was between 23.5 °C to 27.8 °C, the humidity was between 66% and 83% and the highest rainfall was 3,143 mm (*Badan Pusat Statistik, 2019*). The proposed framework runs in stages as shown in Fig. 1 below. Land cover in the study area varies but is dominated by residential areas, public and industrial facilities, Agriculture includes rice fields, fields and plantations, Forest, River and Road.

The proposed framework runs in stages as shown in Fig. 2 below. (1) The LANDSAT 8 OLI image is composed of 11 bands in which each band has different wavelength and function (Table 3). The LANDSAT 8 OLI images obtained from https://earthexplorer.usgs.gov/. (2) The process of geometric correction, radiometer correction and atmospheric correction. (3) Combining bands and calculating use mathematical formulations as in (4)–(10) to form vegetation indices. The generated vegetation indices data with numeric data types. (4) Correlation analysis and temporal regression of vegetation index data were performed to see the correlation and trend of vegetation indices at the same location from time to time (monthly). Each vegetation data indices are divided into two categories i.e., testing data at 70% and learning data at 30% of the total data. (5) Classifying and predicting vegetation indices data using ANN, SVM, RF, k-nn, CART and MARS algorithms. Choose the most optimal algorithm for analysing vegetation index data. (6) Modelling of drought risk areas, forest fire risk areas and spatial connectivity between areas with high risk of drought and land fires using the Getis Ord method. Spatial prediction is done to determine the risk value of drought and land fires in non-sampled areas. (7) Testing of the interpretation of machine learning prediction results imagery with real conditions in the field using the Confusion Matrix method.

The eminences of this framework are: implementing ML to predict a period of the vegetation index value as a data base to determine the drought risk area, and generating information about the drought risk area in the smallest area using the IDW spatial
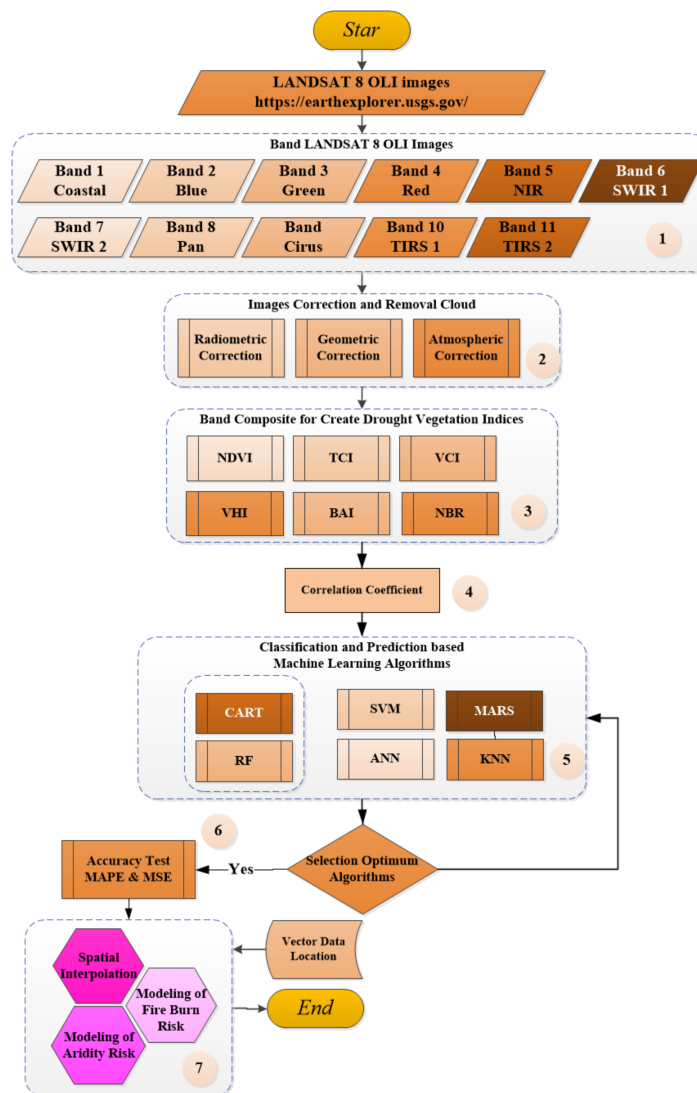
Figure 2 The proposed framework in the study.

interpolation method. This interpolation method provides a function to spatially predict areas which are not the observation points.

RESULTS

The correlation analysis of VI data aims to see the distribution, correlation and direction of the VI data trend in a certain temporal value. Correlation analysis of the observed VI was done using BAI, NBR, NDVI, TCI, VCI and VHI in 2018 and 2019 which was visualized in the form of scatterplot matrices.

The analysis using Pearson Correlation Coefficient method shows that the association between vegetation indices is very strong as indicated by the magnitude of the coefficient r, which is <0.5 or <-0.5, and which some with values >0.3 or >-0.3 are considered
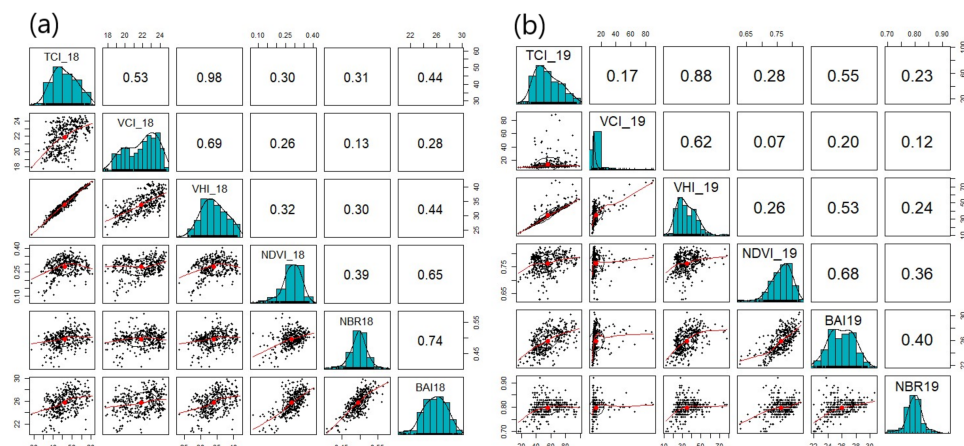
Figure 3 Analysis of the Vegetation Indices Association in 2018 (A) and 2019 (B).

Full-size ⬛ DOI: 10.7717/peerjcs.415/fig-3

as intermediate association and some with values <0.3 or <−0.3 are considered as weak association (Figs. 3A and 3B).

On all VI (Figs. 3A and 3B), line curves represent data distributions (black dots), data centers (red dots) and data trends (red lines) that describe correlation directions i.e., linear and linear curve. Bar chart for each index shows that data is normally distributed in each sample space. VI that correlates with the Greenness Related Vegetation Indices class shows a linear correlation–an increase in the value of an index will be followed by other indices, while VI that correlates with Burn Severity Indices shows a linear curve correlation, which means an increase in the value of an index will be followed by other indices. However, up to certain point, it will tend to decrease. Meanwhile, VHI is formed by VCI and TCI data, so that it has a moderate to strong correlation. The trend of VHI and VCI and TCI shows a linear pattern. Linear curve pattern is observed between VI in the Burn Severity Indices class and other vi. The difference in correlation analysis results on the analyzed VI indicates that each VI has a different sensitivity towards noise (background canopy and atmospheric effect) at the same location.

The VI, which is the drought indicator, is a combination of NDVI (vegetation stress), VCI (time, duration, and drought intensity), TCI (vegetation stress triggered by temperature), and VHI (vegetation stress caused by wetness) (*Jayawardhana & Chathurange, 2020*). The VI is used as a drought-risk indicator as it has strong correlation with precipitation, vegetation canopy greenness, canopy and soil water content, also land covers (*Shi et al., 2020*). The *VI* prediction is done by using machine learning, especially on RF and CART algorithms and comparison is made with ANN, k-nn, SVM, and MARS. Results of *VI* prediction are useful for identifying the risk of drought and land fires in an area. The *VI* prediction is done by using machine learning, especially on RF and CART algorithms and comparison is made with ANN, k-nn, SVM, and MARS.

The results of the RF accuracy analysis and test are shown in Table 4A. The validation testing of prediction results using MAPE shows a low average difference between the predicted results from survey data at 1.04% in all indices. The test using MSE aims to see

Prasetyo et al. (2021), *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.415

10/21

Table 4 Prediction of vegetation index using the RF algorithm and the CART algorithm in 2019–2020.

VI	VI$_{Pred}$ Min	Max	VI$_{Survey}$ Min	Max	MAPE	MSE
(A)						
NDVI	0.389	0.597	0.049	0.405	0.500	0.071
TCI	22.359	78.912	22.359	78.912	0.000	0.000
VCI	13.678	43.638	10.023	57.037	23.492	23.736
VHI	21.375	52.375	17.740	59.987	14.121	3.954
BAI	21.261	30.193	21.215	30.210	0.056	0.000
NBR	0.564	0.716	0.552	0.749	4.406	0.000
(B)						
NDVI	0.426	0.583	0.246	0.583	0.000	0.000
TCI	30.387	71.577	84.639	71.557	0.028	0.000
VCI	14.010	37.593	14.010	38.000	0.003	0.166
VHI	22.503	51.655	22.503	52.000	0.019	0.119
BAI	21.938	28.755	21.938	29.000	0.003	0.060
NBR	0.557	0.706	0.577	0.658	7.295	0.002

the level of similarity between the prediction results with survey data, which the smaller and near zero shows the higher similarity. The average RMSE in this study is 0.05.

The prediction results of vegetation index using the CART algorithm shown in Table 4B. Testing the validation of prediction results using MAPE shows the average difference between the predicted results with survey data is very low at 1.05% in all indices. The average RMSE in this study is 0.05. The result of this experiment indicates that CART algorithm is able to accurately predict vegetation index data.

Validation testing of prediction results using MAPE shows that the average difference between the prediction results and survey data is very low at 5.78% for all indices. The average RMSE in this study is 2.19. The results of this experiment indicate that the ANN algorithm can accurately predict vegetation index data.

The results of vegetation index prediction using k-nn are shown in Table 5B. The validation testing of prediction results using MAPE shows that the average difference between the prediction results and survey data is very low at 6.07% in all indices. The average RMSE in this study is 2.37. The results of this experiment indicate that the k-nn algorithm can accurately predict vegetation index data. The SVM algorithm works by classifying input data to form a multidimensional hyperplane. Hyperplane multidimensional distinguishes vegetation index variable classes namely NDVI, TCI, VCI, VHI, BAI and NBR.

The SVM kernel uses the difference between values to differentiate classes in multidimensional data. For example, if there is a TCI data class, two dimensional spaces will be formed, separated by a hyperplane as a straight line. In vegetation index classes, linear hyperplanes are not easy to form because there are some values that are difficult to

Table 5 (A) Prediction of vegetation index 2019–2020 using the ANN algorithm and the (B) knn algorithm.

VI	VI$_{Pred}$		VI$_{Survey}$		MAPE	MSE
	Min	Max	Min	Max		
(A)						
NDVI	0.331	0.602	0.348	0.615	2.114	0.000
TCI	27.465	76.772	22.359	79.531	3.469	7.612
VCI	13.210	55.443	13.032	57.037	2.795	2.541
VHI	19.801	40.861	17.740	41.711	11.893	0.723
BAI	20.979	30.492	21.214	30.210	0.933	0.080
NBR	0.555	0.695	0.552	0.749	7.210	0.003
(B)						
NDVI	0.396	0.613	0.348	0.615	0.325	0.002
TCI	26.217	77.847	22.359	79.351	1.895	14.884
VCI	13.098	41.098	13.023	57.036	27.944	0.006
VHI	18.591	53.441	17.740	59.591	10.320	0.724
BAI	21.289	30.012	21.220	30.012	0.000	0.005
NBR	0.555	0.734	0.552	0.749	2.003	0.002

Table 6 Prediction of vegetation index using SVM Algorithm (A) and MARS Algorithm (B) in 2019–2020.

VI	VI$_{Pred}$		VI$_{Survey}$		MAPE	MSE
	Min	Max	Min	Max		
(A)						
NDVI	0.351	0.617	0.040	0.504	22.421	0.045
TCI	23.569	78.495	22.358	79.351	1.079	0.732
VCI	13.276	56.479	13.023	57.036	0.977	0.310
VHI	18.701	59.289	17.740	55.711	16.915	12.802
BAI	21.390	29.835	21.213	30.210	1.241	0.140
NBR	0.552	0.741	0.552	0.749	1.068	0.003
(B)						
NDVI	0.359	0.615	0.049	0.405	21.782	0.044
TCI	30.387	71.557	22.359	75.912	9.321	18.966
VCI	13.023	57.037	13.023	57.037	0.000	0.000
VHI	17.740	59.987	17.740	59.987	0.000	0.000
BAI	21.261	30.193	21.215	30.210	0.056	0.000
NBR	0.551	0.726	0.552	0.749	3.071	0.001

include in these classes. The solution is to add a dimension using polynomial, sigmoid, and radial basis kernel functions (Table 6A).

Validation testing of prediction results using MAPE shows that the average difference between the prediction results and survey data is very low at 6.28% in all indices. Tests using MSE aims to see the level of similarity between the results of predictions and survey data, the smaller and near zero, the higher similarity. The average RMSE in this study is

Table 7   Average MAPE and MSE on the Machine Learning Algorithm Test.

Algorithms	MAPE (%)	MSE
ANN	5.70	2.19
k-nn	6.07	2.07
SVM	6.28	2.01
RF	1.04	0.05
CART	1.05	0.05
MARS	5.34	4.15

2.01. The results of this experiment indicate that the SVM algorithm can accurately predict vegetation index data.

The prediction results using MARS algorithm can be seen in Table 6B. Validation testing of prediction results using MAPE shows that the average difference between the predicted results and survey data is very low at 5.34% in all indices. Tests using MSE aims to see the level of similarity between the results of predictions and survey data, the smaller and near zero, the higher similarity. The average RMSE in this study is 4.15. The results of this experiment indicate that the MARS algorithm can accurately predict vegetation index data.

The experimental results show that the most efficient algorithms in predicting vegetation index are RF and CART. RF has a difference between the predicted results and 1.04% survey data (MAPE) and the smallest value close to zero is 0.05 (MSE). CART has a difference between the predicted results and 1.05% survey data (MAPE) and the smallest value close to zero is 0.05 (MSE) (Table 7).

Based on the results of the machine learning algorithm validation test, it can be seen that the most effective and efficient methods are RF and CART (Table 7). The next discussion presents an interpolation of the value of vegetation index and its status at a non-sample point of special survey on RF and CART. Figure 4 shows the spatial prediction of NDVI in 2019 to 2020 with IDW method. The experimental results show that in 2019, most regions in Central Java have high vegetation land coverage with NDVI >0.21. High NDVI is an indicator of high chlorophyll capacity contained in mesophyll canopy leaves. In contrast, in 2020, it is revealed that most regions in Central Java have low vegetation land coverage with NDVI <0.21. The low NDVI value is an indicator of vegetative pressure which has an impact on reducing chlorophyll content. Based on the spatially prediction, the NDVI dynamics between 2018 and 2019, can be inferred that the green vegetation biomass, LAE, photosynthesis activities, chlorophylls density in the vegetation canopy, and the amount of annual and monthly rainfall in 2020 are very low (*Verdin et al., 1999*). On Fig. 4A, it can be seen that the area which has low level of greenness is very wide, with NDVI <0.21, which means that there is stress pressure in the vegetation canopy as the result of low precipitation level, reduced water content in the canopy, and low water level in the soil. The prediction results show that the drought risk consistently happened in the most of Central Java area,
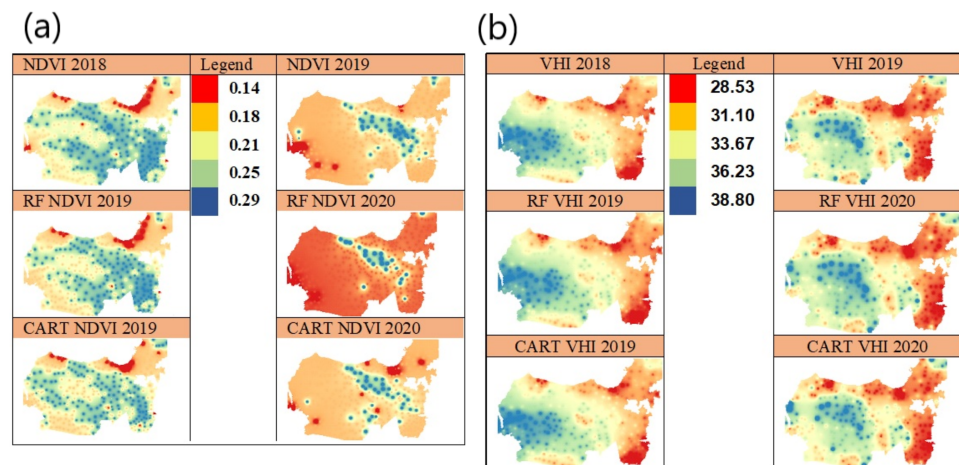
Figure 4 Spatial prediction using IDW in 2019–2020 of NDVI vegetation index (A) and spatial prediction of VHI vegetation index (B).

Full-size 🖼 DOI: 10.7717/peerjcs.415/fig-4

and, in the contrary, the areas that did not experience drought are also consistent since 2018.

VHI consists of VCI, TCI, and α components which are the parameters to quantify VCI and TCI contributions. VCI is composed from NDVI value, while the TCI is composed of LST value, which all of them are calculated using monthly data time series period between 2018 and 2019.Vegetation land coverage can be seen from the dominance of regions that do not indicate drought with a VHI value >33.67 in 2019. In 2020 shows a decrease in the areas that have land coverage with thick vegetation and an increase in the areas which indicate drought shown by a VHI value <31.10 (Figs. 4B). Figure 4B shows that the VHI prediction from 2019-2020 is very low, which means that the vegetation in the most of study areas contain chlorophyll, the low leaf equivalent water thickness and LEA, mainly caused by low precipitation, and this condition indicates drought biophysically (*Tripathi et al., 2013*). VHI is composed of two components, namely, TCI index which is a thermal indicator and VCI index which is an indicator of soil surface moisture. TCI value describes the thermal surface of a ground caused by solar radiation reflectance. The TCI value is calculated based on the land surface temperature (LST) analyzed from the time series imagery pixel (*Karamzadeh et al., 2017*). The low TCI value is an indicator of an increased stress and vegetative growth pressure associated with drought phenomena (*Gidey et al., 2018*). Figure 5A shows that TCI can be accurately predicted using RF and CART algorithms and visualized spatially with IDW.

VCI is an indicator of drought level based on soil surface humidity. The VCI >50% represents normal conditions; between 35%–50% represents moderate drought and <35% represents high drought (Fig. 5B (*Dutta et al., 2015*)).

VCI that is measured in seasonal or annual terms in time series showing a value <20% indicates the occurrence of long drought. RF and CART algorithms can be used to accurately predict TCI using IDW function as shown in Fig. 5B. Figure 5B shows the potential for
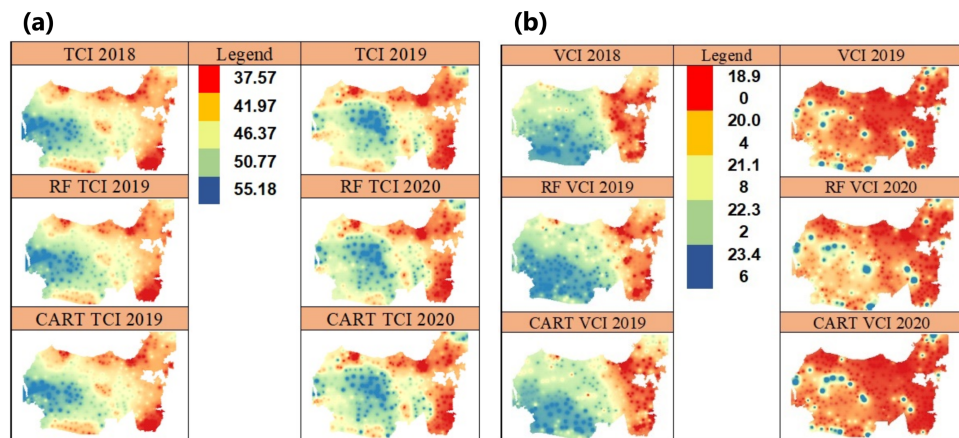
Figure 5 Spatial prediction in 2019–2020 using the IDW of TCI vegetation index (A) and the VCI vegetation index (B)

Full-size ☒ DOI: 10.7717/peerjcs.415/fig-5

high to extreme drought i.e., <20.04. In this study, three indicators are used to see the areas before and after land fires as a result of long drought using BAI and NBR. Fire detection can be done through pixel extraction by identifying R, NIR and SWIR signals. Signal R can show contrasting color patterns and separate pixels that represent healthy vegetation and non-vegetation lands based on chlorophyll concentrations in plants. The newly burnt area reflects very low NIR and R signals and reflects a very high SWIR signal.

BAI is a vegetation index that has a threshold value which is calculated based on pixel comparison on time series that represents burnt and unburned areas (Fig. 6A) (*Dalposso et al., 2013*; *Chuvieco et al., 2013*). BAI is the latest method which has a high sensitivity towards the influence of soil types and the existing land use on the surface of the earth (Table 8A).

The computational experiment shows that the study area has a ΔBAI value of between −3.971 to −2.376, which is interpreted as a region where there are no land fires. This region shows the growth of new vegetation. Prediction using machine learning shows the algorithm that produces the closest value to ΔNBR1819 survey data which is ΔSVM and ΔCART algorithms (Table 8B). NBR is a vegetation index that is used to detect fires based on spectral changes which are interpreted as changes in the amount of chlorophyll content in plants, moisture, and charcoal or ash in the soil (*Parks, Dillon & Miller, 2014*).

The ΔNBR is VI which shows energy differences between healthy vegetation areas and stressed vegetation areas caused by water stress as a result of fire and is used to determine the potential of regrowth. The severity degree of the burned areas is calculated based on the differences between before the fire and post-fire vegetation in the same area. ΔNBR aims to normalize spatial variation in vegetation cover before the fire as this method is very accurate in other studies (*Wulder et al., 2009*). The severity degree is notated with *differenced (d)* or Δ or *delta*. The result of Δ calculation shows that the severity level, vegetation survival, and delayed vegetation mortality. The NBR value is between +1 and −1, and the higher the value shows the burned areas. The experimental results show that
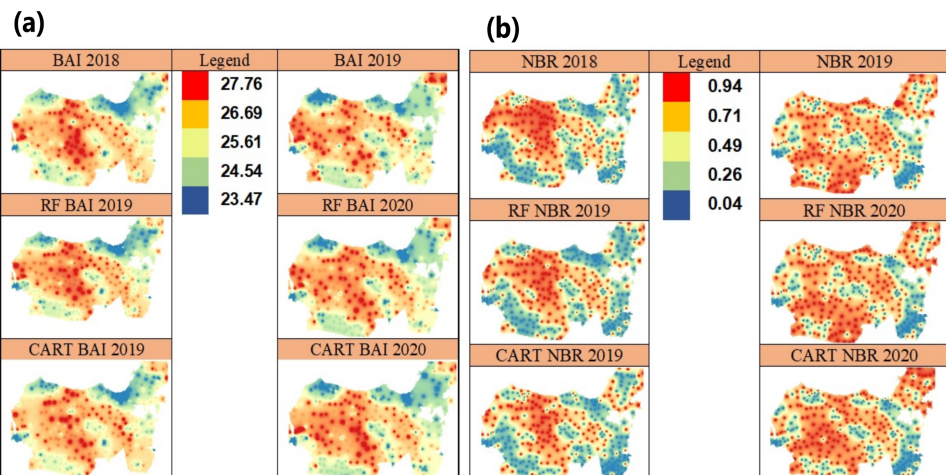
Figure 6 Spatial prediction of the BAI vegetation index (A) and the NBR vegetation index in 2019–2020 using IDW (B).

Full-size 🖼 DOI: 10.7717/peerjcs.415/fig-6

Table 8 (A) Prediction of BAI values with Machine Learning algorithm and (B) ∆NBR values with Machine Learning algorithm.

(A)

Algorithms	Min	Average	Max
∆ANN	−11.787	0.074	3.845
∆k-nn	−6.037	−0.016	6.742
∆SVM	−3.617	−0.242	2.506
∆RF	−2.647	−0.272	5.208
∆CART	−3.067	−0.283	2.376
∆MARS	−6.923	−0.283	5.028
∆BAI1819	−3.971	−0.216	2.376

(B)

∆ANN	−0.426	−0.362	−0.204
∆k-nn	−0.402	−0.302	−0.212
∆SVM	−0.410	−0.301	−0.210
∆RF	−0.381	-0302	−0.237
∆CART	−0.410	−0.303	−0.244
∆MARS	−0.397	−0.302	−0.236
∆NBR1819	−0.412	−0.303	−0.208

the study area has a ∆NBR value of between −0.020 to −0.412 which is interpreted as a region that has just begun to grow. Prediction using machine learning shows two ∆SVM and ∆CART algorithms that produce the closest value to ∆NBR1819 survey data (Fig. 6B), (Table 8B). The ∆NBR which is calculated from 2018 to 2019 shows a typical pattern of no fire and pattern of lands that start to be vegetating which can reflect signal and is depicted on image pixel.

Prasetyo et al. (2021), *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.415

16/21

CONCLUSIONS

The correlation coefficient analysis concerning the results of vegetation index observation namely, BAI, NBR, NDVI, TCI, VCI and VHI in 2018 and 2019 shows that the association between vegetation indices is very strong, with coefficient r is <0.5 or <-0.5, some of which have values >0.3 or >-0.3 as medium association and some have values <0.3 or <-0.3 as weak association. The experimental results show that the most efficient algorithms in predicting vegetation index are RF and CART. RF has a difference between the predicted results and 1.04% survey data (MAPE) and the smallest value close to zero is 0.05 (MSE). CART has a difference between the predicted results and 1.05% survey data (MAPE) and the smallest value close to zero is 0.05 (MSE). Spatial prediction shows that the vegetation index of machine learning prediction result data is almost the same size as the data from field observations. NDVI indicates high vegetation land coverage with NDVI >0.21, whereas in 2020 data show that most areas indicate low vegetation land coverage with NDVI <0.21. Data in 2019 show that VHI >33.67 which indicates that there is no drought, whereas in 2020 there is a decrease of areas with thick vegetated land coverage and an increase in areas indicating drought indicated by a VHI value <31.10. VCI values in the study area varied which VCI $>50\%$ shows a normal condition, VCI between 35%–50% shows a moderate level of drought and VCI $<35\%$ shows a high level of drought. Experiments using machine learning show that dBAI values in the study area are between -3.971 up to -2.376 which indicate that there is no land fire and there is new vegetation growth. DNBR values are between $-0, 208$ to -0.412 which indicate that vegetation starts to grow in the area. Prediction using machine learning shows that the ΔSVM and ΔCART algorithms produce the closest values to the ΔNBR observation data in 2018 and 2019. In future research, we will compare the performance of the framework between LANDSAT 8 OLI image data and Sentinel 2 image data. We will also compare the drought risk areas from vegetation indices analysis with meteorological methods such as SPI, SPEI and PDSI. The results of this comparison can be selected methodology that has high validation and accuracy in the software framework.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Sri Yulianto Joko Prasetyo conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, authored or reviewed drafts of the paper, machine learning algorithms, and approved the final draft.
- Kristoko Dwi Hartomo conceived and designed the experiments, performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Mila Chrismawati Paseleng performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Data and code are available in the Supplementary Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.415#supplemental-information.

REFERENCES

Akhand K, Nizamuddin M, Roytman L. 2018. An artificial neural network-based model for predicting boro rice yield in bangladesh using avhrr-based satellite data. 16–25 DOI 10.5923/j.ijaf.20180801.04.

Badan Pusat Statistik. 2019. Provinsi Jawa Tengah Dalam Angka 2019. p. 87. *Available at* https://jateng.bps.go.id/publication/2019/08/16/fcb9efa7796cdbc491325688/provinsi-jawa-tengah-dalam-angka-2019.html.

Bayissa Y, Tadesse T, Demisse G, Shiferaw A. 2017. Evaluation of satellite-based rainfall estimates and application to monitor meteorological drought for the Upper Blue Nile Basin, Ethiopia. *Remote Sensing* **9(7)**:1–17 DOI 10.3390/rs9070669.

Belayneh A, Adamowski J. 2013. Drought forecasting using new machine learning methods. *Journal of Water and Land Development* **18(9)**:3–12.

Çelik MA, Gülersoy AE. 2018. Climate classification and drought analysis of Mersin. *Manisa Celal Bayar Üniversitesi Sos. Bilim. Derg.* **16(1)**:1–26 DOI 10.18026/CBAYARSOS.411475.

Chuvieco E, Opazo S, Sione W, Valle Hd, Anaya J, Bella CD, Cruz I, Manzo L, López G, Mari N, González-Alonso F, Morelli F, Setzer A, Csiszar I, Kanpandegi JA, Bastarrika A, Libonati R. 2013. Global burned-land estimation in Latin America using modis composite data. *Ecological Applications* **18(1)**:64–79 DOI 10.1890/06-2148.1.

Dalposso GH, Uribe-Opazo MA, Mercante E, Lamparelli RAC. 2013. Spatial autocorrelation of ndvi and gvi indices derived from landsat/tm images for soybean crops in the western of the state of Paraná in 2004/2005 crop season. *Engenharia Agrícola* **33(3)**:525–537 DOI 10.1590/S0100-69162013000300009.

Deputi Bidang Pencegahan dan Kesiapsiagaan. 2016. Kajian risiko bencana jawa tengah 2016–2020. *Available at* https://ppid.bpbd.jatengprov.go.id/naskah-akademis-dan-kajian/.

Du TLT, Du Bui D, Nguyen MD, Lee H. 2018. Satellite-based, multi-indices for evaluation of agricultural droughts in a highly dynamic tropical catchment, Central Vietnam, Water (Switzerland). *Water* **10(5)**:659 DOI 10.3390/w10050659.

Dutta D, Kundu A, Patel NR, Saha SK, Siddiqui AR. 2015. Assessment of agricultural drought in Rajasthan (India) using remote sensing derived Vegetation Condition Index (VCI) and Standardized Precipitation Index (SPI). *Egyptian Journal of Remote Sensing and Space Science* **18(1)**:53–63 DOI 10.1016/j.ejrs.2015.03.006.

Elhag KM, Zhang W. 2018. Monitoring and assessment of drought focused on its impact on sorghum yield over sudan by using meteorological drought indices for the period 2001–2011. *Remote Sensing* **10(8)**:1–21 DOI 10.3390/rs10081231.

Fang P, Zhang X, Wei P, Wang Y, Zhang H, Liu F, Zhao J. 2020. The classification performance and mechanism of machine learning algorithms in winter wheat mapping using Sentinel-2 10 m resolution imagery. *Applied Sciences* **10(15)**:5075 DOI 10.3390/app10155075.

Garnero G, Godone D. 2013. Comparisons between different interpolation techniques. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives)* **40(5W3)**:139–144 DOI 10.5194/isprsarchives-XL-5-W3-139-2013.

Gherghina CS, Maftei C, Filip C. 2011. Assessment of multi-spectral vegetation indices using remote sensing and grid computing. *International Journal of Computer Mathematics* **5(4)**:468–475.

Gidey E, Dikinya O, Sebego R, Segosebe E, Zenebe A. 2018. Using Drought indices to model the statistical relationships between meteorological and agricultural drought in Raya and its environs, Northern Ethiopia. *Earth Systems and Environment* **2(2)**:265–279 DOI 10.1007/s41748-018-0055-9.

Gouveia CM, Trigo RM, Beguería S, Vicente-Serrano SM. 2017. Drought impacts on vegetation activity in the Mediterranean region: an assessment using remote sensing data and multi-scale drought indicators. *Global and Planetary Change* **151**:15–27 DOI 10.1016/j.gloplacha.2016.06.011.

Guo Y, Han S, Li Y, Zhang C, Bai Y. 2018. K-Nearest Neighbor combined with guided filter for hyperspectral image classification. *Procedia Computer Science* **129**:159–165 DOI 10.1016/j.procs.2018.03.066.

Hashim H, Abd Latif Z, Adnan NA. 2019. Urban vegetation classification with ndvi threshold value method with very high resolution (VHR) Pleiades Imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives)* **42(4/W16)**:237–240 DOI 10.5194/isprs-archives-XLII-4-W16-237-2019.

Huntingford C, Jeffers ES, Bonsall MB, Christensen HM, Lees T, Yang H. 2019. Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters* **14(12)**:124007 DOI 10.1088/1748-9326/ab4e55.

Jayawardhana WGNN, Chathurange VMI. 2020. Investigate the sensitivity of the satellite-based agricultural drought indices to monitor the drought condition of paddy and introduction to enhanced multi-temporal drought indices. *Journal of Remote Sensing and GIS* **9**:272 DOI 10.35248/2469-4134.20.9.272.

Kamal M, Jamaluddin I, Parela A, Farda NM. 2000. Comparison of Google Earth Engine (GEE)-based machine learning classifiers for mangrove mapping. In: *40th Asian Conf. Remote Sensing, ACRS 2019; Progress of Remote Sensing Technology for Smart Future*. 1–8.

Karakacan Kuzucu A, Bektas Balcik F. 2017. Testing the potential of vegetation indices for land use/cover classification using high resolution data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals)* **4(4W4)**:279–283 DOI 10.5194/isprs-annals-IV-4-W4-279-2017.

Karamzadeh S, Musaoğlu N, Türkeş M, Tanik AG. 2017. Using satellite-based indices for monitoring drought effects on the Buyuk Menderes River Basin. In: *8th atmospheric sciences symposium—01-04 November 2017 Istanbul - Turkey, 2018*, 414–419.

Kullarni VY, Sinha PK. 2013. Random forest classifier: a survey and future research directions. *International Journal of Advanced Computer Science* **36(1)**:1144–1156.

Lary DJ, Alavi AH, Gandomi AH, Walker AL. 2015. Machine learning in geosciences and remote sensing. *Geoscience Frontiers* **7(1)**:3–10 DOI 10.1016/j.gsf.2015.07.003.

Li Z, Xin X, Tang H, Yang F, Chen B, Zhang B. 2017. Estimating grassland LAI using the Random Forests approach and Landsat imagery in the meadow steppe of Hulunber, China. *Journal of Integrative Agriculture* **16(2)**:286–297 DOI 10.1016/S2095-3119(15)61303-X.

Maxwell AE, Warner TA, Fang F. 2018. Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing* **39(9)**:2784–2817 DOI 10.1080/01431161.2018.1433343.

Mika J, Horváth S, Makra L, Dunkel Z. 2005. The Palmer Drought Severity Index (PDSI) as an indicator of soil moisture. *Physics and Chemistry of the Earth* **30(1–3)**:223–230 DOI 10.1016/j.pce.2004.08.036.

Nam WH, Tadesse T, Wardlow BD, Hayes MJ, Svoboda MD, Hong E-M, Pachepsky YA, Jang M-W. 2018. Developing the vegetation drought response index for south korea (Vegdri-skorea) to assess the vegetation condition during drought events. *International Journal of Remote Sensing* **39(5)**:1548–1574 DOI 10.1080/01431161.2017.1407047.

Parks SA, Dillon GK, Miller C. 2014. A new metric for quantifying burn severity: the relativized burn ratio. *Remote Sensing* **6(3)**:1827–1844 DOI 10.3390/rs6031827.

Paseleng MC, Joko Prasetyo SY, Hartomo KD. 2018. Computer assisted learning on aridity disaster learning using SIMIA (Satellite Imagery for Modelling Index of Aridity. In: *The 2rd international seminar on innovation in mathematics and mathematics education 2018*.

Raczko E, Zagajewski B. 2017. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne

hyperspectral APEX images. *European Journal of Remote Sensing* **50(1)**:144–154 DOI 10.1080/22797254.2017.1299557.

Shi S, Yao F, Zhang J, Yang S. 2020. Evaluation of temperature vegetation dryness index on drought monitoring over Eurasia. *IEEE Access* **8**:30050–30059 DOI 10.1109/ACCESS.2020.2972271.

Tripathi R, Sahoo RN, Gupta VK, Sehgal VK, Sahoo PM. 2013. Developing Vegetation Health Index from biophysical variables derived using MODIS satellite data in the Trans-Gangetic plains of India. *Emirates Journal of Food and Agriculture* **25(5)**:376–384 DOI 10.9755/ejfa.v25i5.11580.

Verdin J, Funk C, Klaver R, Roberts D. 1999. Exploring the correlation between southern africa ndvi and pacific sea surface temperatures: results for the 1998 maize growing season. *International Journal of Remote Sensing* **20(10)**:2117–2124 DOI 10.1080/014311699212380.

Wang M, Au K, Ailamaki A, Brockwell A, Faloutsos C, Ganger GR. 2004. Storage device performance prediction with CART models. In: *Proc. - IEEE Comput. Soc. Annu. Int. Symp. Model. Anal. Simul. Comput. Telecommun. Syst. MASCOTS.* 588–595 DOI 10.1109/MASCOT.2004.1348316.

Wang S, Baig MHA, Liu S, Wan H, Wu T, Yang Y. 2018. Estimating the area burned by agricultural fires from Landsat 8 Data using the Vegetation Difference Index and Burn Scar Index. *International Journal of Wildland Fire* **27(4)**:217–227 DOI 10.1071/WF17069.

Wang Y, Zhang C, Meng FR, Bourque CPA, Zhang C. 2020. Evaluation of the suitability of six drought indices in naturally growing, transitional vegetation zones in Inner Mongolia (China). *PLOS ONE* **15(5)**:1–15 DOI 10.1371/journal.pone.0233525.

Wulder MA, White JC, Alvarez F, Han T, Rogan J, Hawkes B. 2009. Characterizing boreal forest wildfire with multi-temporal Landsat and LIDAR data. *Remote Sensing of the Environment* **113(7)**:1540–1555 DOI 10.1016/j.rse.2009.03.004.

Yang L, Jia K, Liang S, Liu J, Wang X. 2016. Comparison of four machine learning methods for generating the glass fractional vegetation cover product from modis data. *Remote Sensing* **8(8)**:1–16 DOI 10.3390/rs8080682.

Zaraza Aguilera MA. 2020. Classification of land-cover through machine learning algorithms for fusion of sentinel-2A and planetscope imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives)* **42(3/W12)**:361–368 DOI 10.5194/isprs-archives-XLII-3-W12-2020-361-2020.

Zhengmin X, Yuping H, Gongjin Z, Fuqiang W, Xiaoping D. 2014. Agricultural drought characteristics identification and analysis of Henan Province in China. *BioTechnology: An Indian Journal* **24(10)**:407–412.