Journal of Geoscience and Environment Protection, 2022, 10, 265-281
https://www.scirp.org/journal/gep
ISSN Online: 2327-4344
ISSN Print: 2327-4336

Application of Parametric and Non Parametric Classifiers for Assessing Land Use/Land Cover Categories in Cocoa Landscape of Juaboso and Bia West Districts of Ghana

Emmanuel Donkor[1*], Edward Matthew Osei Jnr[2], Stephen Adu-Bredu[3], Samuel A. Andam-Akorful[2], Efiba Vidda Senkyire Kwarteng[2], Lily Lisa Yevugah[4]

[1]Resource Management Support Centre of Forestry Commission, Kumasi, Ghana
[2]Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[3]Forestry Research Institute of Ghana, Kumasi, Ghana
[4]University of Energy and Natural Resources, Sunyani, Ghana
Email: *dadad105@yahoo.com

How to cite this paper: Donkor, E., Jnr, E. M. O., Adu-Bredu, S., Andam-Akorful, S. A., Kwarteng, E. V. S., & Yevugah, L. L. (2022). Application of Parametric and Non Parametric Classifiers for Assessing Land Use/Land Cover Categories in Cocoa Landscape of Juaboso and Bia West Districts of Ghana. *Journal of Geoscience and Environment Protection, 10,* 265-281.
https://doi.org/10.4236/gep.2022.1011018

Received: October 12, 2022
Accepted: November 27, 2022
Published: November 30, 2022

Abstract

Satellite image classification has been used for long time in the field of remote sensing since classification results are used in environmental research, agriculture, climate change and natural resource management. The cocoa landscape of Ghana is complex and diverse in nature, composing of mixture of closed forest, open forest, settlements, croplands and cocoa farms which make mapping the landscape difficult. The purpose of this research is to assess and compare the classification performances of three machine learning classifiers: Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN) and a statistical classification algorithm: Maximum Likelihood (ML) to know which classifier is best suited for mapping the cocoa landscape of Ghana using Juaboso and Bia West districts of Ghana as study area. A representative sampling approach was adopted to collect 1246 sample points for the various Land Use/Land Cover (LULC) types. These sample points were divided at random into 869 which form 70% for classification and 377 which constitute 30% of the total sample points for validation. The Stacked sentinel-2 image, classification data and validation data storing the identities of the LULC classes were imported in R to run supervised classification for each classifier. The classification results show that the highest overall accuracy and kappa statistics were produced by the support vector machine (86.47%, 0.7902); next is the artificial neural network (85.15%, 0.7700), followed by the random forest (84.08%, 0.7559) and finally the maximum likelihood (78.51%, 0.6668). The final LULC map produced under this study can be used to monitor cocoa driven deforestation especially in the gazetted forest and game reserves. This map will also be very useful in the national forest monitoring framework for the REDD + cocoa landscape project.

Keywords

Support Vector Machine, Random Forest, Artificial Neural Network, Maximum Likelihood, Image Classification, Cocoa Landscape

1. Introduction

Remote sensing has been an important source of land cover data during the last three decades (Foody et al., 2004). Improvement in satellite technology has made it possible to acquire land cover information over wide areas at varying spatial, spectral and radiometric resolutions (Hopkins et al., 1988).

The maximum likelihood, minimum distance to mean, parallelepiped, Mahalanobis distance and the box classifier are some of the traditional statistics-based classifiers used in remote sensing (Yu et al., 2014).

Machine learning methods, such as Support Vector Machine (SVM), Random Forest (RF), Decision Trees (DTs), Artificial Neural Network (ANN) and K-nearest neighbours (K-NN), have become common image classifiers as technology has evolved. Some research works have been done to compare the machine learning algorithms with traditional statistical classifiers, and they have been found to improve classification accuracy (Rogan et al., 2002). Traditional statistical classifiers are parametric algorithms. The major limitation of parametric classifiers is their reliance on the data statistical distribution. Also, they have low accuracy for image classification, whereas non parametric classifiers which are machine learning methods do not depend on data assigned to any specific statistical distribution (Caetano, 2009; Mountrakis et al., 2011).

Maximum likelihood (ML), Random forest (RF), Support vector machine (SVM) and Artificial neural network (ANN) classifiers have been chosen for this study, because they are extensively used in image classification (Zagajewski et al., 2021; Saeed et al., 2015).

Maximum likelihood (ML) is one of the simplest but commonly used statistical classification algorithm, in which a pixel with the maximum likelihood is classified into the corresponding class (Saeed et al., 2015). There are several reasons why the maximum likelihood classifier is so popular; first, the maximum likelihood decision rule is naturally appealing, because the most likely outcome among candidates is chosen (Bolstad & Lillesand, 1991). Additionally, covarying data, this frequently occurrence with satellite image data, can be easily accommodated by maximum likelihood classification. Finally, it has been demonstrated that maximum likelihood classifier, which takes variability into account, performs well across a variety of cover types (Lillesand & Kiefer, 1987). A study was conducted by comparing maximum likelihood, support vector machines and random forest techniques using RapidEye image for land cover mapping in the municipality of San Pelayo of Colombian Caribbean. It was found that, support vector machines produced the highest classification accuracy of 81.32%, followed by random forest 78.92% and finally maximum likelihood 68.95%. Though maximum likelihood produced the least classification accuracy, it was able to correctly classified infrastructure which was one of the classification classes better than the other two techniques (Valero et al., 2019) and this could be the ability of maximum likelihood to consider variability.

Random forest (RF) is one of the most widely used machine learning algorithms (Breiman, 2001). This algorithm is appealing since it is used for both classification and regression tasks (Woznicki et al., 2019). It is easy to use, efficient and accurate (Meltzer, 2021). Due to its versatility, RF has been applied in a variety of Earth scientific applications, such as modeling land-use (Araki & Yamamoto, 2018), land-cover (Nitze & Cawkwell, 2015) and modeling forest cover (Betts et al., 2017). Rodriguez-Galiano et al. (2012) examined RF to decision trees and found that RF provided a high accuracy of 92%, outperforming decision trees of accuracy of 83%. The ensemble architecture of RF, which trains multiple decision trees on different subsets of the training data, is thought to account for its improved accuracy.

Support vector machine (SVM) has been shown to outperform other classifiers due to its overall high capacity to simplify complex features (Shao & Lunetta, 2012). Support vector machine was able to achieve high overall accuracy of 88% in a land cover classification utilizing Landsat-8 and using six land-cover classes (Goodin et al., 2015). In order to map paddy rice in China in 2015, Mansaray et al. (2019) examined the effect of training sample size on the overall accuracies of SVM and RF. It was discovered that SVM and RF achieved overall accuracies of 91.8% and 89.2%, respectively.

Artificial neural network (ANN) has become a popular tool in the analysis of remotely sensed data (Mas & Flores, 2008). The ability of ANN to learn on its own and handle complicated issues is one of the reasons it has grown so popular (Di Franco & Santurro, 2021). Artificial neural network has been used in several land cover classification studies including using ANN, SVM and ML with IKONOS image for land cover mapping in Shahriar city of Iran. The classification results showed that, the overall accuracy and kappa coefficient of ANN (87.75%, 0.820) was better than that of SVM (85.57%, 0.819) and ML (78.36%, 0.729) (Saeed et al., 2015). Also comparing classification results of neural network called back propagation neural (BPN) and extended delta bar delta (EDBD) network with parallelepiped, minimum distance and maximum likelihood using Landsat 8 to classify land cover types in Minnesota of United States of America. The classification results revealed that the neural network performed best among the classifiers with overall accuracy and kappa of 95.07%, 0.935 respectively, followed by maximum likelihood (90.77%, 0.882), minimum distance (84.24%, 0.803), parallelepiped (69.23%, 0.612) (Zhang & Chang, 2015).

Maximum Likelihood (ML) is a supervised classification algorithm which is based on the Bayes theorem, assumes the reflectance values for each class in each band is normally distributed. During the ML classification, a given pixel has a probability that belongs to a particular class. As a result, the discriminant function is used to calculate each pixel's probability, and each pixel is then allotted to the class with the highest probability (Kulkarni, 2016). ML classifier has shown to perform effectively across a variety of land cover types as it takes variability into accounts (Lillesand & Kiefer, 1987).

Random Forest (RF) is an ensemble classifier, which means "union of parts". Random Forest uses more decision trees and makes prediction from each decision tree and selects the best outcome by means of voting (Breiman, 2001). One-third of the samples, known as the out-of-bag (OOB) samples, are excluded at random from each new training set that is created to help the tree grow. The tree is constructed using the remaining samples in the bag. The model performance can be evaluated using the OOB samples (Nguyen et al., 2015). Random forest is very flexible, has very high accuracy and also works better than a single decision tree. It does not suffer from the over fitting problem (Breiman, 2001).

Support Vector Machine (SVM) idea was developed by Cortes and Vapnik in 1995, which is a supervised learning method usually utilized in remote sensing applications. The main aim of SVM is to find the best hyperplane that divides the training data into several groups (Mountrakis et al., 2011).

Originally, Support Vector Machine (SVM) was to identify a linear class boundary. In order to overcome this restriction, the feature space was projected to a higher dimension on the grounds that a linear boundary might be present in a higher dimensional feature space. This projection to a higher dimensionality is known as the kernel trick. Kernel increases the number of dimensions in non-separable issues to make them separable. As a result of this, SVM becomes more powerful, adaptable and precise (Maxwell et al., 2018).

Artificial Neural Network works like the human brain and the building blocks are neurons. Each neuron has synaptic weights, which are specific coefficients that link it to other neurons. During training, information is sent to these joining points (Mijwil, 2018). Artificial Neural Network can learn complex configurations, taking into consideration any nonlinear complex relationship between the independent and the dependent variables (Jamali, 2021).

The High Forest Zone (HFZ) of Ghana, which contains the cocoa landscape, comprises 8.2 million hectares amounting to 34% of the country's total land area, with vegetation varying from wet evergreen to dry semi-deciduous (Forestry Commission, 2016; Indufor, 2015). Ghana's HFZ is made up of a complex web of forest, cocoa farms, croplands and human settlements (National REDD+ Secretariat, Forestry Commission, 2017). Implementing forest monitoring systems at the landscape level forms part of the HFZ's climate-smart and sustainable landscape activities. As prescribed by the Intergovernmental Panel on Climate Change, wall-to-wall mapping is essential for the proper execution of these forest monitoring systems (Mitchell et al., 2017).

Land cover maps that are precise and current are essential for environmental research, climate change monitoring and natural resource management (Pelletier et al., 2019). The cocoa landscape of Ghana is complex and diverse in nature, composing of mixture of closed forest, open forest settlements, croplands and cocoa farms making the mapping of the landscape very difficult. Some studies have been carried out in mapping the cocoa landscape of Ghana using one classification algorithm for each study (Benefoh et al., 2018; Ashiagbor et al., 2020). However, other classification algorithms for mapping the cocoa landscape of Ghana have not been fully explored. Hence there is the need to explore the performances of other image classifiers to know which algorithm is best suited for the classification of the cocoa landscape of Ghana as well as other countries with similar cocoa landscapes like Ghana. Benefoh et al. (2018) used Landsat 8 optical dataset applying image-fusion on vegetation indices (VIs) and digital elevation model (DEM) using maximum likelihood algorithm to detect and distinguish cocoa plantation from forest and other land use classes in the Krokosua Hills Forest Reserve catchment of Ghana. Also, in the Juaboso-Bia cocoa landscape of Ghana, Ashiagbor et al. (2020) used Sentinel-1 and Sentinel-2 satellite images to map mono cocoa, cocoa agroforestry, forest lands and other land use classes using random forest classifier.

The aim of this research is to assess and compare the performances of SVM, ANN, RF and ML classifiers to know which classifier is best suited for mapping the cocoa landscape of Ghana using the Juaboso and Bia West districts of Ghana as the study area.

2. Materials and Methods

2.1. Study Area

The study was carried out in Juaboso and Bia West districts in the Western North region of Ghana. The Western North region is the leading cocoa producing region in Ghana. Juaboso and Bia West districts are among the highest cocoa producing districts in the Western North region of Ghana. The study area is situated between latitude 6˚13'N to 6˚50'N and longitude 2˚40'W to 3˚16'W, covering an area of 2571.26 square kilometres or 257,126 hectares (**Figure 1**).

With a mean annual temperature between 25.5˚C and 26.5˚C, the area has a tropical climate marked by warm temperatures. The annual rainfall levels are between the ranges of 1250 - 2000 mm with June and October as its peak months (Ghana Statistical Service, 2014). The rainy and dry seasons are experienced within the study area; the wet season is from April to October, while the dry season lasts from November to March. Numerous food and commercial crops, particularly cocoa, are favoured by the comparatively long rainy season (Ghana Statistical Service, 2014). The elevation ranges between 137 - 594 m above sea level with Krokosua Hills in North West - South West part with the rest of the area on relatively lower elevation. The soils are primarily Oxysols and Ochrosols with Birimian and Hornblende as the parent rocks (Ghana Statistical Service, 2014).

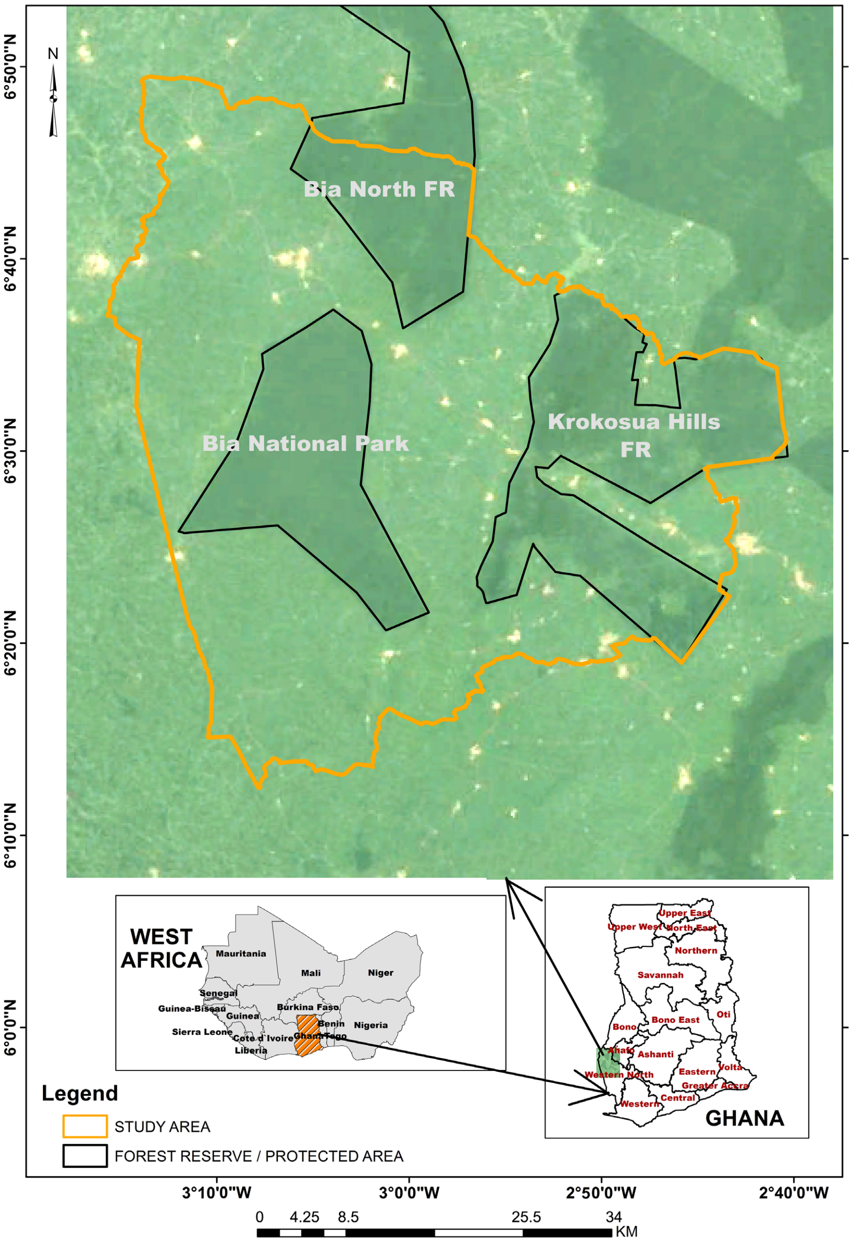

Figure 1. Map of the study area showing the cocoa mosaic landscape of Juaboso and Bia West districts of Ghana. (Image source: Landsat and Copernicus from goggle earth engine).

The study area falls within Moist Evergreen (ME), Moist Semideciduous North West (MSNW) and Moist Semideciduous South East (MSSE) subtypes ecological zones (Hall & Swaine, 1981). There are two forest reserves and one game reserve (protected area) in the study area. The forest reserves are Krokosua Hills and a portion of Bia Tributaries North and Bia National Park (protected area) all are under the administration of the Ghana Forestry Commission (Forestry Commission, 2016). The rest of the area is covered by farmlands mostly cocoa and communities in relatively low lying areas (Ghana Statistical Service, 2014).

2.2. Data Acquisition

Sentinel-2 images for 2020 with cloud cover of less than 10% were downloaded from the Copernicus Open Access Hub (https://scihub.copernicus.eu/). Sentinel-2 images are delivered in tiles, with each tile bearing a distinct name and the study area falls within tiles S2A-MSIL1C-20200105T103421-T30NVM, S2A-MSIL1C-20200204T103221-T30NWN and S2A-MSIL1C-20201210T103429-T30NVN. These images were downloaded as zip files and extracted into their various bands with each image having 13 bands.

Field data for the classification of the image were collected using representative sampling method, where points were collected based on the dominance of the land use/land cover types. At each location, the coordinates were picked using handheld GPS, the land use description at that point and the adjoining land use were recorded on the field sheet.

Data were collected from the following land use/land cover classes; closed forest, open forest, cocoa, settlement/bare surface and other vegetation (Table 1). In all, one thousand, two hundred and forty-six (1246) points were collected. These were made up of; 133 for closed forest, 152 in open forest, 728 for cocoa, 121 in settlement/bare surface and 112 for other vegetation. These sample points were divided at random into 869 which form 70% for classification and 377 which constitute 30% of the total sample points for validation.

2.3. Sentinel-2 Image Pre-Processing

The sentinel-2 images that were downloaded were in Level-1C processing format had only undergone geometric and radiometric corrections but were not atmospherically corrected. The images were converted into Level-2A product and atmospheric correction done using Sen2C or processor (Drusch et al., 2012; Müller-Wilm, 2018). Ten (10) bands; 2, 3, 4, 5, 6, 7, 8, 8A, 11 and 12 were

Table 1. Description of LULC types for the image classification.

LULC Type	Definition
Closed forest	Woody vegetation with a minimum mapping unit of 1 hectare and a canopy cover greater than 60% at a height of 5 m. Closed forest is mostly found in gazetted forest reserves and national Park with a small portion in off reserve sacred groves.
Open forest	Open forest areas are those that have a canopy cover within 15% to 60%. The low canopy cover may be due to excessive timber logging, mining and other environmental factor like bush fires.
Cocoa	Farmlands cultivated with cocoa. It includes cocoa farms without trees (mono cocoa) and those with trees (cocoa agroforestry).
Other Vegetation	Include annual food crop farm, fallowland and other tree crops like oil palm, citrus and rubber.
Settlement/Bare surface	These include areas with no vegetation such as human settlements, barren lands and mined-out areas.

stacked for each tile to produce composite image. The stacked images were mosaicked to form one composite image. The study area shapefile was used to subset the area of interest from the composite image and haze correction applied.

2.4. Supervised Image Classification and Accuracy Assessment

Supervised image classification was carried out using sentinel-2 imagery to determine the LULC types of the research area. The classification was performed using maximum likelihood, random forest, support vector machine and artificial neural network algorithms in R software. Stacked sentinel-2 image, training data and validation data in polygon shapefiles storing the identity for each land cover type were imported into R. Caret, Rstoolbox, rgdal and raster packages were imported into R for the classification. Sentinel-2 image was imported as raster brick using the brick function in raster package.

Maximum Likelihood (ML) classification was done using "rasclass" package in R to train and fit ML model. Random Forest (RF) classification was executed by the "randomForest" package in R. Support Vector Machine (SVM) classification was done using "e1071" package and artificial neural network classification using "neuralnet" package in R.

Using the training data in combination with a classifier, the pixel values in the training area for every band in sentinel-2 were extracted and stored in a data frame with its corresponding LULC class ID (Table 2) to train and fit the model. The classification was carried out separately for each classifier and saved the output classified image file. The output classified images were filtered to remove the speckles from the classified images to enhance their appearance. The final maps were prepared and area for the landuse classes was calculated.

Accuracy assessment was performed for each classifier by using the classified image and 377 validation points in R to generate the confusion matrix, the overall accuracy and the kappa. User's Accuracy (UA) and Producer's Accuracy (PA) were calculated from the confusion matrix.

3. Results

3.1. Image Classification

Supervised classification was used to categorize the research area into five (5)

Table 2. Class ID and LULC name.

Class ID	LULC Name
1	Closed forest
2	Open forest
3	Cocoa
4	Other Vegetation
5	Settlement/Bare surface

LULC classes; closed forest, open forest, cocoa, other vegetation and settlement/bare surface using the four classifiers with combined maps shown in Figure 2.

Maximum likelihood classifier map shows five classes; closed forest, open forest, cocoa, other vegetation and settlement/bare surface. Other vegetation class has a small class area, hence appearing patchy on the map. Random forest, support vector machine and artificial neural network classifiers maps display all the five LULC classes well; this implies these classifiers separated all the classes well under this study.

A summary of the LULC classes areas for the four classifiers is presented in (Table 3) with its bar chart (Figure 3).

3.2. Accuracy Assessment

The accuracy assessment based on the classified images in R generated the confusion matrix, overall accuracy and kappa.

The confusion matrix and the accuracy report for the four classifiers are presented in Table 4.

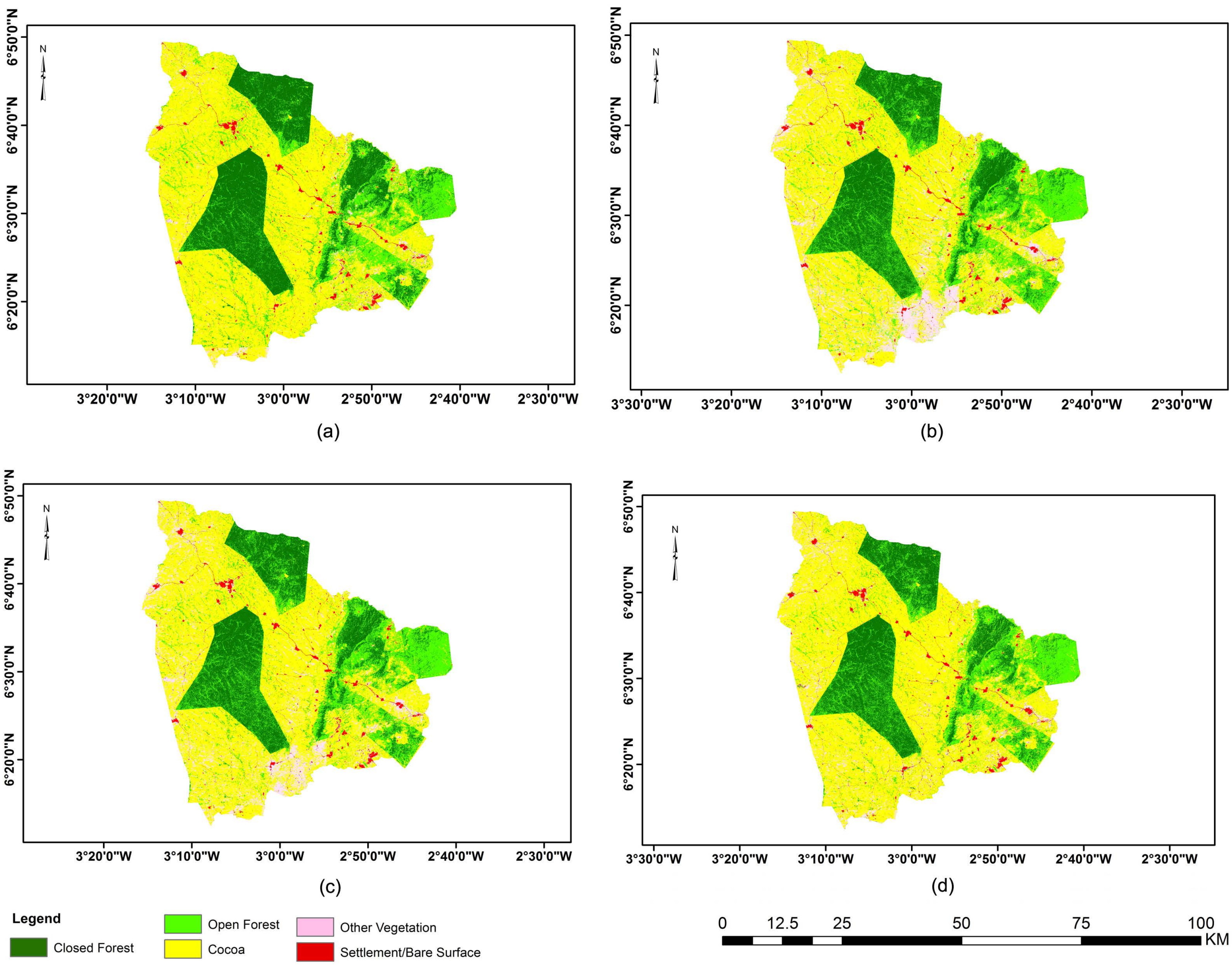

Figure 2. LULC maps of the study area using (a) ML; (b) RF; (c) SVM; (d) ANN.

Table 3. LULC areas for each classifier.

CLASSIFIER	ML		RF		SVM		ANN	
LULC Class	Area (ha)	Area (%)	Area (ha)	Area (%)	Area (ha)	Area (%)	Area (ha)	Area (%)
Closed Forest	57057.40	22.19	47314.91	18.40	48913.36	19.02	48010.46	18.67
Open Forest	43908.38	17.08	47958.84	18.65	43637.70	16.97	44070.35	17.14
Cocoa	143767.99	55.91	127943.8	49.76	135329.8	52.63	145006.76	56.40
Other Vegetation	7829.25	3.04	28645.41	11.14	24808.61	9.65	15127.17	5.88
Settlement/Bare Surface	4562.98	1.78	5262.99	2.05	4436.46	1.73	4911.26	1.91
TOTAL	**257126**	**100**	**257126**	**100**	**257126**	**100**	**257126**	**100**

Table 4. Confusion matrix and accuracy report for the four classifiers.

LULC	Closed Forest	Open Forest	Cocoa	Other Vegetation	Settlement/Bare Surface	Total	PA (%)	UA (%)
				Maximum Likelihood Classifier				
Closed Forest	36	20	6	1	0	63	81.82	57.14
Open Forest	8	30	14	1	0	53	60	56.6
Cocoa	0	0	184	16	9	209	89.32	88.04
Other Vegetation	0	0	2	19	4	25	51.35	76
Settlement/Bare Surface	0	0	0	0	27	27	67.5	100
Total	44	50	206	37	40	377		
Overall Accuracy				78.51%				
Kappa Statistics				0.6668				
				Random Forest Classifier				
Closed Forest	34	10	0	1	0	45	77.27	75.56
Open Forest	10	38	0	0	0	48	70	79.17
Cocoa	0	2	188	9	5	204	91.26	92.16
Other Vegetation	0	0	17	27	5	49	72.97	55.1
Settlement/Bare Surface	0	0	1	0	30	31	75	96.77
Total	44	50	206	37	40	377		
Overall Accuracy				84.08%				
Kappa Statistics				0.7559				
				Support Vector Machine Classifier				
Closed Forest	34	9	0	1	0	44	77.27	77.27
Open Forest	10	39	0	0	0	49	78	79.59
Cocoa	0	1	195	8	6	210	94.66	92.86
Other Vegetation	0	1	11	28	4	44	75.68	63.64

Continued

Settlement/Bare Surface	0	0	0	0	30	30	75	100
Total	44	50	206	37	40	377		
Overall Accuracy				86.47%				
Kappa Statistics				0.7902				
				Artificial Neural Network				
Closed Forest	38	10	0	0	0	48	86.36	79.17
Open Forest	6	38	3	0	0	47	76.00	80.85
Cocoa	0	2	191	15	1	209	92.72	91.39
Other Vegetation	0	0	8	22	7	37	59.46	59.46
Settlement/Bare Surface	0	0	4	0	32	36	80.00	88.89
Total	44	50	206	37	40	377		
Overall Accuracy				85.15%				
Kappa Statistics				0.7700				

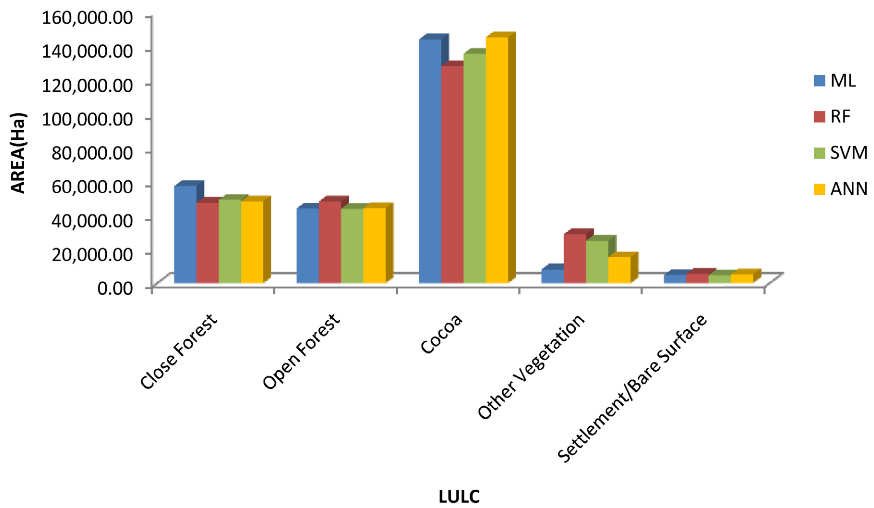

Figure 3. LULC classes areas per classifier.

The highest overall accuracy is 86.47% for the SVM classifier, followed by ANN (85.15%), RF (84.08%) and the least is ML (78.51%). In addition, the kappa statistics of 0.7902 is the highest for SVM, next is ANN (0.7700), followed by RF (0.7559) with ML having the least (0.6668).

The Kappa is the agreement between the model prediction and observed (Delgado & Tibau, 2019). It provides a more accurate indicator of the overall performance of the classifier. This is due to the possibility of a simple accuracy can be skewed if the class distribution is also skewed. Van Ness et al. (2008) considered kappa more than 0.75 as excellent and between 0.4 to 0.75 to be fair to good, hence SVM, ANN and RF are excellent classifiers while ML is a good classifier per this study.

Overall accuracy is somewhat inadequate for summarizing the accuracy of LULC classification. This is because the overall accuracy generated from the accuracy assessment is an average value that does not tell whether the error was evenly distributed across the LULC classes. Consequently, two other measures are often used, which are the producer's accuracy and the user's accuracy. Producer's accuracy indicates for a given class the proportion of the reference data that are classified correctly. User's accuracy calculates for a given class how many pixels are actually what the classification claims they are (Rwanga & Ndambuki, 2017). From (**Table 4**) the overall accuracy for ML classifier is 78.51% with closed forest having 81.82% and 57.14% as PA and UA respectively. This means that 81.82% of the closed forest area has been identified correctly with 57.14% identified as truly closed forest from classification perspective. Using the ML classifier the highly reliable LULC class associated with this classification is the cocoa class from the producer's accuracy and user's accuracy viewpoint with PA (89.32%) and UA (88.04%). Similarly using RF, SVM and ANN classifiers, the classification was able to map the cocoa class very well based on the producer's accuracy and user's accuracy with the highest in SVM (94.66% and 92.86%) for PA and UA respectively.

4. Discussion

The major LULC in the study area is cocoa as obtain from all four (4) classifiers, with the highest area obtained for ANN (56.40%) and the least in RF (49.76%). After cocoa, closed forest is the next in terms of area coverage with the highest area occurring in ML (22.19%) and the least in RF (18.40%). Open forest follows with RF (18.65%) as the highest and SVM (16.97%) as the least in terms of area coverage in this LULC class. Other vegetation follows with RF (11.14%) as the highest with the least in ML (3.04%) as an area in this class. The smallest area is the settlement/bare surface LULC class with RF having the highest area (2.05%) and the least in SVM (1.73%).

Support vector machine produced the highest overall accuracy and kappa of 86.47%, 0.7902 respectively, followed by ANN (85.15%, 0.7700), RF (84.08%, 0.7559) and the least is ML (78.51%, 0.6668). Support vector machine has the ability to handle minimal training data sets and usually produce higher classification accuracy (Bouaziz et al., 2017). Khatami et al. (2016) revealed that SVM was the best among numerous classifiers, including random forest, neural network and decision trees.

Random forest and artificial neural network also performed very well and their performances are closed to that of SVM. Each decision tree is constructed with random forest using a subset of the features. This is favourable because each decision tree may make a precise classification determination that is based only on useful features and the decision trees perform voting to come out with the final classification (Tian et al., 2016). Artificial Neural Network performs supervised classification using small data and the ability to integrate multiple types of data in the study, because there are no assumptions about the data used (Mas & Flores, 2008).

Maximum likelihood classifier performance is fairly good, as it was able to segregate closed forest, open forest, cocoa and settlement/bare surface LULC classes with some misclassification in the other vegetation class. The inability of ML to classify the other vegetation very well may be as a result of the mixed and complex environment of the landscape. Parametric classifiers such as ML is not best suited for complex systems (Mishra, 2018).

The classification results in a research earlier conducted by Benefoh et al. (2018) in the Krokosua Hills forest reserve catchment of Ghana using maximum likelihood method gave an overall accuracy of 82.6% and a kappa of 0.73. Also Ashiagbor et al. (2020) classification results in the Juaboso-Bia cocoa landscape of Ghana using sentinel-2 bands and its Vegetation Indices (VIs) with random forest classifier produced overall accuracy of 79.02% and kappa of 0.748.

One drawback observed in this study is imbalanced classification, thus, where the training dataset is biased or skewed towards a class or classes. Imbalanced classifications present a challenge for predicting models because the majority of machine learning algorithms for classification were built on the premise that there should be an equal number of samples in each class. As a result, models perform poorly when making predictions, especially for the class with small training samples (Browniee, 2019).

A total of 869 training samples were used for the classification, with closed forest class constituting 10.70%, open forest 12.20%, cocoa 58.45%, settlement 9.67% and other vegetation 8.98%. From **Figure 2**, other vegetation class was not well represented and this is due to the small number of training samples used for the classification and cocoa class was visibly represented because more samples of cocoa was used.

5. Conclusion and Recommendation

This research has demonstrated the comparative ability of SVM, ANN, RF and ML classifiers to map the cocoa landscape of the Juaboso and Bia West districts of Ghana. Based on overall classification accuracy, kappa statistics, producer's accuracy and user's accuracy, SVM is the best among the four classifiers for mapping LULC categories of the study area. Hence SVM classifier map could be used as the final LULC map for the study area. Also classification accuracy of the SVM, ANN and RF methods were close to each other and higher than ML, which indicates that non-parametric algorithms like machine learning techniques can deliver more precise results than parametric algorithm like maximum likelihood classifier. The reason is that, the method of the mapping function is not assumed by non-parametric classifiers.

The final LULC map produced in this research provides useful information on the area used for cocoa farming in the study area with 53% of the total land mass under cocoa cultivation. The LULC map can be used to monitor cocoa driven deforestation especially in the gazetted forest and game reserves. This map will also be very useful in the national forest monitoring framework for the REDD + cocoa landscape project in Ghana.

It is recommended that, in future using machine learning algorithms to perform supervised image classification for complex ecosystems like the cocoa landscape of Ghana, the training samples to be used should be almost the same for each class in order minimize the problem of imbalanced classification.

Conflicts of Interest

The authors declare no conflict of interest with respect to the publication of this paper.

Author Statement

Emmanuel Donkor (ED): Conceptualization, Investigation, Formal Analysis, Writing-Original Draft, Writing-Review & Editing, Visualization.

Edward Matthew Osei Jnr (EMO): Writing-Original Draft, Writing-Review & Editing.

Stephen Adu-Bredu (SAB): Formal Analysis, Writing-Original Draft, Writing-Review & Editing.

Samuel A, Andam-Akorful (SAAA): Formal Analysis, Writing-Original Draft, Writing-Review & Editing, Supervision.

Efiba Vidda Senkyire Kwarteng (EVSK): Formal Analysis, Writing-Original Draft, Writing-Review & Editing.

Lily Lisa Yevugah (LLY): Formal Analysis, Writing-Original Draft, Writing-Review & Editing.

References

Araki, S., Shima, M., & Yamamoto, K. (2018). Spatiotemporal Land Use Random Forest Model for Estimating Metropolitan NO_2 Exposure in Japan. *Science of the Total Environment, 634,* 1269-1277. https://doi.org/10.1016/j.scitotenv.2018.03.324

Ashiagbor, G., Forkuo, E. K., Asante, W. A., Acheampong, E., Quaye-Ballard, J. A, Boamah, P., Yakubu, M., & Foli, E. (2020). Pixel-Based and Object-Oriented Approaches in Segregating Cocoa from Forest in the Juabeso-Bia Landscape of Ghana. *Remote Sensing Applications: Society and Environment, 19,* Article ID: 100349. https://doi.org/10.1016/j.rsase.2020.100349

Benefoh, D. T., Villamor, G. B., Van Noordwijk, M., Borgemeister, C., Asante, W. A., & Asubonteng, K. O. (2018). Assessing Land-Use Typologies and Change Intensities in a Structurally Complex Ghanaian Cocoa Landscape. *Applied Geography, 99,* 109-119. https://doi.org/10.1016/j.apgeog.2018.07.027

Betts, M. G., Wolf, C., Ripple, W. J., Phalan, B., Millers, K. A., Adam, D., Butchart, S. H. M., & Levi, T. (2017). Global Forest Loss Disproportionately Erodes Biodiversity in Intact Landscapes. *Nature, 547,* 441-444. https://doi.org/10.1038/nature23285

Bolstad, P. V., & Lillesand, T. M. (1991). Rapid Maximum Likelihood Classification. *Photogrammetric Engineering & Remote Sensing, 57,* 67-74.

Bouaziz, M., Eisold, S., & Guermai, E. (2017). Semiautomatic Approach for Land Cover

Classification: A Remote Sensing Study for Arid Climate in Southeastern Tunisia. *Euro-Mediterranean Journal for Environmental Integration, 2,* Article No. 24. https://doi.org/10.1007/s41207-017-0036-7

Breiman, L. (2001). Random Forests. *Machine Learning, 45,* 5-32. https://doi.org/10.1023/A:1010933404324

Browniee, J. (2019). *Python Machine Learning.* https://machinelearningmastery.com/hyperparameters-for-classification-machine-learning-algorithms

Caetano, M. (2009). *Image Classification.* ESA Advanced Training Course on Land Remote Sensing.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning, 20,* 273-297. https://doi.org/10.1007/BF00994018

Delgado, R, & Tibau, X.-A. (2019). Why Cohen's *Kappa* Should Be Avoided as Performance Measure in Classification. *PLOS ONE, 14,* e0222916. https://doi.org/10.1371/journal.pone.0222916

Di Franco, G., & Santurro, M. (2021). Machine Learning, Artificial Neural Networks and Social Research. *Quality & Quantity, 55,* 1007-1025. https://doi.org/10.1007/s11135-020-01037-y

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B. et al. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment, 120,* 25-36. https://doi.org/10.1016/j.rse.2011.11.026

Foody, M. G., & Mathur, A. (2004). Toward Intelligent Training of Supervised Image Classifications: Directing Training Data Acquisition for SVM Classification. *Remote Sensing of Environment*, *93*, 107-117. https://doi.org/10.1016/j.rse.2004.06.017

Forestry Commission (2016). *Ghana REDD+ Strategy.* https://www.forestcarbonpartnership.org/sites/fcp/files/2015/April/Ghana%20National%20REDD%2B%20Strategy%20Final.pdf

Ghana Statistical Service (2014). *District Analytical Report: Juaboso and Bia West Districts.* https://www2.statsghana.gov.gh/docfiles/2010_District_Report/Western/Juaboso.pdf https://www2.statsghana.gov.gh/docfiles/2010_District_Report/Western/Bia%20West.pdf

Goodin, D. G., Anibas, K. L., & Bezymennyi, M. (2015). Mapping Land Cover and Land Use from Object-Based Classification: An Example from a Complex Agricultural Landscape. *International Journal of Remote Sensing, 36,* 4702-4723. https://doi.org/10.1080/01431161.2015.1088674

Hall, J. B., & Swaine, M. D. (1981). Distribution and Ecology of Vascular Plants in a Tropical Rain Forest: Forest Vegetation in Ghana. In *Geobotany* (Vol. 1). Springer Nature. https://doi.org/10.1007/978-94-009-8650-3

Hopkins, P. F., Maclean, A. L., & Lillesand, T. M. (1988). Assessment of Thematic Mapper Imagery for Forestry Applications under Lakes States Conditions. *Photogrammetry Engineering & Remote Sensing, 54,* 61-70.

Indufor, O. (2015). *Development of Reference Emissions Levels and Measurement, Reporting and Verification System in Ghana.* https://www.forestcarbonpartnership.org/system/files/documents/Ghana%20MRV%20Final%20Report%20%28ID%2067024%29.pdf

Jamali, A. (2021). Improving Land Use Land Cover Mapping of a Neural Network with Three Optimizers of Multi-Verse Optimizer, Genetic Algorithm, and Derivative-Free Function. *The Egyptian Journal of Remote Sensing and Space Science, 24,* 373-390.

https://doi.org/10.1016/j.ejrs.2020.07.001

Khatami, R., Mountrakis, G., & Stehan, V. S. (2016). A Meta-Analysis of Remote Sensing Research on Supervised Pixel-Based Land-Cover Image Classification Processes: General Guidelines for Practitioners and Future. *Remote Sensing of Environment, 177,* 89-100. https://doi.org/10.1016/j.rse.2016.02.028

Kulkarni, A. D. (2016). Random Forest Algorithm for Land Cover Classification. *International Journal on Recent and Innovation Trends in Computing and Communication, 4,* 2321-8169.

Lillesand, T. M., & Kiefer, R. W. (1987). *Remote Sensing and Image Interpretation* (2nd ed.). John Wiley & Sons.

Mansaray, L. R., Wang, F., Huang, J., Yang, L., & Kanu, A. S. (2020). Accuracies of Support Vector Machine and Random Forest in Rice Mapping with Sentinel-1A, Landsat-8 and Sentinel-2A Datasets. *Geocarto International, 35,* 1088-1108. https://doi.org/10.1080/10106049.2019.1568586

Mas, J. F., & Flores, J. J. (2008). The Application of Artificial Neural Networks to the Analysis of Remotely Sensed Data. *International Journal of Remote Sensing, 29,* 617-663. https://doi.org/10.1080/01431160701352154

Maxwell, A. E., Warner, T. A., Fang, F., Maxwell, A. E., & Warner, T. A. (2018). Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *International Journal of Remote Sensing, 39,* 2784-2817. https://doi.org/10.1080/01431161.2018.1433343

Meltzer, R. (2021). *What Is Random Forest?* https://careerfoundry.com/en/blog/data-analytics/what-is-random-forest

Mijwil, M. M. (2018). *Artificial Neural Networks Advantages and Disadvantages.* University of Baghdad.

Mishra, S. (2018). *Parametric and Non-Parametric Methods.* https://www.linkedin.com/pulse/parametric-non-parametric-methods-satya-mishra

Mitchell, A. L., Rosenqvist, A., & Mora, B. (2017). Current Remote Sensing Approaches to Monitoring Forest Degradation in Support of Countries Measurement, Reporting and Verification (MRV) Systems for REDD+. *Carbon Balance and Management, 12,* Article No. 9. https://doi.org/10.1186/s13021-017-0078-9

Mountrakis, G., Im, J., & Ogole, C. (2011). Support Vector Machines in Remote Sensing: A Review. *ISPRS Journal of Photogrammetry and Remote Sensing, 66,* 247-259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Müller-Wilm, U. (2018). *S2 MPC: Sen2Cor Configuration and User Manual.* https://step.esa.int/thirdparties/sen2cor/2.5.5/docs/S2-PDGS-MPC-L2A-SUM-V2.5.5_V2.pdf

National REDD+ Secretariat, & Forestry Commission (2017). *Ghana's National Forest Reference Level.* https://redd.unfccc.int/files/ghana__modified_frl_november_10_2017_clean.pdf

Nguyen, T.-T., Huang, J. Z., & Nguyen, T. T. (2015). Unbiased Feature Selection in Learning Random Forests for High-Dimensional Data. *Research and Development of Advanced Computing Technologies, 2015,* Article ID: 471371. https://doi.org/10.1155/2015/471371

Nitze, I., Barrett, B., & Cawkwell, F. (2015). Temporal Optimisation of Image Acquisition for Land Cover Classification with Random Forest and MODIS Time-Series. *International Journal of Applied Earth Observation and Geoinformation, 34,* 136-146. https://doi.org/10.1016/j.jag.2014.08.001

Pelletier, C., Webb, G. I., & Petitjean, F. (2019). Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sensing, 11,* Article No. 523. https://doi.org/10.3390/rs11050523

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67,* 93-104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L., & Fischer, C. (2002). Land-Cover Change Monitoring with Classification Trees Using Landsat TM and Ancillary Data. *Photogrammetric Engineering & Remote Sensing, 69,* 793-804. https://doi.org/10.14358/PERS.69.7.793

Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. *International Journal of Geosciences, 8,* 611-622. https://doi.org/10.4236/ijg.2017.84033

Saeed, O., Hamid, E., & Ahmadi, F. F. (2015). Using Artificial Neural Network for Classification of High Resolution Remotely Sensed Images and Assessment of Its Performance Compared with Statistical methods. *American Journal of Engineering, Technology and Society, 2,* 1-8.

Shao, Y., & Lunetta, R. S. (2012). Comparison of Support Vector Machine, Neural Network, and CART Algorithms for the Land-Cover Classification Using Limited Training Data Points. *ISPRS Journal of Photogrammetry and Remote Sensing, 70,* 78-87. https://doi.org/10.1016/j.isprsjprs.2012.04.001

Tian, S., Zhang, X., Tian, J., & Sun, Q. (2016). Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China. *Remote Sensing, 8,* Article No. 954. https://doi.org/10.3390/rs8110954

Valero Medina, J. A., & Alzate Atehortúa, B. E. (2019). Comparison of Maximum Likelihood, Support Vector Machines, and Random Forest Techniques in Satellite Images Classification. *Tecnura, 23,* 13-26. https://doi.org/10.14483/22487638.14826

Van Ness, P. H., Towle, V. R., & Juthani-Mehta, M. (2008). Testing Measurement Reliability in Older Populations: Methods for Informed Discrimination in Instrument Selection and Application. *Journal of Aging and Health, 20,* 183-197. https://doi.org/10.1177/0898264307310448

Woznicki, S. A., Baynes, J., Panlasigui, S., Mehaffey, M., & Neale, A. (2019). Development of a Spatially Complete Floodplain Map of the Conterminous United States Using Random Forest. *Science of the Total Environment, 647,* 942-953. https://doi.org/10.1016/j.scitotenv.2018.07.353

Yu, L., Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L. et al. (2014). Meta-Discoveries from a Synthesis of Satellite-Based Land-Cover Mapping Research. *International Journal of Remote Sensing, 35,* 4573-4588. https://doi.org/10.1080/01431161.2014.930206

Zagajewski, B., Kluczek, M., Raczko, E., Njegovec, A., Dabija, A., & Kycko, M. (2021). Comparison of Random Forest, Support Vector Machines, and Neural Networks for Post-Disaster Forest Species Mapping of the Krkonoše/Karkonosze Transboundary Biosphere Reserve. *Remote Sensing, 13,* Article No. 2581. https://doi.org/10.3390/rs13132581

Zhang, S. L., & Chang, T. C. (2015). A Study of Image Classification of Remote Sensing Based on Back-Propagation Neural Network with Extended Delta Bar Delta. *Mathematical Problems in Engineering, 2015,* Article ID: 178598. https://doi.org/10.1155/2015/178598