

Article

Decision Tree Algorithms for Developing Rulesets for Object-Based Land Cover Classification

Darius Phiri ^{1,*}, Matamyo Simwanda ¹, Vincent Nyirenda ² , Yuji Murayama ³  and Manjula Ranagalage ^{3,4} 

¹ Department of Plant and Environmental Sciences, School of Natural Resources, Copperbelt University, P.O. BOX 21692 Kitwe, Zambia; matamyo.simwanda@cbu.ac.zm

² Department of Zoology and Aquatic Sciences, School of Natural Resources, Copperbelt University, P.O. BOX 21692 Kitwe, Zambia; vincent.nyirenda@cbu.ac.zm

³ Faculty of Life and Environmental Sciences, University of Tsukuba, 1-1-1, Tennodai, Tsukuba, Ibaraki 305-8572, Japan; mura@geoenv.tsukuba.ac.jp (Y.M.); manjularanagalage@ssh.rjt.ac.lk (M.R.)

⁴ Department of Environmental Management, Faculty of Social Sciences and Humanities, Rajarata University of Sri Lanka, Mihintale 50300, Sri Lanka

* Correspondence: darius.phiri@cbu.ac.zm

Received: 24 March 2020; Accepted: 15 May 2020; Published: 19 May 2020



Abstract: Decision tree (DT) algorithms are important non-parametric tools used for land cover classification. While different DTs have been applied to Landsat land cover classification, their individual classification accuracies and performance have not been compared, especially on their effectiveness to produce accurate thresholds for developing rulesets for object-based land cover classification. Here, the focus was on comparing the performance of five DT algorithms: Tree, C5.0, Rpart, Ipred, and Party. These DT algorithms were used to classify ten land cover classes using Landsat 8 images on the Copperbelt Province of Zambia. Classification was done using object-based image analysis (OBIA) through the development of rulesets with thresholds defined by the DTs. The performance of the DT algorithms was assessed based on: (1) DT accuracy through cross-validation; (2) land cover classification accuracy of thematic maps; and (3) other structure properties such as the sizes of the tree diagrams and variable selection abilities. The results indicate that only the rulesets developed from DT algorithms with simple structures and a minimum number of variables produced high land cover classification accuracies (overall accuracy > 88%). Thus, algorithms such as Tree and Rpart produced higher classification results as compared to C5.0 and Party DT algorithms, which involve many variables in classification. This high accuracy has been attributed to the ability to minimize overfitting and the capacity to handle noise in the data during training by the Tree and Rpart DTs. The study produced new insights on the formal selection of DT algorithms for OBIA ruleset development. Therefore, the Tree and Rpart algorithms could be used for developing rulesets because they produce high land cover classification accuracies and have simple structures. As an avenue of future studies, the performance of DT algorithms can be compared with contemporary machine-learning classifiers (e.g., Random Forest and Support Vector Machine).

Keywords: land use; remote sensing; spectral mixture analysis; change detection; optical images; Africa

1. Introduction

Object based image analysis (OBIA) has become an effective method of land cover classification of remotely sensed data [1,2]. Unlike traditional pixel-based analysis, OBIA offers an opportunity to develop discrete objects which relate to real world objects through image segmentation [3,4].

The segmentation process reduces within class spectral variations and offers an opportunity to increase classification accuracy, especially when conducted at an appropriate scale [5,6].

The ability to incorporate texture, compaction, and other object related information with spectral information has differentiated OBIA from other methods of classification such as pixel and sub-pixel approaches. Compared to pixel-based image analysis, OBIA is effective in reducing salt-and-pepper effects on thematic maps [7,8]. There are currently a number of segmentation algorithms available in eCognition Developer; however, multiresolution segmentation is the most common method used in land cover classification [1,4,9].

Apart from segmentation, another important component of OBIA is the actual classification of segmented objects [1,2]. Myint et al. [10] explained that there are two ways of assigning classes to segmented objects: (1) employing expert knowledge through rulesets; and (2) using automated classifiers. Under expert knowledge, classification is done by developing rulesets which are based on the thresholds of different object related information. Under the automated classification approach, objects are classified based on contemporary machine-learning classifiers such as Nearest Neighbor (NN), Random Forest (RF), Support Vector Machine (SVM), and classification and regression tree (CART), which have been incorporated into eCognition Developer 9.1 [2,11]. However, the practice of developing rulesets using thresholds of different object-related information remains a common practice in OBIA land cover classification [1,12].

There are many ways of establishing thresholds during ruleset development such as using expert knowledge, trial-and-error, and using binary recursive decision trees (DTs) [10]. Although not a common practice, the implementation of DTs seems more formal in establishing thresholds and the eventual development of effective rulesets [2,3,13]. Here, it is important to note that the statistical packages that have been used for developing rulesets are referred to as DT algorithms [14]. These algorithms are generally referred to as “black box” or “white box” depending on how easily an interpreter can follow the process. Black box algorithms, such as RF, have been used extensively in land cover classification, especially with the advancements in machine learning techniques [15]. On the other hand, simple machine learning DT algorithms such as Rpart, C5.0, and Tree have also been used for land cover classification and establishment of thresholds when developing rulesets [13].

Simple DTs are useful tools for establishing thresholds for developing decision rules for land cover classification of remote sensing data because they are non-parametric and are easy to interpret [10,16]. During classification, many variables are generated based on spectral and textural object-related information which can be used to develop effective rulesets if appropriate techniques such as DTs are applied. Belgiu et al. [17] suggested that DTs can be helpful for selecting the most influential variable and identifying the thresholds for different land cover classes because they are non-parametric and hence ideal for most landscapes.

While decision tree algorithms have been used in different areas associated with land cover classification, these algorithms have not been individually assessed on their effectiveness in establishing thresholds for developing rulesets for OBIA land cover classification. The main aim of this study was to conduct a multiple criteria evaluation of five different machine learning DT algorithms based on their performance when classifying Landsat 8 images. The performance comparison focused on the effectiveness of these five algorithms on handling different sizes of data and how simple each algorithm is to interpret.

2. Material and Methods

2.1. Study Site

The study was conducted on the Copperbelt Province of Zambia (Figure 1) which is located in the northern part of the country (latitude: 12.82° S, longitude: 28.21° E). The area receives between 1000 and 1200 mm of rainfall per annum and experiences temperature ranging 7–35 °C [18,19]. Mining and agriculture are the major economic activities in this area. This area is highly urbanized and has a

population density (62.5 persons per square kilometer) [18,20]. As a result of the non-productivity of the mines, most people practice small-scale shifting cultivation causing rapid land cover change, especially the conversion of forest areas to agriculture and settlements. The Copperbelt Province also has the largest proportion of forest plantations in Zambia that are owned by a parastatal company called Zambia Forest and Forestry Company (ZAFFICO) [18].

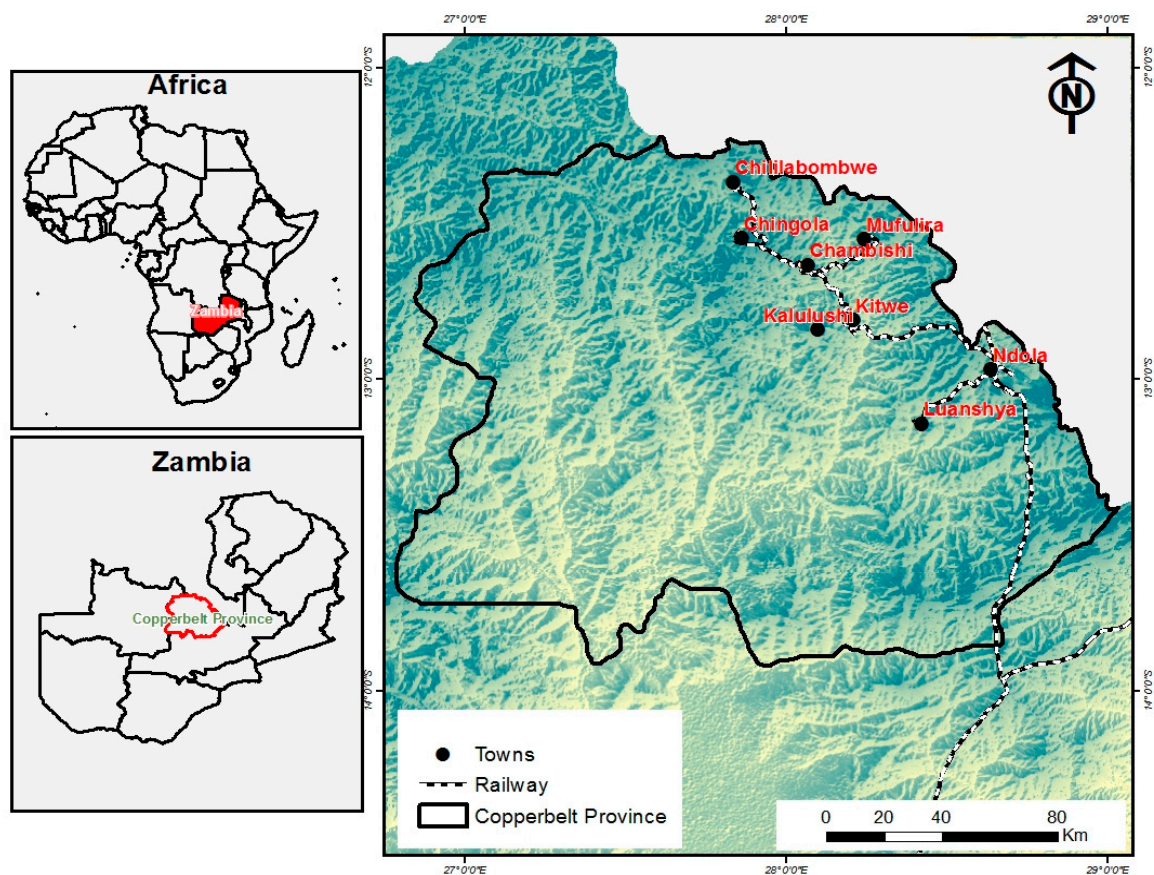


Figure 1. Administrative boundaries of Copperbelt Province of Zambia including rivers and the major towns.

2.2. Datasets

Landsat 8 images, also called Landsat observation land images (OLI) [21], acquired from the United States Geological Survey (USGS) website (<http://glovis.usgs.gov>) were used in this study. The images were taken in 2016 and the September images were selected because, during this period, the study area experiences a dry season (i.e., no rains) and hence has less cloud cover. Landsat 8 has a spatial resolution of 30 m, spectral resolution of 11 bands, temporal resolution of 16 days, and radiometric resolution of 12 bits. In addition, Landsat 8 images have a panchromatic band with a spatial resolution of 15 m [22,23]. In this study, six bands that range from visible to infrared were used. Apart from the Landsat images, the Shuttle Radar Topography Mission (SRTM) Digital Elevation Models (DEM), with a spatial resolution of 30 m, were used for pre-processing and as auxiliary data for classification.

In total, 2600 random points were randomly overlaid on the Landsat images using ArcGIS 10.4 software package [24]. Visual and prior expert knowledge of different land cover classes was used to assign classes to each of the 2600 points. The 2600 random points were assigned one of the 10 land cover types identified on the ground (Table 1 and Table S1). This data was separated into DT training (1000 sample points), land cover classification (1000 sample points), and accuracy assessment sample

for both the DT algorithms and the land cover map (600 sample points). The dataset was distributed to each land cover class following the percentage area covered by each class (see Table 1).

Table 1. Description of the 10 land cover classes used in this study.

No.	Land Cover Class	Description	Area of LC (%)	Training	Land Cover Classification	Validation
1	Bare land	Areas without any vegetation such as rocks and sandy areas	5.00	50	50	30
2	Dry Agriculture	Harvested areas with little green vegetation	6.16	62	62	37
3	Grassland	Areas which are dominated by grass and small shrubs	15.90	159	159	95
4	Irrigated Crops	Areas under irrigated systems such as pivot centers	6.02	60	60	36
5	Plantation Forests	Exotic forests areas	6.03	60	60	36
6	Primary Forests	Undisturbed or intact natural forests	23.46	235	235	141
7	Secondary Forests	Natural forests which are/were disturbed	20.24	202	202	121
8	Settlement	Built-up areas	5.54	55	55	33
9	Waterbodies	Lakes, rivers, and dams	5.00	50	50	30
10	Wetlands	Vegetation around water bodies	5.98	60	60	36
			100	1000	1000	600

Note that there were 1000 independent training samples for the DT and 1000 for land cover classification, while there were 6000 validation sample points for both DTs algorithm and land cover maps. The total number of samples was 2600.

2.3. Methods

2.3.1. Pre-processing

Pre-processing included the correction of images from atmospheric effects and topographic variation. This process converts digital numbers (DN) into ground reflectance values, which are more useful for image analysis. To ensure consistency during analysis, all images were projected to the Universal Transverse Mercator (UTM) projection system Zone 35S and World Geodetic System 84 (WGS 84) datum. Automated ATCOR 3, available in PCI Geomatics (PCI Geomatics, Ontario, Canada), was used for haze removal, atmospheric correction, and topographic correction by incorporating a 30 m digital elevation model.

2.3.2. Image Segmentation

Segmentation creates spectrally homogenous objects which can be related to real objects on the ground [25,26]. Past research recognized the challenges in establishing the optimal segmentation parameters [2,5]. Thus, the segmentation parameters which are scale (Sc), shape (Sh) and compaction (Cm) are commonly established by using trial-and-error methods [4,25,27]. Drăguț et al. [5] proposed a formal method of establishing optimal levels of scale factors using an Estimation of Scale Parameter (ESP) tool. For this study, the ESP tool indicated 12 for Sc, 0.2 for Sh, and 0.8 for Cm. With these scale parameters, multiresolution algorithm in eCognition Developer 9.1 (Trimble Navigation Ltd., Sunnyvale, California) was used to segment the images into spectrally homogeneous objects.

2.3.3. Sample Selection and Feature Extraction

After segmentation, the next step was to select sample objects using the 1000 random training points and extract object-related information. Several object-related feature values were developed based on spectral indices (Table 2), DEM values, spectral values of each band, and grey level co-matrix (GLCM). This process was done in eCognition Developer 9.1 by using the “assign class by thematic layer” and “classified image object to sample” tools. The samples and extracted object-related information were then exported to a spreadsheet. To assess the performance of the DT algorithms in land cover classification, the DTs’ accuracies and the classification accuracy of the final thematic maps were considered. It is important to note that we refer to the accuracy derived from cross-validation of the DTs after training as DT accuracy while the final accuracy of thematic maps, which was derived from accuracy assessment, is referred to as land cover classification accuracy or thematic map accuracy.

Table 2. Selected spectral indices used for developing DT rulesets.

Spectral Indices	Formula	Common Application(s)	References
Normalized Difference Vegetation Index (NDVI)	$ND \frac{NIR-Red}{NIR+Red}$	Measure density, greenness, and health of vegetation	Liao et al. [28]; Zhu et al. [29]
Enhanced Vegetation index (EVI)	$EVI = \frac{2.5 * (NIR-Red)}{(NIR+6*Red-7.5*Blue+1)}$	Corrects soil background signals and reduce atmospheric effects	Huete et al. [30]
Green Normalized Difference Vegetation Index (GNDVI)	$VI = \frac{NIR-Green}{NIR+Green}$	Similar to NDVI, but more sensitive to chlorophyll	Gitelson [31]
Green Ratio Vegetation Index (GRVI)	$GR \frac{NIR}{Green}$	Discriminating vegetation canopy based on level of photosynthesis	Sripada et al. [32]
Leaf Area Index (LAI)	$LAI = (3.618 * EVI - 0.118)$	Estimation of foliage cover and productivity	Atzberger et al. [33]
Simple Ratio (SR)	$SR = \frac{NIR}{Red}$	Used just as NDVI	Birth et al. [34]
Non-Linear Index (NLI)	$NLI = \frac{NIR^2-Red}{NIR^2+Red}$	Assumes non-linear relationship of vegetation parameters	Goel et al. [35]
Optimized Soil Adjusted Vegetation Index (OSAVI)	$OSAV = \frac{1.5 * (NIR-Red)}{NIR+Red+0.16}$	Used for soil variation from low vegetation cover	Rondeaux et al. [36]
Soil Adjusted Vegetation Index (SAVI)	$SAVI = \frac{(NIR-Red)}{(NIR+Red)} (1+L); L = 0.5$	Analyze soil and vegetation relationship	Huete [37]
Renormalized Difference Vegetation Index (RDVI)	$RDVI = \frac{(NIR-Red)}{\sqrt{NIR-Red}}$	Used to indicate vegetation health and productivity	Roujean et al. [38]
Normalized Burn Ratio (NBR)	$NBR = \frac{(NIR-SWIR)}{(NIR+SWIR)}$	Monitoring burnt areas in large areas	Key et al. [39]; Garcia et al. [40]
Ferrous Minerals Ratio	$FMR = \frac{SWIR}{NIR}$	Indicates iron bearing surfaces	Segal [41]
Iron Oxide Ratio (IOR)	$IOR = \frac{Red}{Blue}$	Indicates rocks that have been subjected to oxidation	Segal [41]
Normalized Difference Built-Up Index (NDBI)	$NDBI = \frac{SWIR-NIR}{SWIR+NIR}$	Detections of urban areas	Zha et al. [42]
Normalized Difference Snow Index (NDSI)	$NDSI = \frac{Green-SWIR1}{Green+SWIR1}$	Snow cover detection	Salomonson et al. [43]
Ratio vegetation Index (RVI)	$RVI = \frac{Red}{NIR}$	An inverse of the simple ratio	Silleos et al. [44]
Specific leaf area vegetation index (SLAVI)	$SLAVI = \frac{NIR}{Red+NIR}$	Estimations of foliage cover and productivity	Silleos et al. [44]
Normalized difference Water index (NDWI)	$NDWI = \frac{Green-NIR}{Green+NIR}$	Water detection	Gao [45]

2.3.4. Decision Tree Algorithms

DTs have been used in image-based classification because they are non-parametric and can be interpreted easily [13,14,46]. In OBIA, establishing decision rulesets is an important step towards land cover classification. However, this stage requires thresholds related to classes, which can be established by using knowledge-based methods or simple DTs. The knowledge-based approach can be complex,

especially when many land covers and decision variables are involved. Here, the focus was on binary recursive DTs, which use response variables to split trees until there is no possibility of further splitting. The performance of five DT algorithms (Table 3), Rpart, Tree, Party, C5.0, and Ipred, was assessed in this study using a multiple criteria approach [47]. The assessment included three components: (1) assessing the accuracy of DTs accuracy on clustering of training data; (2) assessing the accuracy of land cover classification; and (3) examining the simplicity (e.g., Tree diagram and number of variables) of the structure of the DTs.

Table 3. Description of DT algorithms used in the study.

No.	Name	Description
1	Rpart	Recursive partitioning for classification and regression (CART)
2	Party	Condition classification and regression
3	Tree	Classification and regression (CART) tree report misclassification
4	C5.0	Boosting and bagging decision tree and rule based for pattern recognition
5	Ipred	Involves bagging and resampling in classification

2.3.5. Assessing DT Accuracy

The DT algorithms were trained using the information sample extracted from the samples. After training the DT algorithms, an independent dataset (600 points) was used to cross-validate the DTs which were produced during training. The comparison was made by using the predicted results from the DTs and the independent data, which was used as reference data. Accuracies in terms of percentage were derived from the comparison for all the five DT algorithms. A comparison was made on the accuracies of the DTs when different sample sizes were used to train the DT algorithms. The samples (i.e., 1000 sample points) were divided into ten samples in multiples of 100. Thus, the smallest sample size was 100, and the largest sample size was 1000 (Table 4). These samples were established by randomization and selecting the specific number of samples. The independent samples were then used to assess the accuracy of DTs over increased sample sizes. A Kruskal–Wallis non-parametric test was then used to assess the significant difference among the DT accuracies of the five algorithms because of the limited classification attempts and non-normality of the data as suggested by Li et al. [2].

Table 4. Samples used in assessing the DT accuracy based on an increasing sample size.

Sample No.	Sample Size
1	100
2	200
3	300
4	400
5	500
6	600
7	700
8	800
9	900
10	1000

After making the 10 classifications based on the sample size for each DT, the best result for each DT algorithm was used in developing rulesets for land cover classification. Rulesets were developed from output summaries and DT diagrams. The rulesets were then implemented in eCognition Developer 9.1 process tree for land cover classification to produce thematic maps for the five DTs.

2.3.6. Assessing Thematic Map Accuracy

The land cover classification is complete only after the accurate assessment is done on thematic maps [1,48,49]. The independent sample of 600 random validation points was used to build confusion matrices by comparing the classified and reference land cover points. The confusion matrices were used to calculate the users', producers', and overall accuracies (Figure 2). The suitability of the five land cover classification outputs was compared using user's and producer's accuracies. The general accuracy of the thematic maps was done using the overall accuracy.

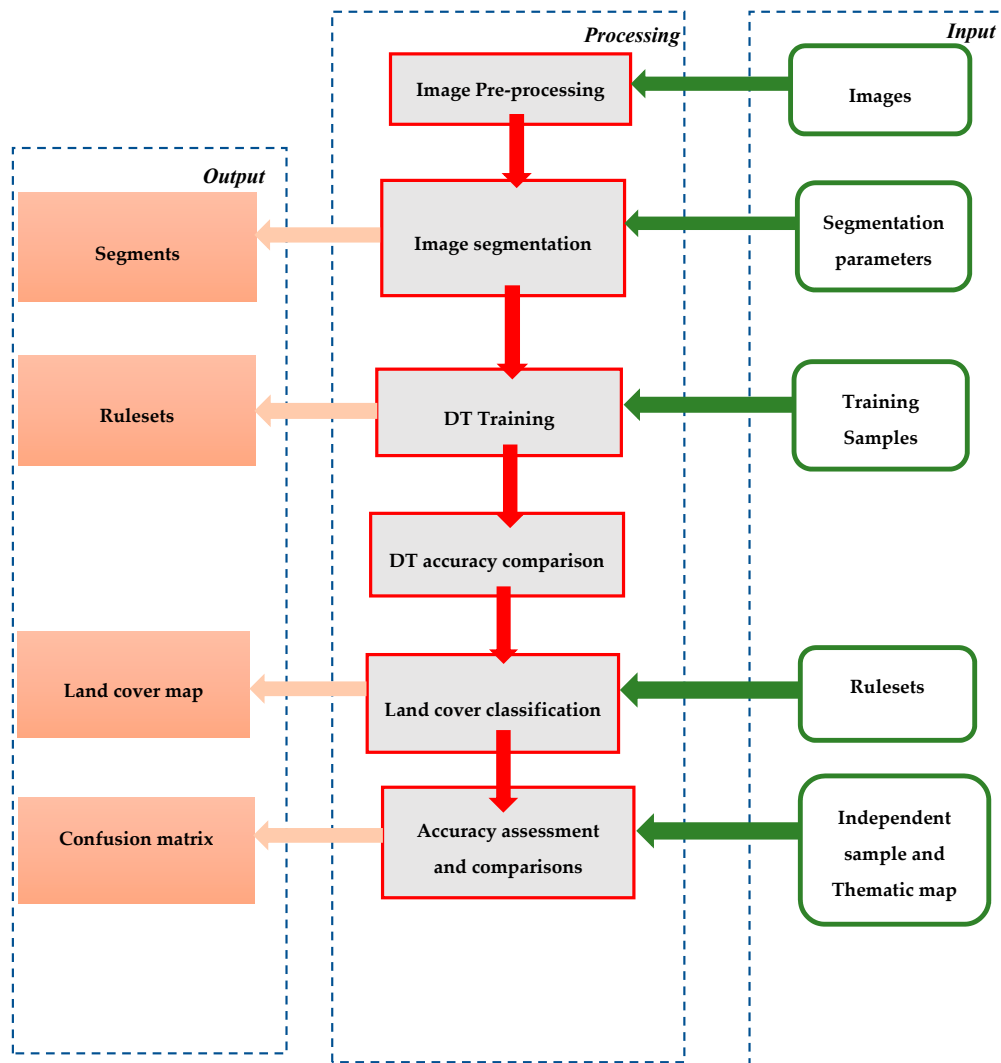


Figure 2. Flow chart of the steps used in the study.

3. Results

3.1. DT Accuracy

The DT algorithms were tested by using predicted results against an independent sample through cross-validation. The DT accuracies from this assessment showed that C5.0 had the highest (83%) mean DT accuracy, while Party had the least accuracy (77%) (Figure 3). However, these differences were not statistically significant when tested using a Kruskal–Wallis non-parametric test (p -value > 0.05) [50].

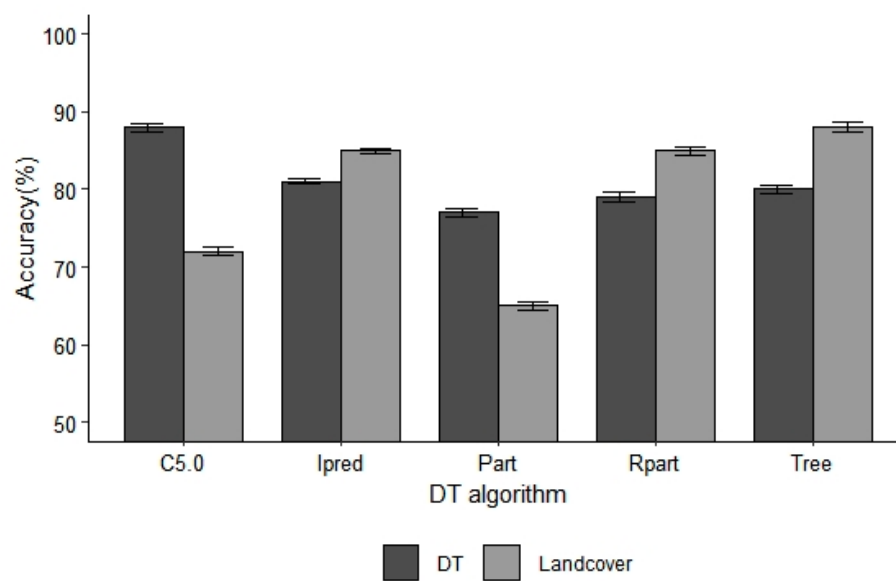


Figure 3. The mean DT accuracy of the five decision trees using from cross-validation.

The results of the performance of DT algorithms with different sample sizes show that the individual accuracy increases with the increase in sample size. Figure 4 shows that C5.0 had the highest DT accuracy of 88% when the largest sample size of 1000 was used, while Rpart had the lowest accuracy of 63% when the smallest sample size was used.

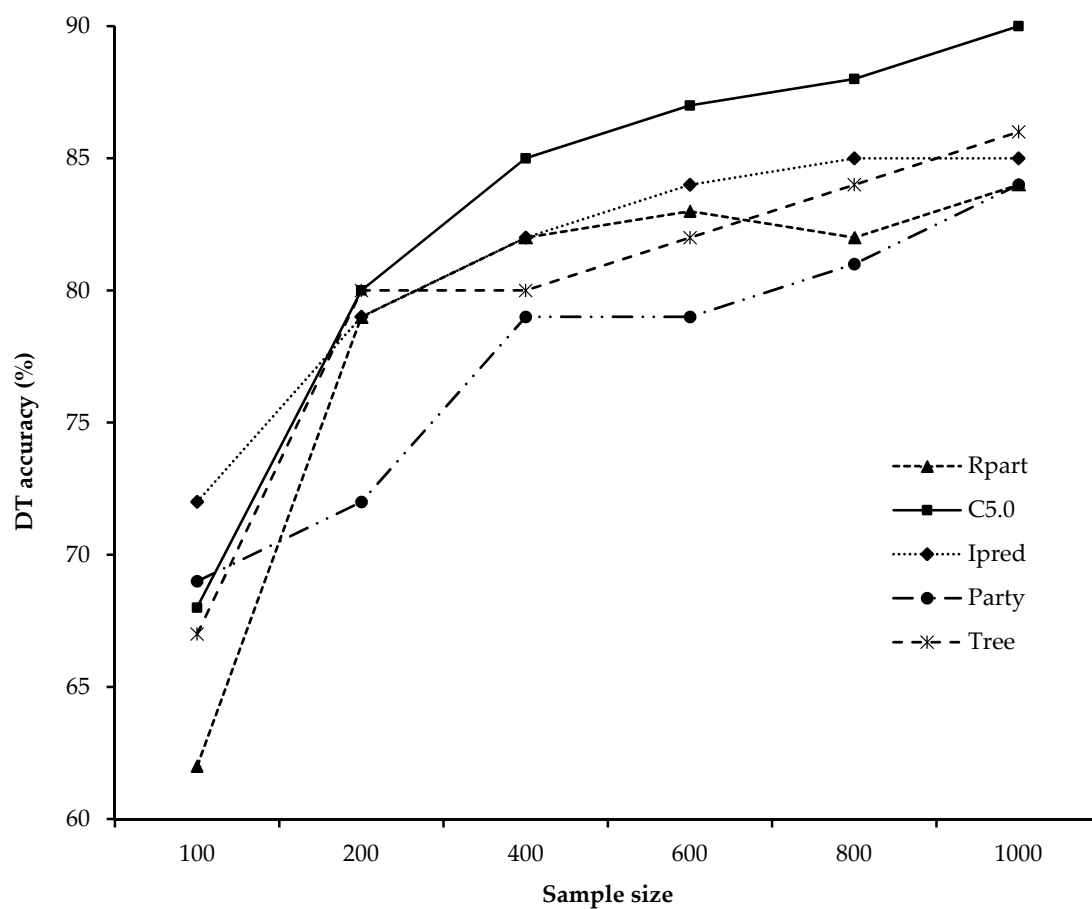


Figure 4. The accuracies of DT algorithms with increasing sample sizes.

3.2. Thematic Map Accuracy

3.2.1. Number of Variables and Classification Accuracy

The efficiency of DT algorithms was also assessed by comparing the variables used and the final thematic map accuracies. In this study, 197 initial variables were considered for training the DTs. Rpart had the fewest variables (8), while C5.0 used 35 variables for building a single DT (Figure 5). Tree and Ipred also had fewer variables (less than 15) and retained high land cover classification accuracies of over 85%. The ability of the DT algorithms to select a minimum number of variables and retain high classification accuracies was important for DT algorithms selection because this simplifies the structure of the DT.

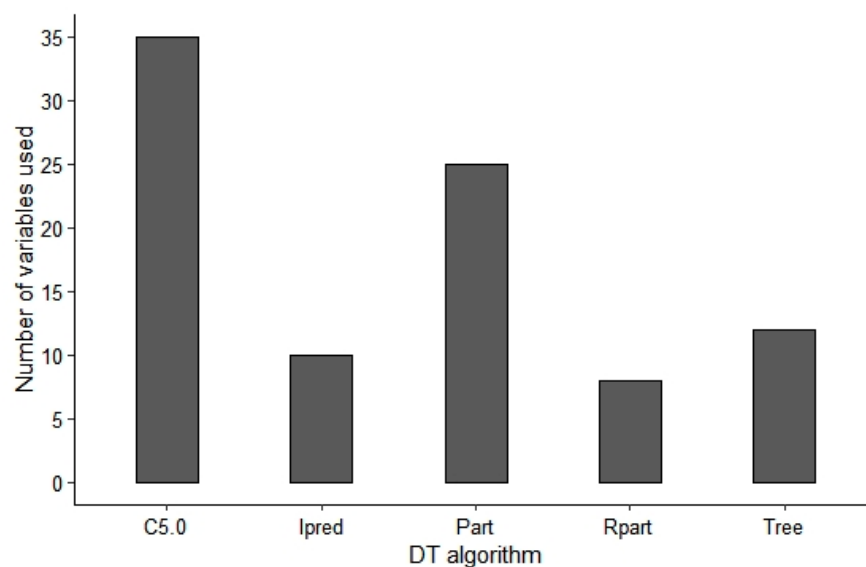


Figure 5. Thematic map classification accuracies (%) and the number of variables used during classification.

3.2.2. Thematic Map Classification Accuracy

The overall accuracy and Kappa coefficient, calculated from the confusion matrices, were used to compare OBIA classification accuracy resulting from the five DT algorithms. The producer's (PA) and user's accuracies (UA) were considered to establish the classification accuracy of each land cover class. The thematic maps showed a variation in the results which were produced. The C5.0 and Part algorithms had small land cover components, especially on bare land, while thematic maps for Rpart, Tree, and Ipred showed continuous sections for bare land (Figure 6).

The results show that the thematic map produced using the Tree algorithm had the highest overall accuracy of 89%, while the Party algorithm thematic map had the lowest overall accuracy of 73% (Table 5 and Figure 6). Rpart, Ipred, and C5.0 had overall accuracies of 88%, 85%, and 74%, respectively. The classification by DT algorithms with relatively low overall accuracies, C5.0 and Party, had very lower PAs and UAs (59–43%) for classes such as bare land and wetlands. Bare land had the lowest PA and UA, 74% and 62%, respectively, for the Rpart algorithm and primary forests had the lowest PA of 66% while the UA accuracy was 96%. Primary and secondary forests had some of the lowest user's and producer's accuracies (39% and 37%, respectively), indicating the challenges in separating the two land covers.

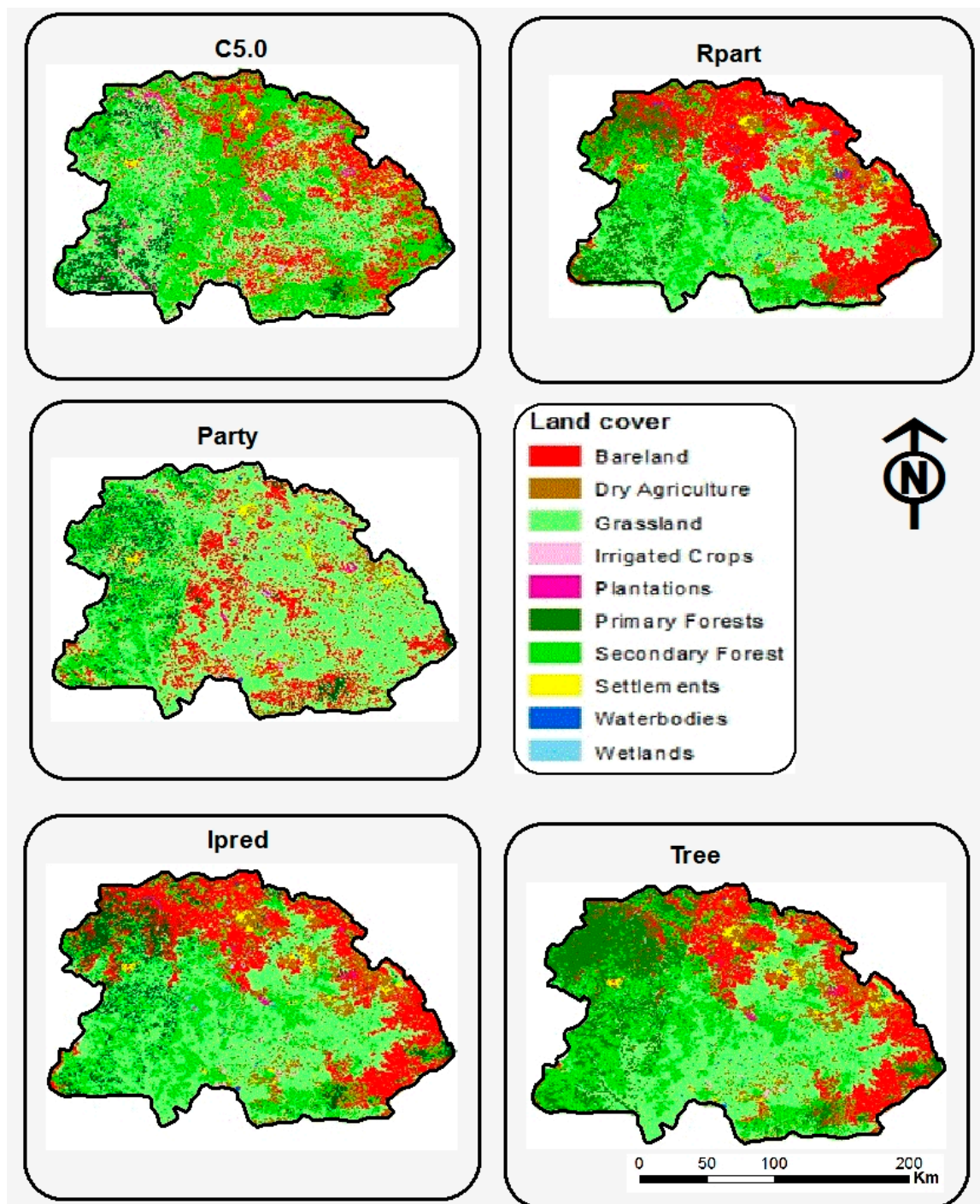


Figure 6. Maps of land cover classification using rulesets developed from the five DT algorithms.

Table 5. Summary of accuracies for land cover classification using Tree, Rpart, Ipred, C5.0, and Party.

Land Cover	Tree		Rpart		Ipred		C5.0		Party	
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
Bare land	93	83	74	62	74	81	43	47	59	58
Dry Agriculture	97	99	97	97	94	95	93	70	91	72
Grassland	92	86	98	98	93	93	78	82	65	95
Irrigated Crops	100	81	100	100	97	100	100	95	100	100
Plantation Forest	100	93	69	94	91	100	82	90	81	54
Primary Forests	66	96	89	89	88	82	80	39	90	80
Secondary Forests	85	82	85	87	87	82	64	84	82	37
Settlements	100	98	98	98	84	88	100	87	92	83
Waterbodies	100	98	91	87	100	98	100	98	100	98
Wetlands	100	88	100	100	68	94	88	88	44	88
Overall accuracy (%)	89		88		85		74		73	
Kappa coefficient (%)	86		84		82		70		70	

3.2.3. Other DT Characteristics

The structure of each algorithm was also considered in terms of graphic output, useful summaries, and the ability to produce ruleset as part of the output. Rpart, Tree, C5.0, and Party have tree diagrams as part of their output. Unlike Rpart and Tree algorithms, C5.0 and Party produced large tree diagrams which are difficult to interpret. This is because of their ability to include many variables when building DTs. On the other hand, Ipred aims at improving the classification results by developing many DT and improving their accuracies. It is difficult to produce a single tree diagram because the Ipred algorithm focuses on many DTs at once [51]. All DT algorithms produce summaries which are useful for developing rulesets; however, C5.0 and Rpart produce comprehensive rulesets for each terminal node.

4. Discussion

4.1. DT Accuracy

The comparison of the five DT algorithms showed that the DT accuracy was not significantly different among these algorithms. However, C5.0 showed high mean DT accuracy across different sample sizes. These findings are in line with the findings of Powers et al. [13], who achieved over 88% DT accuracy using the C5.0 algorithm. This can be attributed to C5.0's ability to integrate more decision variables in developing a DT through boosting and bagging [52,53]. Boosting improves classification by continuously formulating independent trees which are used to correct errors on the final models, while, in bagging, several trees are formulated and the final tree is established by voting for the most accurate variables and splits [46,47,53]. Apart from C5.0, DT algorithms such as Ipred and random forest also use bagging to improve classification accuracy [51,54,55].

In evaluating the performance of DT algorithms, it is important to employ a multiple criteria approach. DeFries et al. [47] used a multiple criteria approach in evaluating DT performances by considering their accuracy, ability to handle noise in the data, computation time and structure of the algorithms. While algorithms the DT accuracy of the five algorithms were not significantly different, the C5.0 had a relatively higher mean DT accuracy than the other four DTs. However, the C5.0 algorithm is very sensitive to noise in the data and they have a larger structure [55]. In the present study, 35 variables were used and 105 nodes were developed for a single C5.0 DT as compared to other algorithms such as Rpart which had 10 variables and 12 nodes.

The DTs algorithm, such as Rpart and Tree, which have simple structures, are effective in selecting and reducing the number of predictor variables [46]. Rodriguez-Galiano et al. [50] reported that the accuracy among different DT algorithms did not differ on land cover mapping in the Mediterranean region; however, the performance of these DTs was sensitive to noise on the data and the size of the sample. Therefore, the performance of the algorithm should not be selected based on the statistical

accuracy alone but rather on multiple criteria which include several considerations and the eventual classification accuracy of the thematic maps.

4.2. Thematic Map Classification Accuracy

This study has shown that DT algorithms are effective tools in developing decision rulesets for land cover thematic maps. The DTs have been used in land cover classification because they are non-parametric in nature and can be used with a number of auxiliary data such as digital elevation models (DEM), spectral indices, and spatial data [56,57]. For example, Im et al. [58] used Light Detection and Ranging (LiDAR) data and OBIA in land cover classification using the C5.0 DT algorithm.

High thematic map classification accuracy can be achieved by using these DTs; however, the accuracy differs from algorithm to algorithm. Due to spectral similarities between some classes such as primary and secondary forests, their accuracies were not as high as the other land cover classes (e.g., water and plantation forests). Phiri [19] reported similar challenges when conducting land cover classification in Zambia. Sharma et al. [56] reported that land cover classification accuracy using CART algorithms was better (>88%) than traditional classification approaches such as Maximum Likelihood and ISODATA, which attained overall accuracies of less than 72%. However, most studies have reported that other machine-learning algorithms such as support vector machines, random forests and neural network produced higher classification accuracies than the DTs used in this study [52,56,57]. Apart from C5.0 algorithm, CART based algorithms, such as Rpart and Tree, are the most commonly used algorithms because of their simplicity and ability to select and reduce variables for classification [16,57].

The high DT accuracy from cross-validation may not result in high land cover classification accuracy because of: (1) the overfitting and saturation of the algorithm; (2) the ability to handle noise on the data; and (3) the size and structure of the algorithm [47]. For example, C5.0 had a high mean DT accuracy similar to other algorithms; however, it had a relatively low land cover classification accuracy (Table 2). This is large because the C5.0 algorithm is susceptible to noise in the data as this algorithm does not have a strong ability to handle outliers and is more prone to overfitting [47,55]. Large DTs need to be pruned in order to reduce the effects of overfitting; however, pruning may affect classification accuracy [13].

In this study, land cover classification accuracy was high for DT algorithms such as Rpart and Tree; these DTs also had simple structures, an ability to deal with noise in the data and high statistical prediction accuracy. Simple algorithms which use a minimum number of decision variables have simple structures, are less saturated, and hence can be easily interpreted [56,57]. Among these DT algorithms, Rpart and C5.0 are commonly used in land cover classification and usually produce high classification accuracies. Powers et al. [13] reported 88% overall accuracy when C5.0 was used for mapping fine scale-industrial disturbance. Another DT algorithm which is simple and has high classification accuracy is the Tree algorithm; however, Rpart is preferable to Tree because it is more flexible and has a lot of supportive packages currently available in R statistical software [55]. When working with DT algorithms for land cover classification, it is important to establish the effects of different tuning parameters such as the number of variables, number of splits, size of a tree and allowable error because they can influence the classification results [6,50]. Future studies could focus more on the influence of different tuning parameters on the classification accuracy of different landscapes and different remote sensing data.

4.3. Selecting the Best DT for Ruleset Development

Choosing the ideal DT algorithm to use for land cover classification should be the most important objective when using these DT algorithms to establish decision rulesets in OBIA land cover classification. It is important to consider all the properties of the DT algorithms such as model accuracy, simplicity, ability to handle different numbers of variables, and sizes of datasets (Table 6). This can be achieved by using a multiple criteria evaluation approach as suggested by DeFries et al. [47]. The focus in

this evaluation should be on the DT algorithms which have high classification accuracies and simple structures and are not susceptible to noise in the data and are easy to interpret.

Table 6. Summary of classification accuracies and other properties of DT algorithms.

DT Algorithm	DT Accuracy (%)	Land Cover Classification Accuracy (%)	Number of Rulesets	Simplicity	Graphic Output	Ruleset Output	Variable Selection
Rpart	79	88	10	✓	✓	✓	✓
Party	77	74	30		✓		
Tree	80	89	12	✓	✓		✓
C5.0	83	77	36		✓	✓	
Ipred	81	86	12				✓

In this study, Tree and Rpart algorithms are recommended for developing decision rulesets, especially when larger numbers of variables are involved as these algorithms have the ability to select a small number of influential variables for classification and hence achieving simplicity and high classification accuracy [47,56]. Ipred algorithm does not differ from Rpart and Tree in most of its functionalities; however, this algorithm was built on the principle of bagging which is difficult to achieve simplicity and extract decision rulesets as it produces several decision trees. To successfully use Ipred, a function (nbagg = 1) which specifies the production of one DT could be employed [51]. The C5.0 and Party should be used when the objective is to include more predictor variables and to produce a high DT accuracy during cross-validation.

Although the overall accuracies achieved in this study were high, user's and producer's accuracies for spectrally similar classes such as primary and secondary forest were low. Therefore, there is a need to define the classes so that they are spectrally similar. In addition, other methods such as non-parametric machine learning classifiers, e.g., Random Forest [15,19,59] and Support Vector Machine [60,61], which have proved to be more effective, can be used during classification in order to achieve higher accuracy.

The current study has only one location and this has the potential to affect the transferability and the generalization of the results. This can be a challenge for future studies and the results from this study can be generalization with some levels of uncertainties. It is important to note that the generalization, transferability, and reproducibility of the results are largely influenced by the type of DT used, the sample size, and the type of input features.

5. Conclusions

In this paper, a systematic comparison of the performance of five DT algorithms on land cover classification using Landsat 8 is presented. The main focus was selecting DT algorithms which have high classification accuracies, simple structures, and are easy to interpret, by using the multiple criteria approach suggested by DeFries et al. [47]. While all algorithms had high mean DT accuracies, it was established that the Tree and the Rpart algorithms were simple, easy to interpret, and not affected by noise from datasets. The results from the Tree and Rpart DT algorithms produce high overall accuracies of over 86%. The C5.0 and Party algorithms were equally good with respect to the overall accuracy; however, they incorporate a large number of decision variables in the output which can be difficult to implement and exhibit the effects of overfitting and saturation. Further analysis showed that Rpart and Tree can select the minimum number of variables and hence retain simple but accurate rulesets. Based on the DT and the land cover accuracy and other important aspects such as number of variable and the simplicity of the DT structure, it can be recommended to use Rpart or Tree in developing rulesets for OBIA land cover classification of Landsat 8 imagery. Going forward, future studies can

compare the performance of these simple DTs with the contemporary machine learning classifiers such as RF and SVM in different geographic locations at multiple time periods.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2220-9964/9/5/329/s1>.

Author Contributions: Conceptualization, Darius Phiri and Matamyo Simwanda; Funding acquisition, Yuji Murayama and Manjula Ranagalage; Writing—original draft, Darius Phiri; and Writing—review and editing, Matamyo Simwanda, Vincent Nyirenda, Yuji Murayama, and Manjula Ranagalage. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partly supported by the JSPS grant 18H00763 (2018–20).

Acknowledgments: We would like to acknowledge the anonymous reviewers for their valuable suggestions that have helped to improve this manuscript as they worked tirelessly during the difficult time of COVID-19.

Conflicts of Interest: As authors, we declare that there is no conflict of interest.

References

1. Kumar, R.; Nandy, S.; Agarwal, R.; Kushwaha, S.P.S. Forest cover dynamics analysis and prediction modeling using logistic regression model. *Ecol. Indic.* **2014**, *45*, 444–455. [\[CrossRef\]](#)
2. Li, M.; Ma, L.; Blaschke, T.; Cheng, L.; Tiede, D. A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 87–98. [\[CrossRef\]](#)
3. Peña-Barragán, J.M.; Ngugi, M.K.; Plant, R.E.; Six, J. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sens. Environ.* **2011**, *115*, 1301–1316. [\[CrossRef\]](#)
4. Kindu, M.; Schneider, T.; Teketay, D.; Knoke, T. Land use/land cover change analysis using object-based classification approach in Munessa-Shashemene landscape of the Ethiopian highlands. *Remote Sens.* **2013**, *5*, 2411–2435. [\[CrossRef\]](#)
5. Drăguț, L.; Tiede, D.; Levick, S.R. ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 859–871. [\[CrossRef\]](#)
6. Phiri, D.; Morgenroth, J.; Xu, C. Four decades of land cover and forest connectivity study in Zambia—An object-based image analysis approach. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *79*, 97–109. [\[CrossRef\]](#)
7. Kelly, M.; Blanchard, S.D.; Kersten, E.; Koy, K. Terrestrial remotely sensed imagery in support of public health: New avenues of research using object-based image analysis. *Remote Sens.* **2011**, *3*, 2321–2345. [\[CrossRef\]](#)
8. Phiri, D.; Morgenroth, J.; Xu, C.; Hermosilla, T. Effects of pre-processing methods on Landsat OLI-8 land cover classification using OBIA and random forests classifier. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 170–178. [\[CrossRef\]](#)
9. Li, M.; Zang, S.; Wu, C.; Deng, Y. Segmentation-based and rule-based spectral mixture analysis for estimating urban imperviousness. *Adv. Space Res.* **2015**, *55*, 1307–1315. [\[CrossRef\]](#)
10. Myint, S.W.; Gober, P.; Brazel, A.; Grossman-Clarke, S.; Weng, Q. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sens. Environ.* **2011**, *115*, 1145–1161. [\[CrossRef\]](#)
11. Phiri, D.; Simwanda, M.; Nyirenda, V. Mapping the Impacts of Cyclone Idai in Mozambique Using Sentinel-2 and OBIA Approach. *S. Afr. J. Geogr.* **2020**. [\[CrossRef\]](#)
12. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [\[CrossRef\]](#)
13. Powers, R.P.; Hermosilla, T.; Coops, N.C.; Chen, G. Remote sensing and object-based techniques for mapping fine-scale industrial disturbances. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 51–57. [\[CrossRef\]](#)
14. Phiri, D.; Morgenroth, J.; Xu, C. Long-term land cover change in Zambia: An assessment of driving factors. *Sci. Total Environ.* **2019**, 134206. [\[CrossRef\]](#)
15. Puissant, A.; Rougier, S.; Stumpf, A. Object-oriented mapping of urban trees using Random Forest classifiers. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 235–245. [\[CrossRef\]](#)
16. Freund, Y.; Mason, L. The alternating decision tree learning algorithm. In Proceedings of the ICML, Sixteenth International Conference on Machine Learning, Bled, Slovenia, 27–30 June 1999; pp. 124–133.
17. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [\[CrossRef\]](#)

18. Phiri, D.; Phiri, E.; Kasubika, R.; Zulu, D.; Lwali, C. The implication of using a fixed form factor in areas under different rainfall and soil conditions for *Pinus kesiya* in Zambia. *South. For. J. For. Sci.* **2016**, *78*, 35–39. [\[CrossRef\]](#)
19. Phiri, D. Monitoring Land Cover Dynamics for Zambia Using Remote Sensing: 1972–2016. Ph.D. Thesis, University of Canterbury, Christchurch, New Zealand, 2019.
20. Kalaba, F.K.; Quinn, C.H.; Dougill, A.J.; Vinya, R. Floristic composition, species diversity and carbon storage in charcoal and agriculture fallows and management implications in Miombo woodlands of Zambia. *For. Ecol. Manag.* **2013**, *304*, 99–109. [\[CrossRef\]](#)
21. Phiri, D.; Morgenroth, J. Developments in Landsat land cover classification methods: A review. *Remote Sens.* **2017**, *9*, 967. [\[CrossRef\]](#)
22. Wulder, M.A.; White, J.C.; Loveland, T.R.; Woodcock, C.E.; Belward, A.S.; Cohen, W.B.; Fosnight, E.A.; Shaw, J.; Masek, J.G.; Roy, D.P. The global Landsat archive: Status, consolidation, and direction. *Remote Sens. Environ.* **2016**, *185*, 271–283. [\[CrossRef\]](#)
23. Poursanidis, D.; Chrysoulakis, N.; Mitraka, Z. Landsat 8 vs. Landsat 5: A comparison based on urban and peri-urban land cover mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *35 Pt B*, 259–269. [\[CrossRef\]](#)
24. ESRI. *ArcGIS Desktop. Release 10.4*; Environment System Research Institute: Redlands, CA, USA, 2016.
25. Hussain, M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 91–106. [\[CrossRef\]](#)
26. Rasuly, A.; Naghdifar, R.; Rasoli, M. Monitoring of Caspian Sea Coastline Changes Using Object-Oriented Techniques. *Procedia Environ. Sci.* **2010**, *2*, 416–426. [\[CrossRef\]](#)
27. Jacquin, A.; Misakova, L.; Gay, M. A hybrid object-based classification approach for mapping urban sprawl in periurban environment. *Landsc. Urban Plan.* **2008**, *84*, 152–165. [\[CrossRef\]](#)
28. Liao, L.M.; Song, J.L.; Wang, J.D.; Xiao, Z.Q.; Wang, J. Bayesian Method for Building Frequent Landsat-Like NDVI Datasets by Integrating MODIS and Landsat NDVI. *Remote Sens.* **2016**, *8*, 452. [\[CrossRef\]](#)
29. Zhu, X.L.; Liu, D.S. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 222–231. [\[CrossRef\]](#)
30. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [\[CrossRef\]](#)
31. Gitelson, A.A.; Merzlyak, M.N. Remote sensing of chlorophyll concentration in higher plant leaves. *Adv. Space Res.* **1998**, *22*, 689–692. [\[CrossRef\]](#)
32. Sripada, R.P.; Heiniger, R.W.; White, J.G.; Meijer, A.D. Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agron. J.* **2006**, *98*, 968–977. [\[CrossRef\]](#)
33. Atzberger, C.; Darvishzadeh, R.; Immitzer, M.; Schlerf, M.; Skidmore, A.; le Maire, G. Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *43*, 19–31. [\[CrossRef\]](#)
34. Birth, G.S.; McVey, G.R. Measuring the color of growing turf with a reflectance spectrophotometer. *Agron. J.* **1968**, *60*, 640–643. [\[CrossRef\]](#)
35. Goel, N.S.; Qin, W. Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: A computer simulation. *Remote Sens. Rev.* **1994**, *10*, 309–347. [\[CrossRef\]](#)
36. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [\[CrossRef\]](#)
37. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [\[CrossRef\]](#)
38. Roujean, J.-L.; Breon, F.-M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* **1995**, *51*, 375–384. [\[CrossRef\]](#)
39. Key, C.; Benson, N. Landscape assessment: Remote sensing of severity, the normalized burn ratio and ground measure of severity, the composite burn index. In *FIREMON: Fire Effects Monitoring and Inventory System* Ogden, Utah: USDA Forest Service, Rocky Mountain Res. Station; USDA Forest Service, Rocky Mountain Research Station: Ogden, UT, USA, 2005.
40. Garcia, M.L.; Caselles, V. Mapping burns and natural reforestation using Thematic Mapper data. *Geocarto Int.* **1991**, *6*, 31–37. [\[CrossRef\]](#)

41. Segal, D. Theoretical basis for differentiation of ferric-iron bearing minerals using Landsat MSS data. In Proceedings of the International Symposium on Remote Sensing of Environment, 2nd Thematic Conference, Remote Sensing for Exploration Geology 1982, Ft. Worth, TX, USA, 6–10 December 1982; Volume II, pp. 949–951.
42. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [\[CrossRef\]](#)
43. Salomonson, V.; Appel, I. Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote Sens. Environ.* **2004**, *89*, 351–360. [\[CrossRef\]](#)
44. Silleos, N.G.; Alexandridis, T.K.; Gitas, I.Z.; Perakis, K. Vegetation Indices: Advances Made in Biomass Estimation and Vegetation Monitoring in the Last 30 Years. *Geocarto Int.* **2006**, *21*, 21–28. [\[CrossRef\]](#)
45. Gao, B.-C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [\[CrossRef\]](#)
46. Punia, M.; Joshi, P.; Porwal, M. Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data. *Expert Syst. Appl.* **2011**, *38*, 5577–5583. [\[CrossRef\]](#)
47. DeFries, R.S.; Chan, J.C.-W. Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification from Satellite Data. *Remote Sens. Environ.* **2000**, *74*, 503–515. [\[CrossRef\]](#)
48. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC Press/Taylor & Francis: Boca Raton, FL, USA, 2009; Volume 2.
49. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [\[CrossRef\]](#)
50. Rodriguez-Galiano, V.F.; Chica-Rivas, M. Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models. *Int. J. Digit. Earth* **2014**, *7*, 492–509. [\[CrossRef\]](#)
51. Peters, A.; Hothorn, T.; Ipred: Improved Predictors. R Package Version 0.9-6. 2017. Available online: <https://CRAN.R-project.org/package=ipred> (accessed on 6 June 2017).
52. Chan, J.C.-W.; Huang, C.; DeFries, R. Enhanced algorithm performance for land cover classification from remotely sensed data using bagging and boosting. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 693–695.
53. Kuhn, M.; Steve, W.; Coulter, N. C50: C5.0 Decision Trees and Rule-Based Models. R Package Version 0.1.0-24. 2015. Available online: <https://CRAN.R-project.org/package=C50> (accessed on 6 June 2017).
54. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* **2012**, *118*, 259–272. [\[CrossRef\]](#)
55. Lantz, B. *Machine Learning with R*, 1st ed.; Packt Publishing: Birmingham, UK, 2013.
56. Sharma, R.; Ghosh, A.; Joshi, P. Decision tree approach for classification of remotely sensed satellite data using open source support. *J. Earth Syst. Sci.* **2013**, *122*, 1237–1247. [\[CrossRef\]](#)
57. Shao, Y.; Lunetta, R.S. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 78–87. [\[CrossRef\]](#)
58. Im, J.; Jensen, J.R.; Hodgson, M.E. Object-based land cover classification using high-posting-density LiDAR data. *GIScience Remote Sens.* **2008**, *45*, 209–228. [\[CrossRef\]](#)
59. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104. [\[CrossRef\]](#)
60. Kranjčić, N.; Medak, D.; Župan, R.; Rezo, M.J.R.S. Support Vector Machine Accuracy Assessment for Extracting Green Urban Areas in Towns. *Remote Sens.* **2019**, *11*, 655. [\[CrossRef\]](#)
61. Huang, C.; Davis, L.S.; Townshend, J.R.G. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749. [\[CrossRef\]](#)

