Journal of Geography, Environment and Earth Science International

25(5): 1-20, 2021; Article no.JGEESI.69347
ISSN: 2454-7352

Mapping Perennial Crops in Africa: A Case Study of Oil Palm in Ghana

Njomaba Elisha[1*], James Nana Ofori[2], Ben Emunah Aikins[3] and Gideon Adzraku[3]

[1]Laboratory of Geoinformation Science and Remote Sensing, Wageningen University and Research, Netherland.
[2]Department of Plant Ecology and Nature Conservation, Wageningen University and Research, Netherland.
[3]Remote Sensing and GIS Unit, Bedeston Technologies, and Consulting Services, Accra, Tema, Ghana.

Authors' contributions

This work was carried out in collaboration among all authors. Author EN designed the study, performed the remote sensing analysis, wrote the protocol and wrote the first draft of the manuscript. Authors JN and BEA managed the literature searches and assisted with the training and validation data collection. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JGEESI/2021/v25i530283
Editor(s):
(1) Dr. Wen-Cheng Liu, National United University, Taiwan.
Reviewers:
(1) Musingo Tito E. Mbuvi, Kenya Forestry Research Institute (KEFRI), Kenya.
(2) Ovat Innocent Ovat, Cross River University of Technology, Nigeria.
Complete Peer review History: http://www.sdiarticle4.com/review-history/69347

Original Research Article

Received 10 April 2021
Accepted 15 June 2021
Published 19 June 2021

ABSTRACT

Forests in Sub-Saharan Africa are experiencing some of the highest rates of deforestation and degradation in the world, with most natural forest species being replaced by cropland and plantation monoculture. In this work, a method was developed that combined the Synthetic Aperture Radar (Sentinel-1) and optical satellite imagery (Sentinel-2) data to accurately map natural forest and perennial crops (oil palm) in Ghana. This was done using all three variables including spatial, spectral, and temporal variables to assess the most important variables in characterizing oil palm and natural forest, as well as the added value of sentinel-1 SAR data in a sentinel-2 optical-based classification. In this workflow, the Gray level co-occurrence matrix (GLCM) was calculated as representing textural/spatial variables, a yearly median composite to represent the spectral variables, and raining and dry season composites of Normalized Difference Vegetation Index

**Corresponding author: E-mail: enjomaba@st.ug.edu.gh;*

(NDVI) and Normalized Difference Moisture Index (NDMI) to represent the temporal variables for the Sentinel-2 data. In terms of the SAR data, rainy and dry season composites of NDVI and NDMI were calculated. With all these variables together, a characterization of the study area was conducted based on reference data of the land use land cover classes including oil palm, natural forests, and croplands (others) using Random Forest classifier. The variable importance of the Random Forest model was investigated to identify the top 10 most important variables. Results from this study showed that spectral variables followed by spatial variables are the most important and need to be considered when characterizing oil palm and natural forest, which is consistent with some pieces of literature. The use of sentinel-2 data achieved an acceptable classification accuracy (75%); whereas, sentinel-1 SAR further increased the accuracy (up to 85%) as compared to sentinel-2 only.

Keywords: Mapping; plantation; forests; oil palm.

1. INTRODUCTION

About 4.7 million ha/year of net forest loss occurs globally and a 3 million ha/year increase in plantation forests has been recorded between the years 2010 and 2020 by the Food and Agriculture Organization of the United Nations (FAO, 2020). Forest changes have been attributed to certain drivers such as the conversion of land use/ land cover (LULC) for the production of commodities, agriculture shifting, forestry, urbanization, and wildfires [1]. The modifications of the land have impacted the environment negatively including the increase in greenhouse gas emissions, disruption of the water cycle, increase in soil erosion, biodiversity loss, and local livelihoods disruption [2]. To overcome these problems, some policies and frameworks have been proposed and implemented from local to global scale levels, with the emphasis on detailed and accurate measurements of forest types [3]. Mapping natural and plantation forests can provide accurate input for the detection of deforestation, climate change, carbon assessment, and detection of biodiversity loss.

Some studies on mapping perennial crops and natural forests have been done and they utilized remote sensing data through different approaches. One typical approach is to use phenological characteristics of some specific plantation types based on time series satellite sensor imagery [3]. An example of a study using this approach adopted the difference in spectral characteristics of the defoliation period of deciduous rubber to separate it from natural forests [4]. Another approach has been to use image processing techniques to enhance the characterization of plantation forests and natural forests. Textural analysis has been specifically used to differentiate the unique spatial pattern of the targeted perennial crops, e.g., oil palm fields from other surrounding land covers [4]. Moreover, vegetation indices were incorporated to amplify the differences in mapping plantations as well as natural forests, such as NDVI, enhanced vegetation index (EVI), soil-adjusted total vegetation index (SATVI), normalized difference tillage Index (NDTI), and land surface water index (LSWI). In addition, the development of radar satellite sensors has been seen its involvement in LULC mapping as well as forest monitoring (The SAR Handbook, 2019). Synthetic Aperture Radar (SAR) images have been used widely in forest mapping at both global and regional scales since they can provide cloud-free structural information sensitive to forest cover [5]. Another approach that recent studies have frequently used is the combination of optical and SAR imagery [6]. The common issue that runs through all this research as outlined above is that the focus is on using a subset of variables only. That is, either spectral, spatial, and temporal variables or a combination of two out of the three variables for such classifications and this creates problems that might affect the classification accuracies. For accurate classification of landcover, there is the need to look at all variables to improve the accuracy of classification. This study was to find out which variables (spatial, spectral, and temporal) variables are of influence when mapping natural forests and perennial crops. This was done by considering all variables (spectral, spatial, and temporal) rather than a subset of such variables.

Also, studies on mapping perennial crops have mainly used single-temporal medium or high-resolution optical images for oil palm plantation recognition. This yielded a low identification accuracy with the former, though much improved when using the latter [7]. However, high-resolution image data have generally been expensive and unable to provide high temporal resolutions, making it difficult to identify and monitor oil palm plantations at high temporal frequencies over large areas. In this context, a combination of sentinel-2 and sentinel-1 is used for monitoring oil plantations in the study area imagery.

The objective of this paper is to increase accuracy in mapping natural forest and perennial crops (oil palm) in Ghana, using remote sensing data. The specific objectives were: (1) assess the most important variable to be considered when characterizing natural forest and perennial crops in Ghana; (2) evaluate the classification performance of the satellite data with sentinel-2 data only and combined sentinel-2 and sentinel-1.

2. MATERIALS AND METHODS

2.1 Study Area

The Ejisu Municipality, Ghana (Fig.1) lies between longitude 1° 15' W–1° 45' W and latitude 6° 15' N–7° 00' N and falls within the forest-dissected plateau terrain of Ghana. The municipality stretches over an area of about 637.2km2. It has a tropical and wet semi-equatorial climate with annual average temperatures ranging from 20^0C in August to 32^0C in March and a mean annual rainfall of 1200mm, mainly from March to July, with high relative humidity during this period. The main agronomical tree crops grown in the area are maize, cocoyam, plantain, cassava, citrus, cocoa, and oil palm. Since the area has a variety of forest types, it was suitable as a testing area for classification.

2.2 Methods

The overall workflow of establishing the high-resolution LULC map for Vietnam is illustrated in Fig. 2.

2.3 Class Definitions

For land cover classification, three main classes were identified from satellite and aerial images with local knowledge of the study area as well as land use land cover maps from other research conducted within the study area. These are the oil palm, natural forest, and others (croplands). The final land cover delineated as others were defined as croplands such as maize, cocoyam, plantain, cassava, citrus, cocoa, and grasslands. An overview of the defined classes is shown in Fig. 3.

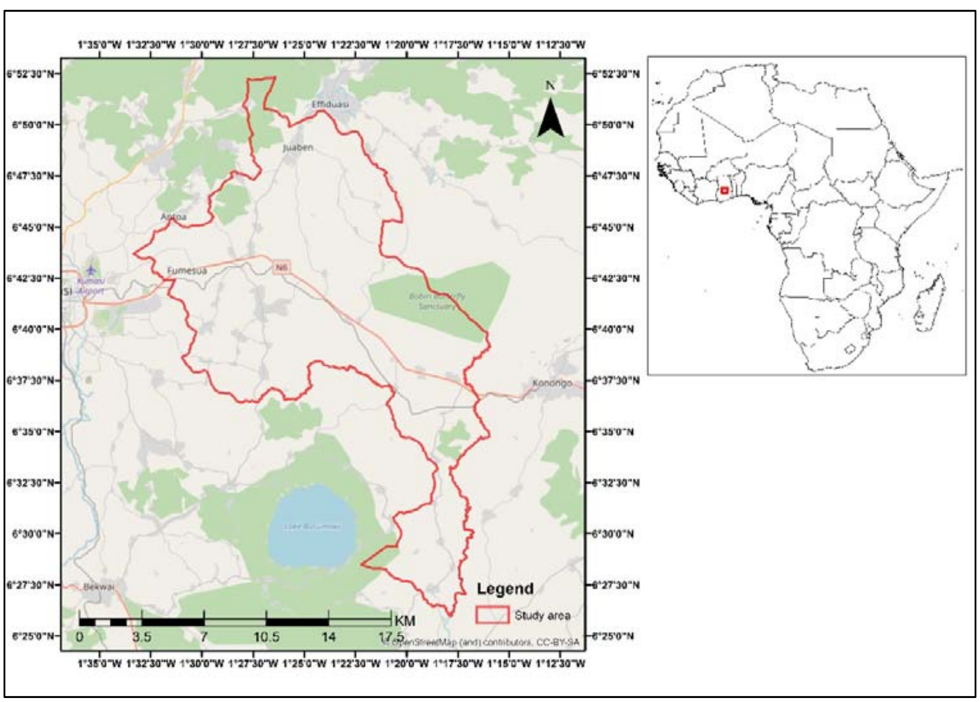

Fig. 1. Map of the study area

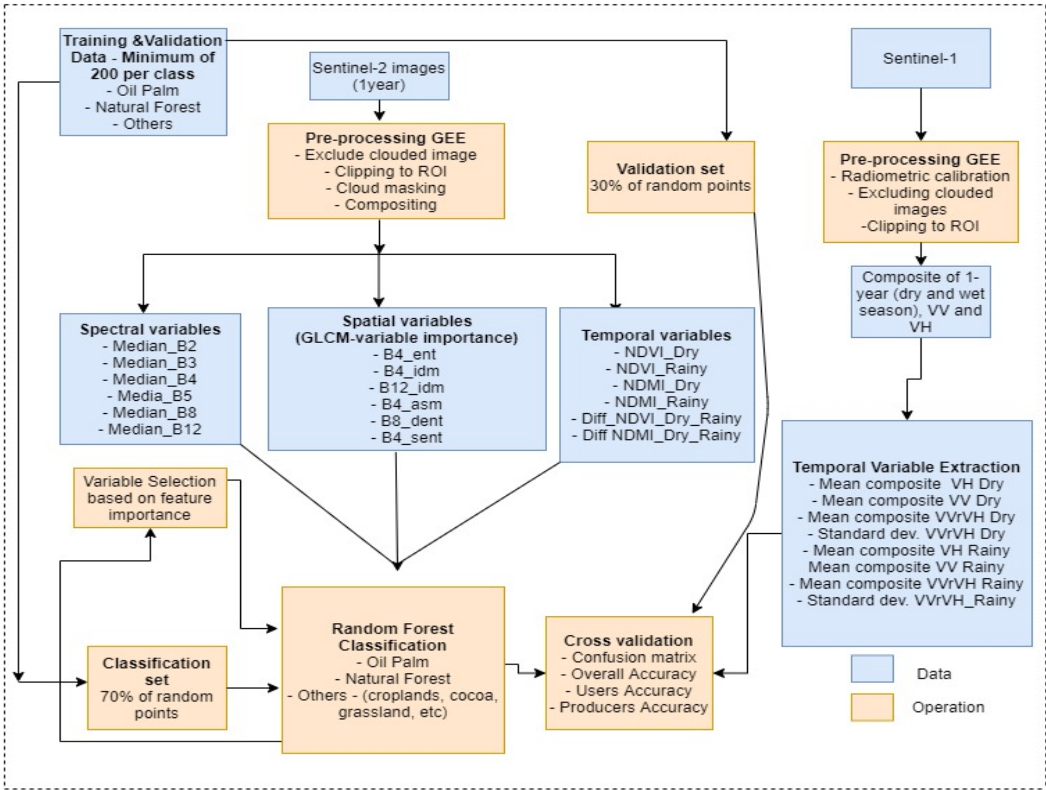

Fig. 2. Workflow for data acquisition, pre-processing, and validation

Fig. 3. Aerial imagery examples of land cover classes within the study area (Google, 2020c)

2.4 Training and Validation Data

To acquire reference samples from the study as training and validation data, we made use of existing landcover datasets, local studies, and satellite images. The land cover dataset used was the Forest cover data of the Copernicus Global Land Service (https://land.copernicus.eu/global/products/lc) which shows the global fraction of vegetation cover (FCover) corresponding to the fraction of ground covered by green vegetation. This helped to limit the extent of the study to only forest cover and aided the collection of the training and validation data. Other sources of reference for the data collection are described in Tables 1 and 2.

The forest cover data from the Copernicus Global Land Services was loaded within the google earth engine for visual assessment and interpretation of the land cover within the study area. Training and validation data were collected in Google Earth Engine (GEE) and overlaid the forest cover data. We collected about 10000 randomly generated sample points to ensure that the entire area was adequately covered. A buffer of 25 meters was specified around all the points generated to avoid placement of multiple random samples within the same pixel resulting in multiple representations of the same training or validation sample. About 960 points out of the 10000 were assessed and created by visual interpretation from google earth imagery and existing knowledge of the area to see which class they represent.

Points for three classes, (200 for oil palm, 380 for natural forest, and 380 for others) were collected. This was because increasing the number of classes for natural forests and others(croplands) gave a better representation of the overall spatial distribution of the mapped classes. The classes in question were labeled (Natural forest, Oil palm plantation, and others). The Others class is made up of croplands such as maize, cocoyam, plantain, cassava, citrus, cocoa, etc. Fig. 4.

2.5 Data Pre-processing and Variable Extraction

2.5.1 Sentinel-2 and sentinel-1 data pre-processing

Sentinel-2 data over a one-year (2018) period was acquired and pre-processed in GEE, where the images were provided as a Level-1C product that represents Top of Atmosphere (TOA) reflectance. The resolution of the images 10-meter for the shorter wavelengths and 20-or 60-meter resolution for longer wavelengths or narrower bandwidths. We selected only images with a cloud cover of less than 2%. As part of the filtering, the following bands which are considered important as far as vegetation analysis are concerned (B2, B3, B4, B5, B8, B12) representing (Blue, Green, Red, Red-Edge, NIR, SWIR) respectively were selected for the analysis. The selection of these bands stern from the fact that, with an increase in data dimension, the computational and storage costs will also be sharply increased.

According to [8], the red and NIR bands are usually regarded as important bands for forest cover estimation since the NIR band shows high reflectance for green vegetation due to its high internal leaf scattering, while the red bands show low reflectance due to chlorophyll absorption with the increase in forest vegetation cover. They further emphasize that the red-edge band has a high correlation with the various physiological vegetation parameters, such as nitrogen content, chlorophyll content, and biomass, hence also an important indicator to describe the status of plant pigments and health. The blue band has low reflectance over the vegetation canopy because of the strong absorption of chlorophyll but it is vital for vegetation monitoring using remote sensing data, making it important for vegetation analysis as well. Hence, justifying our selection of the Blue, Green, Red, Red-Edge, NIR, and SWIR bands.

The sentinel-1 Level-1 Ground Range Detected (GRD) data with a resolution of 10-m and the Interferometric Wide Swath (IW) instrument mode was also acquired and preprocessed in GEE. Image within the study area was acquired in the ascending mode, which means that the side-looking orientation of the sensor is always the same, namely west to east. The sentinel-1 data covered a full 1-year period of 2018 just as in the case of the Sentinel-2 data. This covers a cycle of both rainy and dry seasons (between April and October for rainy and between November and March for dry season). This selection resulted in a total of 38 sentinel-1 images each for both dry season and rainy season with a 3-day revisit time, which is enough for accurate classification. Some radiometric calibration, which computes backscatter intensity using sensor calibration parameters in the GRD metadata, also areas with extremely high and

Table 1. Definition of local studies used as input for the selection of training and validation samples

Author/reference	Description
[9]	Oil Palm Mapping using Support Vector Machine with Landsat ETM+ Data
[10]	Unsustainable Management of Forest in Ghana from 1900-2010
[11]	Effects of Tree-crop Farming on Land-cover Transitions in a Mosaic Landscape in the Eastern Region of Ghana

Table 2. Description of aerial images used as input for the selection of training and validation samples

Dataset	Description	Author/reference
Google Earth images	High-resolution global composition of aerial images and satellite images in different scales and from different dates	Google (2020c)
Microsoft Bing Maps	High-resolution global composition of aerial images and satellite images in different scales	(Microsoft Bing Maps, 2020)

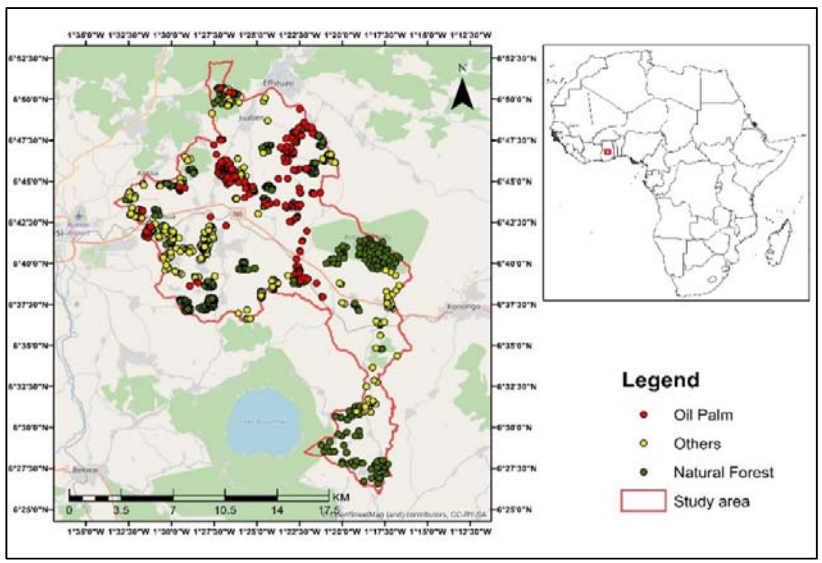

Fig. 4. Training and validation samples collected for classification

low incidence angles were excluded. This resulted in masking out the edges of all images. These images were then clipped to the extent of the study area and generated as a composite.

2.5.2 Spatial (textural) variable extraction from pre-processed sentinel-2 data

Spatial (textural) information relates to the structural features of the target surface and the surrounding environment, which can also reflect spatial variation in land cover. As such, the textural features can be extracted by statistical, structural, and spectral methods [12]. The potential for textural features from satellite images to be used in the identification of crops has been demonstrated to be significant [13]. In this research, the Gray level co-occurrence matrix (GLCM) method of textural feature extraction was used.

2.5.3 Gray level co-occurrence matrix (GLCM)

The GLCM is a classic method of texture (spatial) feature extraction, which is effective in image recognition, segmentation, retrieval, classification, and texture analysis method. It has been extensively employed in many fields and has been continuously improved [14]. The principal concept of GLCM is that the texture information contained in an image is defined by the adjacency relationships that the gray tones in an image have to one another. The matrix element P $(i, j \mid d, \Theta)$ contains the second-order statistical probability values for changes between gray levels i and j at a particular displacement distance d and a particular angle Θ. Instead of using the frequency values in a GLCM directly, it is common practice to normalize them to the range [0, 1] to avoid scaling effects. Resolution of the image, the land cover under consideration, and the scale of the different features. From various literature readings, if the window size is too small, then it does not contain enough information about the area to perform an accurate analysis, also if the size is too large, then it can overlap with other types of ground cover and produce erroneous results. Since the area of study has land cover types that are close to each other and for that matter, there is a possibility of these land cover overlapping, a medium window size was used. A table overview of GLCM equations is presented in Table 3 with their meanings.

2.5.4 Selection of spatial variables from GLCM computations

The variable importance test of the random forest model based on the training data was used to estimate the important variables. This resulted in a list of variables ranked according to their importance. The first six variables according to their mead decrease accuracies were then selected for further analysis to represent spatial variables. An overview of these variables is shown in Table 4.

2.5.5 Temporal variable extraction from sentinel-2 data

In this research, we acquired one year of sentinel-2 data over two seasons: dry and wet based on filtering by the dates for both dry and wet seasons in Ghana. This was provided by the Ghana Meteorological Agency (GMET) as shown in Fig. 4. Filtering based on a selection of the most important bands (Blue, Green, Red, Red-Edge, NIR, SWIR) as far as vegetation cover estimation is concerned was performed. Functions were created to mask out clouds, using Bits 10 and Bits 11 represented as clouds and circus respectively for both dry and wet seasons. From the preprocessed sentinel-2 data, mainly two vegetation indices were performed, namely, the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Moisture. Index (NDMI). These variables were then extracted based on the training data of the 960 points (Fig. 3) for both dry and rainy seasons.

2.6 Spectral Variable Extraction

Spectral features (median) of the selected bands were extracted from the sentinel-2 data in google earth engine. The median is another way to measure the center of a dataset, it is an attractive statistic to use for compositing. This is because the median can take care of outliers, thus, it is not easily affected by outliers. The table below shows the various bands for which their medians were extracted to represent the spectral variable.

2.6.1 Sentinel-1 variable extraction and variable importance

The mean values of VV and VH backscattering values for both rainy and dry season were extracted from the sentinel-1 data and used for the land cover classification, complemented with

Table 3. Overview of GLCM computations, meaning, and their equations

S1. No	Feature	Meaning	Equation
1	asm1, asm2, asm3, asm4, asm5, asm6,	Angular second moment	
2	contrast1,contrast2, contrast3,contrast4, contrast5,contrast6	Contrast	
3	corr1, corr2, corr3, corr4, corr5, corr6	Correlation	
4	var1, var2, var3, var4, var5, var6	Variance	; ;
5	idm1, idm2, idm3, idm4, idm5 idm6	Inverse Difference Moment	
6	savg1, savg2, savg3, savg4, savg5, savg6	Sum Average (Mean)	
7	svar1, svar2, svar3, svar4, svar5, svar6,	Sum Variance	
8	sent1, sent2, sent3, sent4, sent5, sent6	Sum Entropy	
9	ent1, ent2, ent3, ent4, ent5, ent6	Entropy	
10	dvar1, dvar2, dvar3, dvar4, dvar5, dvar6	Difference Variance	
11	dent1, dent2, dent3, dent4, dent5, dent6	Difference Entropy	
12	Imcorr1_1, imcorr1_2, imcorr1_3, imcorr1_4, mcorr1_5, imcorr1_6	Information Measures of Correlation 1	
13	Imcorr2_1, imcorr2_2, imcorr2_3, imcorr2_4, mcorr2_5, imcorr2_6	Information Measures of Correlation 2	
14	maxcorr1, maxcorr2, maxcorr3, maxcorr4, maxcorr5, maxcorr6	Maximal Correlation Coefficient	
15	diss1, diss2, diss3, diss4, diss5, diss6	Dissimilarity	
16	inertia1, inertia2, inertia3, inertia4, inertia5, inertia6	Inertia	
17	shade1, shade2, shade3, shade4, shade5, shade6	Shade	
18	Prom1, prom2, prom3, prom4, prom5, prom6	Prominence	

For detailed explanations of calculations, see Hall-Beyer (2007) and Gonzalez and Woods (1992).Pi,j is the probability of values I and j occurring in adjacent pixels in the original image within the window defining the neighborhood. I and j are the labels of the columns and rows (respectively) of the GLCM

Table 4. Overview of selected spatial variables

Variables	Meaning	Mean Decrease Accuracy
ent.3	Entropy of band 4 (red)	13.593365
idm.3	Inverse Difference Moment of band 4 (red)	13.493740
idm.6	Inverse Difference Moment of band 12 (SWIR)	13.297509
asm.3	Angular second moment of band 4 (red)	13.173867
dent.5	Difference Entropy of Band 8 (NIR)	13.159718
sent.3	Sum Entropy of band 4 (red)	12.894367
imcorr2.6	Information Measures of Correlation 2 of band 12(SWIR)	12.820696
dent.6	Difference Entropy of band 12 (SWIR)	12.151775
imcorr1.6	Information Measures of Correlation 1 of band 12(SWIR)	11.78822
sent.6	Sum Entropy of band 12 (SWIR)	11.601014
diss.5	Difference entropy of band 8 (NIR)	11.108438
idm.5	Inverse difference moment of band 8 (NIR)	10.819887
shade.5	Shade of Band 8 (NIR)	10.385652
shade.4	Shade of band 5 (red edge)	10.317684
imcorr2.3	Information Measures of Correlation 2 of band 8 (NIR)	9.930402
imcorr1.3	Information Measures of Correlation 1 of band 4 (red)	9.888173
asm.6	Angular second moment of band 12 (SWIR)	9.860717
sent.4	Sum Entropy of band 5 (red edge)	9.599164
dent.3	Difference Entropy of band 4 (red)	9.440915
ent.6	Entropy of band 12 (SWIR)	9.299110
var.5	Variance of band 8 (NIR)	9.257354
sent.5	Sum Entropy of band 8 (NIR)	9.089627
corr.5	Correlation of band 8 (NIR)	8.488786
shade.2	Shade of band 3 (green)	8.419940
idm.2	Inverse Difference Moment of band 3 (green)	8.356993
corr.3	Correlation of band 4 (red)	8.292002
imcorr1.1	Information measure of correlation 1 of band 2 (blue)	8.245279
svar.4	Sum Variance of band 5 (red edge)	8.193131
prom.6	Prominence of band 12 (SWIR)	7.966215
ent.4	Entropy of band 5 (red edge)	7.933559

Table 5. Overview of equations for NDVI and NDMI calculation

Index	Description	Equation
NDVI	A measure to indicate the occurrence of vegetation based on the normalized difference in NIR (band 8) and red (band 4) reflectance.	
NDMI	NDMI incorporates the difference between SWIR and NIR bands and is, therefore, more sensitive to moisture content rather than photosynthetic activity	

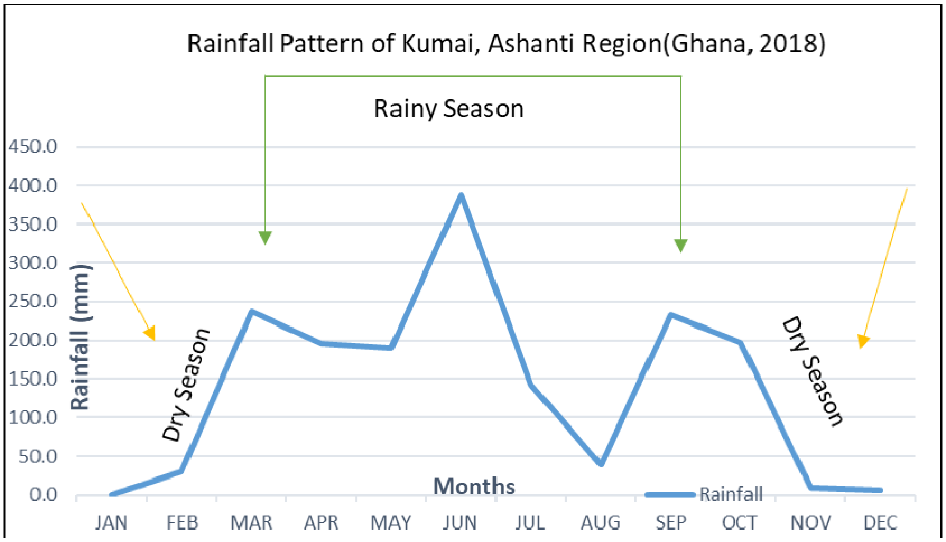

Fig. 5. Rainfall pattern within the study area for 2018

Table 6. Overview of used sentinel-2 temporal variables

Variable	Description
NDVI_dry	A normalized difference Vegetation index to indicate the occurrence of vegetation for the dry season period
NDVI_rainy	A normalized difference Vegetation index to indicate the occurrence of vegetation for the rainy season period
NDMI_dry	A measure to indicate the moisture content within the study area for the dry season period.
NDMI_rainy	A measure to indicate the moisture content within the study area for the rainy season period.
Diff_NDVI_Dry_Rainy	Difference between dry and rainy season of the Normalized Difference Vegetation Index
Diff_NDMI_Dry_Rainy	Difference between dry and rainy season of the Normalized Difference Moisture Index

Table 7. Overview of selected spectral variables for classification

Variables	Meaning
Median_B2	Median of band 2
Median_B3	Median of band 3
Median_B4	Median of band 4
Median_B5	Median of band 5
Median_B8	Median of band 8
Median_B12	Median of band 12

the ratio between VV and VH (VV/VH) as well as the standard deviation for VV and VH, both dry and rainy season. These variables in order of importance as generated with the random forest model are shown in Table 9. The first six variables were then selected to be included in the model created with sentinel-2 data to assess the added value of sentinel-1 to a sentinel-2 based classification of oil palm and natural forest.

2.6.2 Classification method, validation and confusion matrix

The classification was done within R studio (R Core Team, 2018) with Random Forest Algorithm. This was done by using the training data in constructing the decision tree. Two different classifications were performed using the random forest approach. The first classification was performed with variables extracted from the sentinel-2 data as shown in Table 8 whiles the second classification combined variables from sentinel-2 and sentinel-1 data (Table 10). Both classifications were done with 70% of the random sample points used as training data. The remaining 30% were used for validation. The splitting into training and validation was done randomly before the classification was performed. The default parameters for the Random Forest algorithm were maintained in the classification, the number of trees set to 500. To limit the tree depth and prevent the trees from overfitting, the minimum node was set to 5. These parameters were kept the same for all classifications. The random forest algorithm returns three variable importance measures. That is the selection rate of each candidate variable, the Gini index of impurity reduction, and the permutation of the predictor variables as an estimate of importance [15]. The mean decrease accuracy which is a measure of how much the accuracy decreases when a variable is excluded was used to generate this variable importance due to its simplicity. For both classifications with the sentinel-2 data and combining both sentinel-2 and sentinel-1 data, the confusion matrixes showing the overall accuracy, Kappa statistics, producer, and users' accuracy were generated.

3. RESULTS

3.1 Variable Importance: Spatial, Spectral, and Temporal Variables

Figs 6 and 7 show results from the variable importance plot of the random forest model of the top 10 most important variables as far as the spectral, spatial, and temporal variables are concerned. This is given regarding each class, oil palm, Natural forest, and the others class.

In Fig. 6, it is seen that the Median of band 5 (red edge), a spectral variable happens to be the most important variable, followed by spatial variable known as the Sum Entropy of band 5 (red edge), then another spectral variable of median band 4 for each class. The temporal variables appear not to be important as far as oil palm and natural forest characterization are concerned as shown by the results in Figs 6 and 7. Among the ten variables, the temporal variable appears once. Generally, the spectral and spatial variables are important as shown in Fig 5. The same appears in Fig 6, where the most important variable happens to be the spectral variable. However, the temporal variables appear to be a little of importance in this model as seen by the standard deviation for the VV and VH for both rainy and dry seasons and the ratio between the bands for rainy seasons also appearing in the variable importance diagram.

3.2 Land Cover Classification of Oil Palm and Natural Forest

Fig. 8 and 9 illustrate the land-cover classification results highlighting oil palm natural forest, and others (croplands) using the Random Forest model based on spatial, spectral, and temporal variables. Natural forest and other croplands appear to dominate in the study area accounting for approximately 35% and 50% respectively with oil palm only accounting for about 15% as seen in Fig. 9. The general trend from the classified maps shows that almost all oil palm plantations are located in the north-eastern part of the study area, and these are bothered by some natural forest as well as croplands and, most natural forests are located at the eastern portions of the area with few patches at the west and southern portions of the map. Most oil palm lands are also found within forest areas in the study location.

3.2.1 Validation of sentinel-2 and sentinel-1 classification

This study performed two separate validations, that is cross-validation of the model producing OOB error and K-fold validation. The cross-validation of the sentinel-2 classification model used 672 samples and 18 predictors for the three classes. The number of variables randomly sampled as candidates at each split (mtry) was 10. With mtry of 10, the average "in sample" accuracy was about 86%. Therefore, the accuracy of the random forest model is about 86% and the OOB estimate of error rate is 13.99%. The model with combine sentinel-2 and sentinel-1 data also produced quite different results. Again with 672 samples of the training data, 24 predictors for the three classes, cross-validation of 10 folds repeated 5 times produced an accuracy of the optimal model using the largest value. With mtry of 13, the accuracy was about 85%.

Table 11 provides the summary of land cover classification accuracies based on the sentinel-2 data, and Table 12 provides land cover classification accuracies with the combination of sentinel-2 and sentinel-1 data. The table shows quite different accuracy results for both classifications. Table 11 shows that the classification based on only sentinel-2 data yielded an overall classification accuracy of 74%. The individual class accuracies are shown in the table with oil palm recording a very low producer accuracy of 33% and a user accuracy of about 64%. Most of the oil palm classes are misclassified as natural forest and others. Table 12, indicates that combining sentinel-2 and sentinel-1 variables provides a better classification result. The classification procedure based on sentinel-2 and sentinel-1 data combined yields a significant improvement with overall accuracy increasing by 11% from 75% to 85%. There is also an improvement in the individual class accuracies with the accuracy of oil palm increasing from 65% to 82%. The combination of sentinel-2 and sentinel-1 data yielded a better accuracy compared to only sentinel-2 data.

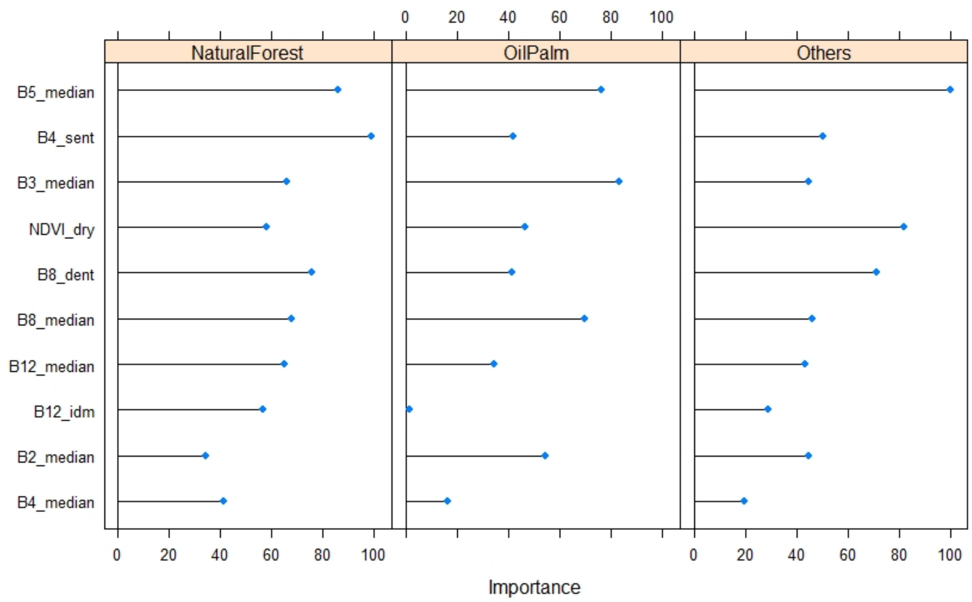

Fig. 6. Variable importance for the three classes using only sentinel-2 data

Table 8. All used variables for sentinel-2 classification

No	Variables	Variable type
1	B4_ent	spatial
2	B4_idm	spatial
3	B12_idm	spatial
4	B4_asm	spatial
5	B8_dent	spatial
6	B4_sent	spatial
7	Median_B2	spectral
8	Median_B3	spectral
9	Median_B4	spectral
10	Median_B5	spectral
11	Median_B8	spectral
12	Median_B12	spectral
13	NDVI_dry	temporal
14	NDVI_rainy	temporal
15	NDMI_dry	temporal
16	NDMI_rainy	temporal
17	Diff_NDVI_Dry_Rainy	temporal
18	Diff_NDMI_Dry_Rainy	temporal

Table 9. Overview of sentinel-1 extracted and used variables

Variables	Meaning	Mean Decrease in Accuracy
VV_VH_SDdry	Standard deviation between VV and VH of sentinel-1 data for the full 2018 period for the dry season.	46.80327
VV_VH_SD_rainy	Standard deviation between VV and VH of sentinel-1 data for the full 2018 period for the rainy season.	43.98977
VVrVH_mean_dry	The ratio between VV and VH for sentinel-1 data for the full 2018 period for the dry season.	43.66699
VVrVH_mean_rainy	The ratio between VV and VH for sentinel-1 data for the full 2018 period for the rainy season.	42.71625
VH_mean_rainy	Mean composite of sentinel-1data of VH for the full 2018 period for the rainy season	40.54810
VV_mean_rainy	Mean composite of sentinel-1data of VV for the full 2018 period for the rainy season	40.46791
VH_mean_dry	Mean composite of sentinel-1data of VH for the full 2018 period for dry season	39.39470
VV_mean_dry	Mean composite of sentinel-1data of VV for the full 2018 period for dry season	38.51086

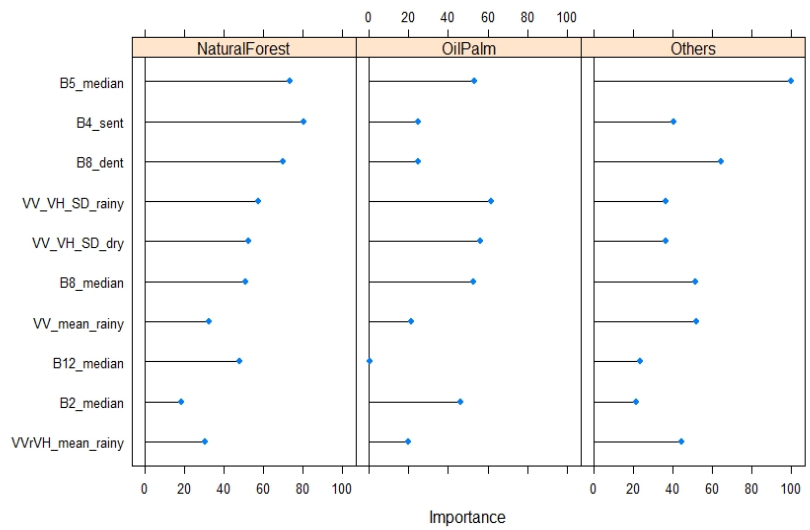

Fig. 7. Variable importance for the three classes using both sentinel-2 and sentinel-1 data

Table 10. Combined variables from sentinel-2 and sentinel-1 classification

No	Sentinel-2 Variables	Variable type
1	B4_ent	spatial
2	B4_idm	spatial
3	B12_idm	spatial
4	B4_asm	spatial
5	B8_dent	spatial
6	dent5	spatial
7	Median_B2	spectral
8	Median_B3	spectral
9	Median_B4	spectral
10	Median_B5	spectral
11	Median_B8	spectral
12	Median_B12	spectral
13	NDVI_dry	temporal
14	NDVI_rainy	temporal
15	NDMI_dry	temporal
16	NDMI_rainy	temporal
17	Diff_NDVI_Dry_Rainy	temporal
18	Diff_NDMI_Dry_Rainy	temporal
No	Sentinel-1 variables	Variable type
19	VV_VH_SDdry	temporal
20	VV_VH_SD_rainy	temporal
21	VVrVH_mean_dry	temporal
22	VVrVH_mean_rainy	temporal
23	VH_mean_rainy	temporal
24	VV_mean_rainy	temporal

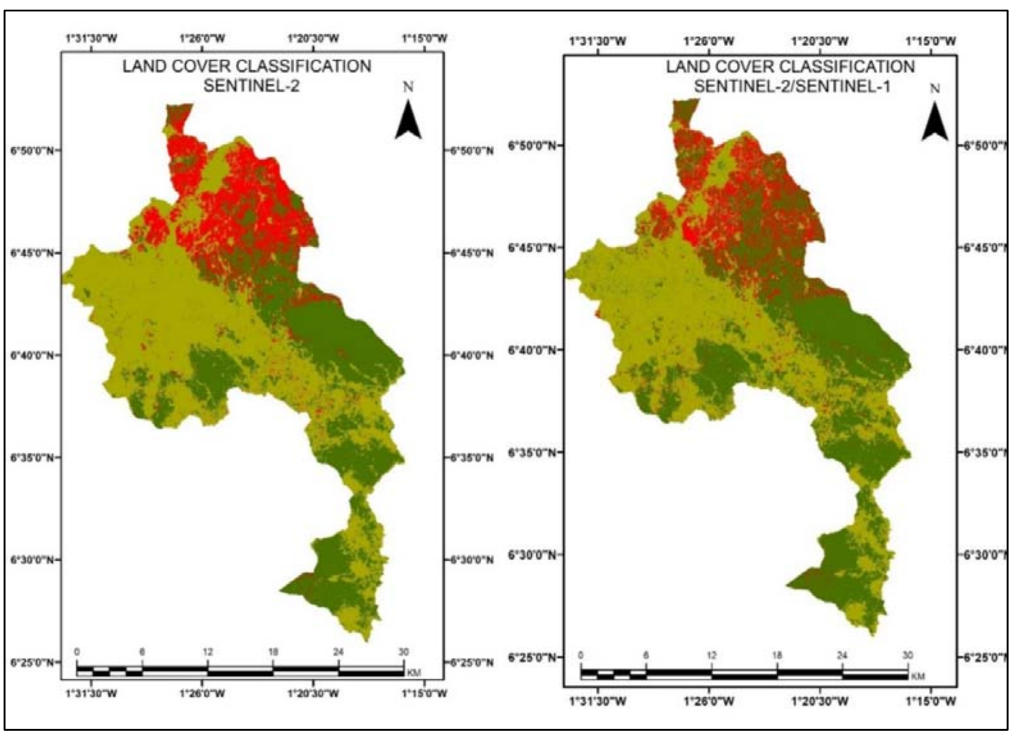

Fig. 8A. Classification results of oil palm, natural forest, and others with sentinel-2 data(left) and sentinel-2 combined with sentinel-1 data (right)

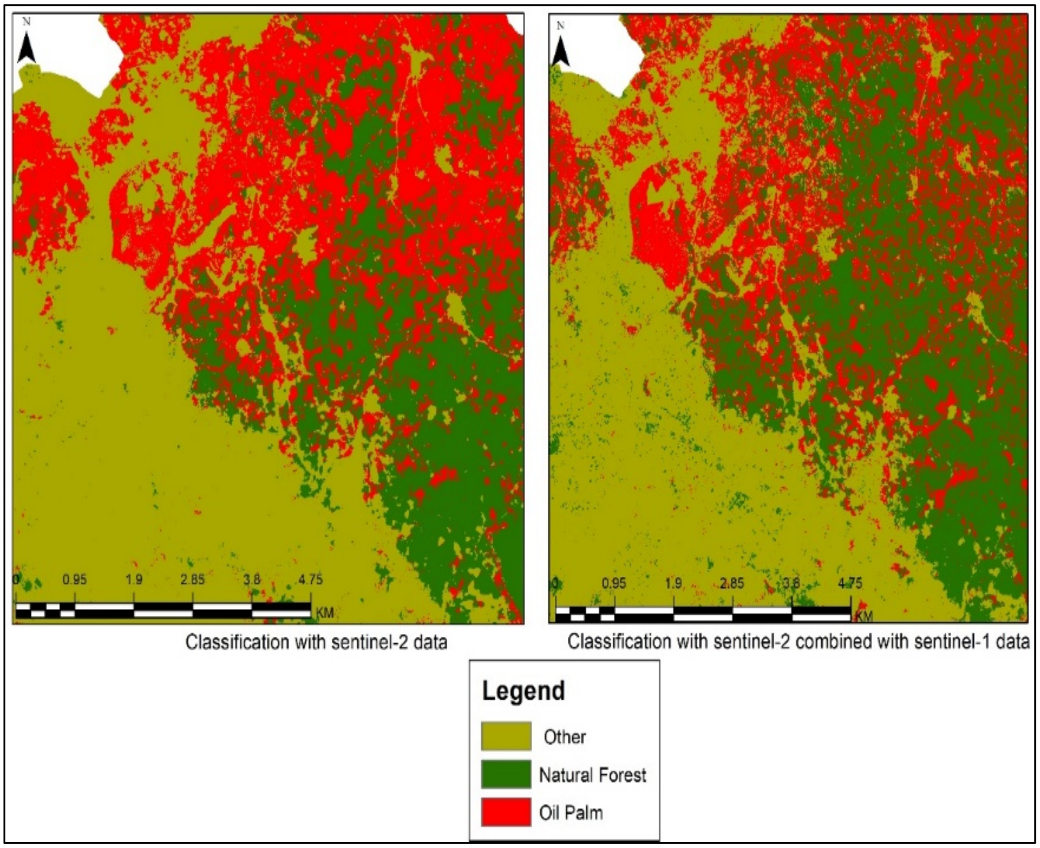

Fig. 8B. Areas of main differences in classification

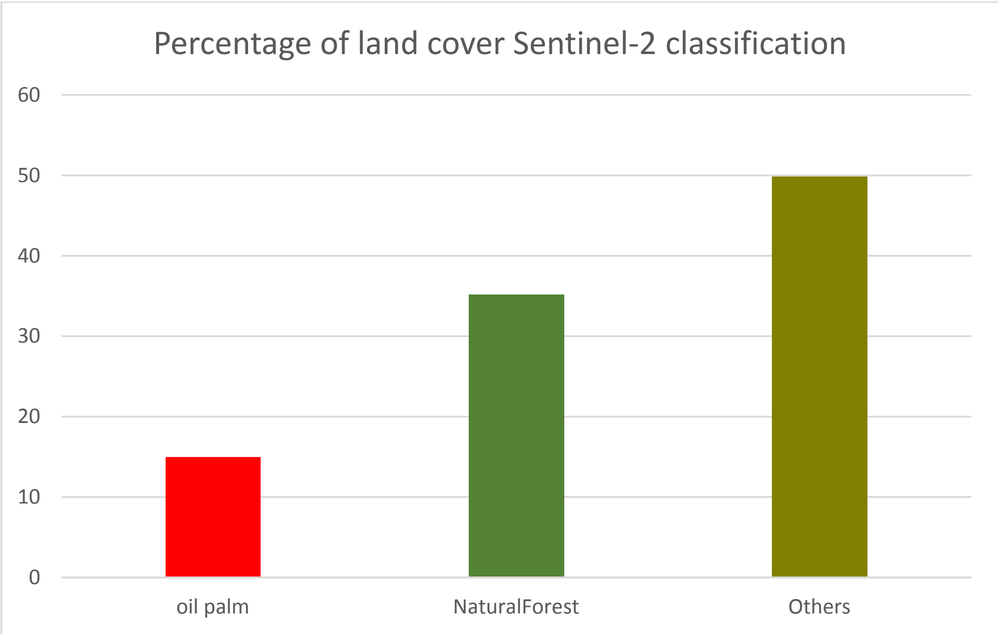

Fig. 9. Percentage area of land cover sentinel 2

Table 11. Confusion matrix of validation data (sentinel 2)

Classification	Reference					
		Oil palm	**NatualForest**	**Others**	**Total**	**User's Accuracy (%)**
	OilPalm	22	8	4	34	64.71
	NaturalForest	26	100	17	143	69.93
	Others	12	9	93	111	83.78
	Total	60	114	114	288	
	Producer's Accuracy (%)	33.66	87.72	81.58		
	Overall (%)					74.65

Table 12. Confusion matrix of classification using validation data (sentinel-2 and sentinel-1 data combined)

Classification	Reference					
		Oil palm	**NatualForest**	**Others**	**Total**	**User's Accuracy(%)**
	OilPalm	42	5	3	51	82.35
	NaturalForest	9	102	10	121	84.10
	Others	9	7	101	117	86.32
	Total	60	114	114	288	
	Producer's Accuracy (%)	70.00	89.47	88.60		
	Overall (%)					85.10

4. DISCUSSION

Variable importance plays an important role to understand the contribution of an individual feature or group of individual features in the classification task [16]. Variable importance was therefore computed for the random forest classification of both sentinel-2 and sentinel-1 data.

The spectral information is often the most important variable in land-cover classification, especially for medium and coarse spatial resolution images [17]. Another study conducted by Vuolo et al., 2018 found that mono-temporal sentinel-2 red-edge and SWIR were important for mapping both tree species and crop types.

This was confirmed in the study as shown in Figs 5 and 6 that the Median band 5 (red edge) was the most important spectral variable. Another important spectral variable that showed up in the results was also the median of band 8 (NIR). This can be attributed to the presence of the red edge band. In research conducted by [18], where they tested all sentinel-2 band combinations and evaluated the significance of each band by counting the frequency each band appeared in the top 5% of the band combinations. They concluded the importance of spectral regions as NIR, red edge, visible, and SWIR.

However, in high-resolution images, spatial information becomes an important feature in improving land-cover classification [18]. This study also confirmed the importance of spatial features, especially for the separation of natural forest from oil palm (Fig. 6). That is the Sum Entropy of Band 5 (red edge) and the Difference Entropy of Band 12 (SWIR). Texture or spatial variables have often been extracted using GLCM based on either a spectral image or a fixed window size such as 5 X 5. And the response of using texture variables depends on specific land cover types and patch sizes. They may therefore be effective for some land-cover types but not good for others [19] For this research since the study area is made up of land cover types that are close to each other, a fixed window size of 5 X 5 was used to ensure accurate results. Kapidura, [20], in his work comparing the different methods of textural analysis for their efficiency for forest classification, concluded that the combination of spectro-textural classification-based approach produces a good accuracy, hence should be considered in the characterization of natural and plantation forests.

In another study conducted by [21] to demonstrate the performance of sentinel-2 in forest mapping, accurate discrimination of the forest was arrived at and attributed mainly to the high spatial resolution, in this case, spatial variables available from the 10m sentinel-2 bands and the capacity of sentinel-2 to include red-edge bands which also agrees with the results from this research.

The research also identified the contribution of sentinel-1 in general to the sentinel-2 classification of plantation and natural forest. As represented in Fig 7, the sentinel-1 variables such as the standard deviation between VV and VH of the sentinel-1 data for the full 2018 period for the dry season and standard deviation between VV and VH of sentinel-1 data for the full 2018 period for the rainy season were shown as important variables in the overall variable importance for the characterization of plantation and natural forest. This is explained by the inclusion of the sentinel-1 data for classification purposes. Data from sentinel-1 were demonstrated to be useful for land cover classification as a possibility to complement the cloud-coverage areas [22]. These results also confirm the work of [22] where sentinel-1 data was used to show results of delineated forest areas in an Australian study site.

In literature, overall accuracy derived from the confusion matrix and kappa coefficient has been widely used for evaluation purposes [16]. In this work, for the estimation of accuracy of the classification, Overall accuracy and Kappa coefficient were computed by setting the parameters for testing. Combining sentinel-2 and sentinel-1 data improved the results with an overall accuracy increased from 75% to 85% as well as Kappa Coefficients also increasing from 0.6 to 0.7 compared to sentinel-2 only. This is confirmed by a study conducted by [23] where the combination of sentinel-2 and sentinel-1 data helped improve classification accuracies. A ground-truth validation of the sample points collected was also done to accurately classify the land cover into their respective classes.

In terms of classification output, the research produced a LULC map that is of good quality with a higher resolution (10m). Also, while most previous LULC maps of Ghana categorized forest and plantation as one class, this map has three different classes (natural forests, oil palm plantation, and others which represent croplands such as cocoa, maize, rubber, cassava, among others). This can offer better support for forest monitoring initiatives such as REDD+, LULC change, national forest inventory, etc.

5. CONCLUSION

The study utilized remote sensing data to create a high-resolution LULC map that aimed to distinguish natural forest and plantation forest (oil palm) in Africa, with Ghana as a case study. This approach comprised of using only sentinel-2 data for classification first and then combining sentinel-2 with sentinel-1 data.

In characterizing natural and plantation forests the research decided to utilize all three data variables (spatial, spectral, and temporal variables) in assessing the most important variable to be considered for such characterization. The spectral variables followed by spatial variables are most important and should be considered when characterizing plantation and natural forest. The use of only one or two of the three variables is good for oil palm characterization, however, including all spatial, spectral, and temporal variables gives a very detailed analysis of land cover characterization and increases the accuracy. Combining sentinel-2 and sentinel-1 data improved the results with an overall accuracy increased from 75% to 85% as well as Kappa Coefficients also increasing from 0.6 to 0.7 compared to sentinel-2 only. The study, therefore, confirms the feasibility of producing accurate LULC maps in the era of big data or Earth observation. There is also the need to produce such maps in historical periods to have spatially explicit insights into the constraints of plantation and natural forest dynamics with socio-economic and policy backgrounds in Ghana.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Curtis PG, Slay CM, Harris NL, Tyukavina A, Hansen MC. Classifying drivers of global forest loss. Science. 2018;361(6407):1108–1111. Available:https://doi.org/10.1126/science.aau3445
2. Barlow J, Lennox GD, Ferreira J, Berenguer E, Lees AC, Nally R. Mac, Thomson JR, Ferraz S FDB, Louzada J, Oliveira VHF, Parry L. Ribeiro De Castro Solar R, Vieira ICG, Aragaõ LEOC, Begotti RA, Braga RF, Cardoso TM, Jr RCDO, Souza CM, Gardner TA. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. Nature. 2016;535(7610):144–147. Available:https://doi.org/10.1038/nature18326
3. Hoang TT, Truong VT, Hayashi M, Tadono T, Nasahara KN. New JAXA high-resolution land use/land cover map for Vietnam aiming for natural forest and plantation forest monitoring. Remote Sensing. 2020;12(17):1–25. Available:https://doi.org/10.3390/RS12172707
4. Xiao C, Li P, Feng Z. Monitoring annual dynamics of mature rubber plantations in Xishuangbanna during 1987-2018 using Landsat time-series data: A multiple normalization approach. International Journal of Applied Earth Observation and Geoinformation. 2019;77:30–41. Available:https://doi.org/10.1016/j.jag.2018.12.006
5. Truong, Van Thinh, Thanh Tung Hoang, Duong Phan Cao, Masato Hayashi, Takeo Tadono, K. NN. JAXA annual forest cover maps for vietnam auxiliary data. Remote Sensing, 2019;11(20):2412.
6. Gutiérrez-Vélez VH, DeFries R. Annual multi-resolution detection of land cover conversion to oil palm in the Peruvian Amazon. Remote Sensing of Environment. 2013;129:154–167.
7. Torbick N, Ledoux L, Salas W, Zhao M. Regional mapping of plantation extent using multisensor imagery. Remote Sensing. 2016;8(3). Available:https://doi.org/10.3390/rs8030236
8. Wang B, Jia K, Liang S, Xie X, Wei X, Zhao X, Yao Y, Zhang X.. Assessment of Sentinel-2 MSI spectral band reflectances for estimating fractional vegetation cover. Remote Sensing. 2018;10(12):1–20. Available:https://doi.org/10.3390/rs10121927
9. Kwesi NI. Oil palm mapping using support vector machine with Landsat etm + data. 2012;1–68.
10. Emmanuel Quacou I. Unsustainable Management of Forests in Ghana from 1900-2010. International Journal of Environmental Monitoring and Analysis,

2016'4(6):160.
Available:https://doi.org/10.11648/j.ijema.20160406.14
11. Asubonteng K, Pfeffer K, Ros-Tonen M, Verbesselt J, Baud I. Effects of tree-crop farming on land-cover transitions in a mosaic landscape in the eastern region of Ghana. Environmental Management. 2018;62(3):529–547.
Available: https://doi.org/10.1007/s00267-018-1060-3
12. Villa P, Bresciani M, Bolpagni R, Pinardi M, Giardino C. A rule-based approach for mapping macrophyte communities using multi-temporal aquatic vegetation indices. Remote Sensing of Environment. 2015;171:218–233.
Available:https://doi.org/https://doi.org/10.1016/j.rse.2015.10.020
13. Sun C, Bian Y, Zhou TJP. Using of multi-source and multi-temporal remote. remote, multi-temporal region, subtropical agriculture. 2019;19:1–23.
Available: https://www.mdpi.com/1424-8220/19/10/2401/htm
14. Zhang X, Cui J, Wang W, Lin C. A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm. Sensors. 2017;17(7):1474.
15. Adjorlolo C, Cho MA, Mutanga O, Ismail R. Optimizing spectral resolutions for the classification of C3 and C4 grass species, using wavelengths of known absorption features. Journal of Applied Remote Sensing. 2012;6(1):063560–063561.
Available:https://doi.org/10.1117/1.jrs.6.063560
16. Saini R, Ghosh SK. Crop classification on single date sentinel-2 imagery using random forest and suppor vector machine. *XLII*(November). 2018;20–23.
17. Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing. 2007;28(5):823–870.
Available:https://doi.org/10.1080/01431160600746456
18. Richter K, Hank TB, Vuolo F, Mauser W, D'Urso G. Optimal exploitation of the sentinel-2 spectral capabilities for crop leaf area index mapping. Remote Sensing. 2012;4(3):561–582.
Available:https://doi.org/10.3390/rs4030561
19. Lu, Dengsheng, Li G, Moran E, Dutra L, Batistella M. The roles of textural images in improving land-cover classification in the Brazilian Amazon. International Journal of Remote Sensing. 2014;35(24):8188–8207.
Available:https://doi.org/10.1080/01431161.2014.980920
20. Kupidura P. The comparison of different methods of texture analysis for their efficacy for land use classification in satellite imagery. Remote Sens. 2019;11.
21. Puletti N, Chianucci F, Castaldi C. Use of Sentinel-2 for forest classification in Mediterranean environments. Ann. Silvic. Res. 2018;42:32–38.
22. Balzter H, Cole B, Thiel C, Schmullius C. Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests. Remote Sens. 2015;7:14876–14898.
23. Hasan S, Shi W, Zhu X, Abbas S. Monitoring of land use/land cover and socioeconomic changes in South China over the last three decades using landsat and nighttime light data. Remote Sens. 2019;11:1–23.
Available:https://doi.org/10.1016/j.rse.2012.10.033

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/69347