

Land Cover Land Use (LCLU) Classification Methods in Semi-Arid Botswana

Rejoice Tsheko*

Department of Agricultural Engineering and Land Planning, Botswana University of Agriculture and Natural Resources, Botswana

ABSTRACT

This paper presents Land Cover Land Use (LCLU) detected from a Landsat 8 (OLI) using two classification schemes namely Maximum Likelihood Algorithm (MLA) and Artificial Neural Networks (ANNs). Analysis was carried out using two, three and eight features (surface reflectance and indices). For all classifications, the overall accuracy and kappa statistic varied from 93.81% and 0.89 to 99.38% and 0.99, respectively. The highest classification accuracies were obtained by either using all eight features or two features (indices only) for both classification schemes. This demonstrates the importance of Normalized Difference Vegetation Index (NDVI) and Normalized Difference Buildup Index (NDBI) in LCLU mapping. The two indices are robust enough to be used to detect shrubs, trees, water, and buildup in a satellite image. Further, the ANNs classifier is also robust enough to be used for this classification. Although the MLA classifier used both the mean values and variance of the features, the ANNs classifier only used the mean values of the features. This is a demonstration of data fusion in a normalized scale -1.0 to 1.0. This work also demonstrates that acceptable classification accuracies can be achieved with fewer spectral channels.

Keywords: Land Cover Land Use (LCLU); Maximum Likelihood Algorithm (MLA); Artificial Neural Networks (ANNs); Normalized Difference Vegetation Index; Normalized Difference Buildup Index

INTRODUCTION

Production of land cover land use (LCLU) maps by classifying satellite image is now a standard practice in Earth Observation (EO). In addition to the production of the LCLU database, there is also a need for LCLU change detection to be quantified. LCLU mapping and monitoring land use changes is important for energy balances, carbon budget and hydrological cycle studies. Moreover, nowadays forecasting or predicting future LCLU is also a very important aspect of LCLU studies. Satellite images normally vary from low, medium to high spatial resolution and high, medium to low temporal resolution. These variations in spatial and temporal resolutions result in tradeoffs such as detecting small areas (spatial) and the possibility to have cloud and clouds shadows (temporal) in images especially for time series analysis. To overcome some of the limitations from these satellite image datasets, some researchers have explored using a combination (data fusion) of Landsat and MODIS time-series data with varying success.

In the detection of LCLU classes, either full pixel or subpixel schemes are used for the classification using statistical, regression, expert classifier, artificial neural networks, and other more sophisticated algorithms. LULC database is used for resources management, policy implementation, and natural disasters [1].

Desertification is an issue of environmental concern in semi-arid to arid parts of the world. Desertification is defined as the depletion or destruction of the biological capability of the land (in arid, semi-arid and dry sub-humid areas) leading to desert-like condition caused by climate variation and human activities. Desertification is a serious problem in arid and semi-arid environments which cover 40% of the global land surface and account for about 15% of the human populated. Some studies show that 76.1% of Kazakhstan land is desertification sensitive with 3.8% desertification (Hu). There have been different studies which estimated Africa to have a desertification of 43% (Reich). Land degradation in Botswana was estimated at 25% in 1980 (during drought) to 6.5% in 1994 (during a wet year)

*Corresponding to: lanning, Botswana University of Agriculture and Natural Resources, Botswana, Tel: 2673650349; E-mail: rtsheko@buan.ac.bw

Received date: June 30, 2021; Accepted date: October 09, 2021; Published date: October 19, 2021

Citation: Tsheko R (2021) Land Cover Land Use (LCLU) Classification Methods in Semi-Arid Botswana. J Remote Sens GIS. 10:p497

Copyright: © 2021 Tsheko R. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Landsat data has been used extensively to study land cover changes varying from deforestation, fire scars, agriculture monitoring, urban sprawl, and wetlands loss Jensen; Spruce, Zhu & Woodcock). Since and, Moderate Resolution Imaging Spectroradiometer (MODIS) and Sentinel-2 datasets have been used for monitoring vegetation and land cover changes respectively. Optical remote sensing has long been and is promising as a systematic and cost effective way of monitoring bare ground cover and gain. There is lack of accurate assessment of the status, change, and trend of desertification, this hinders developing global actions to prevent and eradicate the problem especially in one of the most seriously affected countries like China (Yang).

Studies have shown that accuracy assessment results from homogeneous land cover classes are generally higher than those from heterogeneous classes. In the arid to semi-arid areas such as the Kalahari, bare area is actually 30% herbaceous and 70% bare and this mix pixels are also present in i) rural areas where bare areas, buildup, crop fields and grazing pasture land classes are not distinct hence tend to generating a spectral mixture, ii) during the winter months, the agricultural fields and grazing lands are bare (no crops, leafless trees and shrubs, and no grass cover which is normally annual iii) urban buildup land use has a lot of open bare areas (unpaved). These mixed pixels tend to overestimate bare land class or confuse LCLU classes. Because of these mixed pixels which result in inconsistencies in estimating LCLU, we examine the use of classic Maximum Likelihood Algorithm (MLA) and an Artificial Neural Network (ANN) to classify Landsat 8 imagery and detect LCLU classes around Greater Gaborone area in South Eastern Botswana. Using a 30m pixel size is a challenge to accurately detect bare areas (found in all the LCLU) hence in this study only four LCLU classes are considered.

STUDY AREA

The study area is the greater Gaborone area in Botswana (Figure 1), a relatively flat landlocked country in Southern Africa of size 581, 730 km² with a population of about 2.024 million people with a population growth rate of 1.88% ((Central Statistics Office), (Statistics Botswana)). The country faces reoccurring droughts and a threat of desertification resulting from communal grazing which accounts for 71% of the country's land. The greater Gaborone area is the area south east of Botswana which comprise of the capital city Gaborone and several major towns and villages accounting for over 38% of the total population of Botswana ((Statistics Botswana). The annual climate ranges from months of dry temperate weather during winter (April-September) humid subtropical weather interspersed with drier periods of hot weather during summer (October-March). Annual rainfall, brought by winds from the Indian Ocean, averages 500 mm, for the study area [2].

The general vegetation of the country is savanna grassland with yellow or light brown grass cover (turning green after rains) and woody plants. The savanna ranges from acacia shrub savanna in the southwest through acacia thorn bush and tree savanna "parkland" into denser woodland and eventually forest as one moves north and east [3].



Figure 1: Study Area (24°17'53.84"S, 25°17'50.66"E to 25°17'51.14"S, 26°12'1.72"E)

METHODOLOGY

Landsat 8 (OLI) (2020) scenes (Path 172/Row 077) acquired 31st January 2020 was used in this study. A radiometric correction was done to convert the digital numbers (DNs) into scaled surface reflectance values and remove haze and other artifacts on the image. Preprocessing of the image included dark objects subtraction and calibration using parameters from the Landsat 8 metadata files. This process had been used to reduce over-estimating bare soil and build-up due to pixel saturation. All the atmospheric corrections were done using ENVI software Radiometric Correction Module © Exelis Visual Information Solutions [4].

Features Extraction

Four (4) LCLU classes were defined for the study area, these include tree (trees dominated including riverine), buildup (settlements, rural and urban centers), shrubs (include fallow agricultural land), and water (manmade reservoirs). The features (Table 1) are mean reflectance values of following six (6) bands and two (2) indices. These features were calculated and extracted for each of the four LCLU classes above [5].

Table 1: Features/bands used.

	Band or feature	Wavelength (nm)
1	Band 2 (blue)	452-512
2	Band 3 (green)	533-590

3	Band 4 (red)	636-673
4	Band 5 (nir)	851-879
5	Band 6 (swir1)	1566-1651
6	Band 7 (swir2)	2107-2294
7	Normalized difference vegetation index (ndvi)	
8	Normalized difference bare index (ndbi)	

Classification

Two (2) supervised image classification schemes were used namely Artificial Neural Networks (ANNs) and Maximum Likelihood Algorithm (MLA). Four (4) tests were carried out namely 1) using all the eight (8) features in Table 1 referred to as ALL, 2) using only three (3) features (red, green & blue) bands referred to as RGB, 3) using three (3) features (nir, ndbi and ndvi) referred to as NNN and finally using only two (2) features (ndbi and ndvi) referred to as NN [6].

Artificial Neural Networks

Artificial neural networks (ANNs) computational tools that are extensive used in solving many complex real-world problems including classification in remote sensing. Some of the advantages of ANNs include nonlinearity, high parallelism, fault, and noise tolerance, and learning and generalization capabilities. In this study, the Java Neural Network Simulator (JavaNNS) which is a simulator for neural networks developed at the Wilhelm-Schickard-Institute for Computer Science (WSI) was used. It is based on the Stuttgart Neural Network Simulator (SNNS) 4.2 kernel, with a graphical user interface written in Java set on top of it. The implementation included random weights as the initializing function, topological order as the updating function, backpropagation as the learning function and a normalized remapping function [7].

When using all the eight (8) features (ALL), neural network architecture with eight (8) inputs, two (2) hidden layers with eight (8) nodes and four (4) outputs was used. For the two (RGB & NNN) three (3) features each, two neural networks with three (3) inputs, two (2) hidden layers with six (6) nodes and four (4) outputs were used as shown in Error! Reference source not found.. For the two features, the architecture was similar to that of three inputs except that there were two inputs instead of three. The interfacing between the datasets and JavaNNS was done using the "Interface to SNNS" software, which includes the routines "input2snns" and "output2snns" distributed by Tor Vergata University, Dipartimento di Informatica Sistemi e Produzione, Via del Politecnico, 1 - 00133 Rome, ITALY [8].

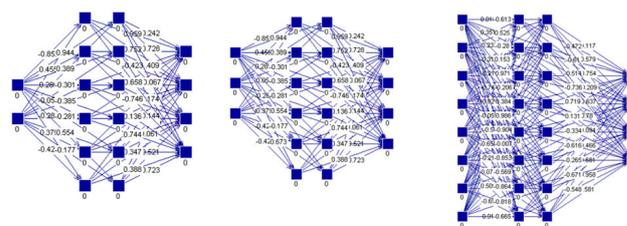


Figure 2: Neural Network Architectures

Training sites were identified in the satellite image and the mean feature values were calculated for each training site and used in the classification. For each LULC class, thousand (1000) pixels were used to train the artificial neural network as shown in Table 2 below. The training sites (or regions of interest ROIs) were selected using the ENVI software. These datasets were split into two (training and validation) sets and each set was processed using the input2snns routine. The output files from the input2snns were the training and validation patterns sets and their associated statistical files which were used as inputs to the JavaNN Simulator and output2snns routine. The JavaNN Simulator produced the classifier (*.net file), this classifier was then used in the output2snns routines to classify the whole image (using the statistics file and the image data). The error graph (error of training set and error of validation set) was used to monitor the network training and also avoiding overtraining the classifier [9].

Table 2: Training and validation pixels

LCLU Class	Training pixels	Validation pixels
Tree dominated	1000	610
Shrubs	1000	673
Buildup	1000	363
Water	1000	241

Maximum likelihood classification

Maximum Likelihood Estimation is a probabilistic framework for solving the problem of density estimation. It involves maximizing a likelihood function to find the probability distribution and parameters that best explain the observed data. A pixel is assigned to the LCLU class with the highest likelihood or labelled as unclassified if the probability value is below a set threshold. The images 8 features (ALL), 3 features (RGB), 3 features (NNN) and 2 features (NN) were classified using the supervised maximum likelihood algorithm on the same training areas (ROIs). The MLA implementation was done using ENVI software Supervised Classification Module © Exelis Visual Information Solutions [10].

Accuracy assessment

For validation, a total of 485 random pixels were selected from the Landsat 8 image with the aid from the following: Landsat 8 panchromatic band, google map and field visits. The field visit was important as LULC is highly dynamic (google map versus

Landsat 8 acquisition date) especially in urban areas. Random samples were created in ENVI software and then processed in ArcGIS where the points (class name, ID, X and Y coordinates) were converted to KMZ. The KMZ file was then loaded in google map[11].

Error Matrices

Once the classification of the data sets was achieved, the following matrices (overall accuracy OA, Kappa Coefficient KC, Producer Accuracy PA and User Accuracy UA) were calculated for all the tests [12].

RESULTS AND DISCUSSION

The mean feature values for all the four (4) LULC classes are shown in Table 3 below and notably the spectral signature and ratios of shrubs are like those of trees and the ndvi and ndbi values of the water are negative [13].

Table 3: Mean Features values

	Shrubs	Water	Buildup	Trees
Feature	Mean	Mean	Mean	Mean
blue	0.01106	0.02162	0.03257	0.00455
green	0.02116	0.03714	0.042	0.00916
red	0.0289	0.04042	0.0561	0.0074
nir	0.12803	0.01969	0.09668	0.11166
swir1	0.10243	0.00583	0.10733	0.04913
swir2	0.06599	0.00401	0.09509	0.02107
ndvi	0.6319	-0.2666	0.26325	0.87691
ndbi	0.56049	-0.6375	0.32125	0.74412

Figure 3 and Figure 4 below show results of classification (MLA and ANNs) using all the eight features (reflectance and reflectance ratios).

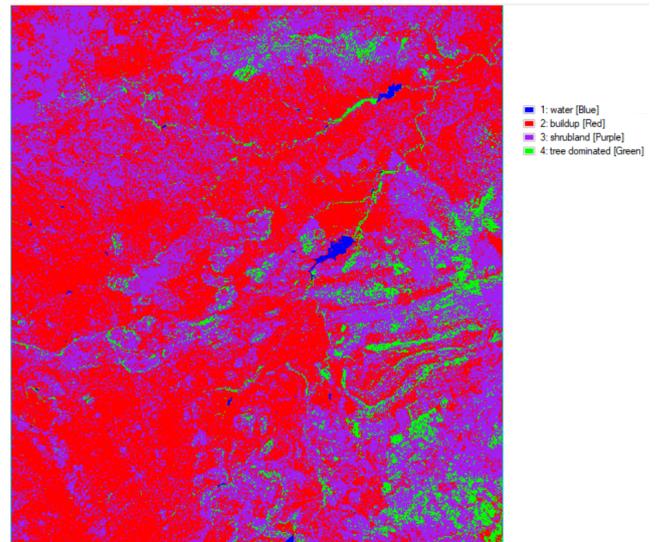


Figure 3: All Eight Features used in the classification (MLA)

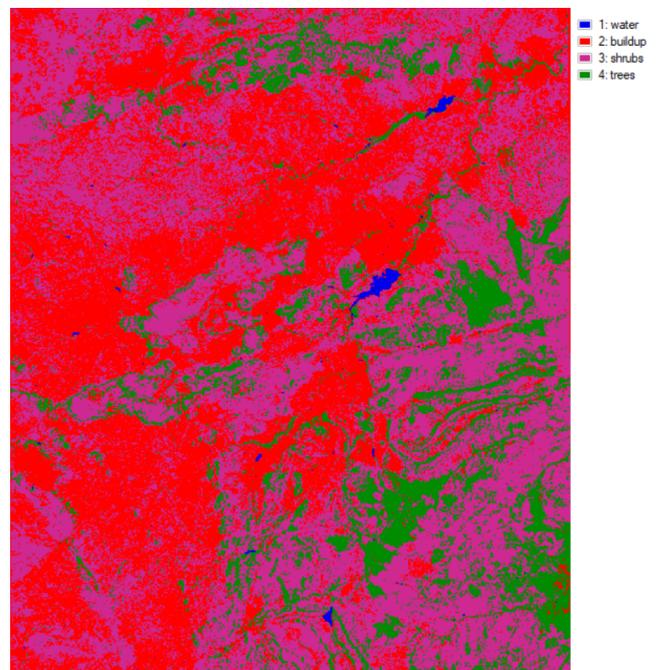


Figure 4: ALL Eight Features used in the classification (ANNs)

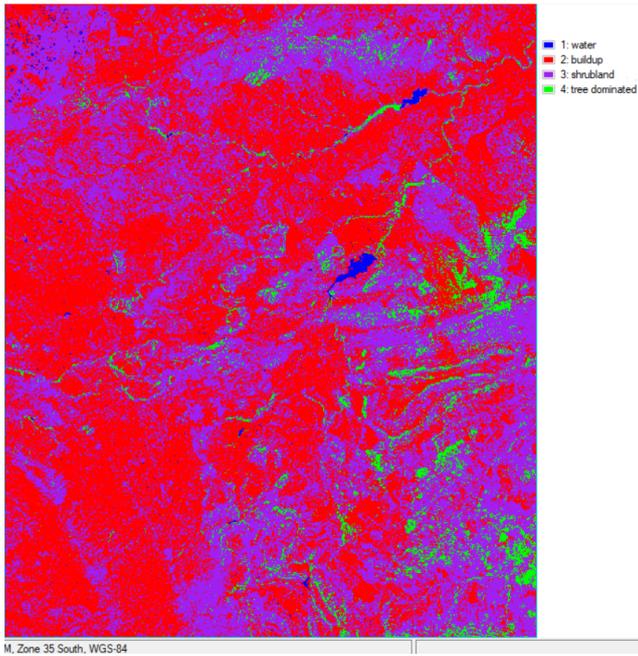


Figure 5: RGB spectral band Features used in the classification (MLA)

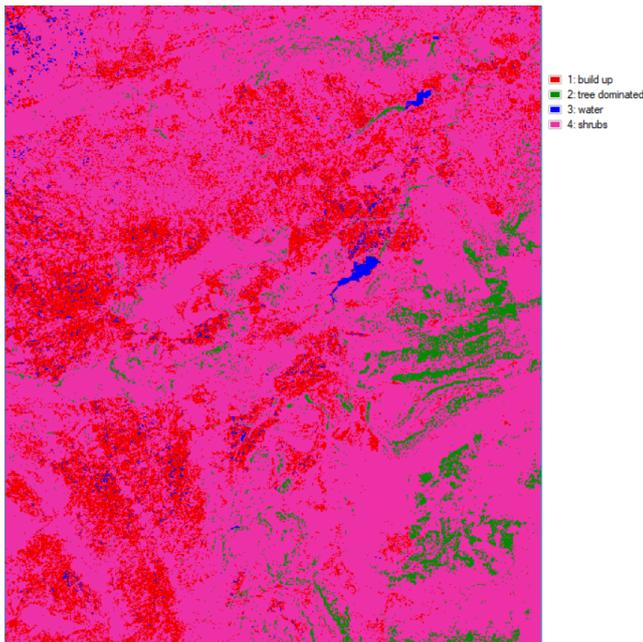


Figure 6: RGB spectral band Features used in the classification (ANNs)

Figure 7 and Figure 8 below show results of classification (MLA and ANNs) using all the nir, ndbi and ndvi features.

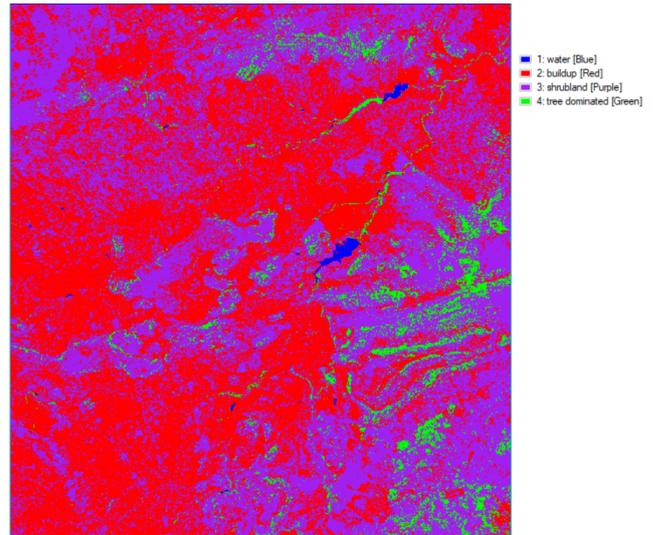


Figure 7: Near Infrared-NDVI-NDBI Features used in the classification (MLA)

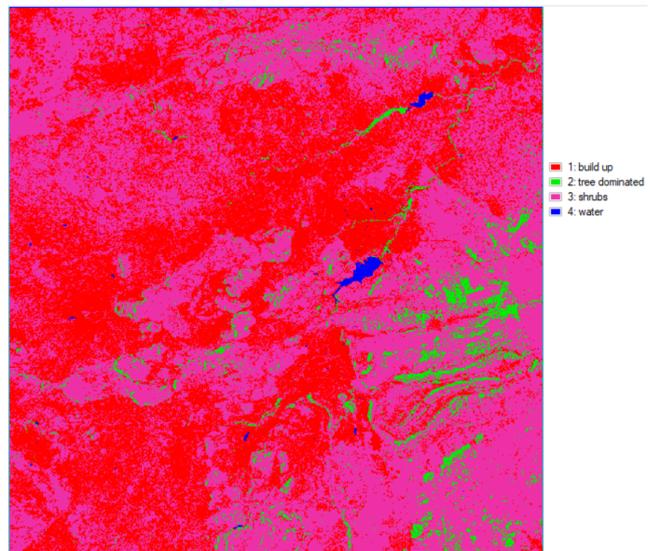


Figure 8: Near Infrared-NDVI-NDBI Features used in the classification (ANNs)

Figure 9 and Figure 10 below show results of classification (MLA and ANNs) using all the ndbi and ndvi features only.

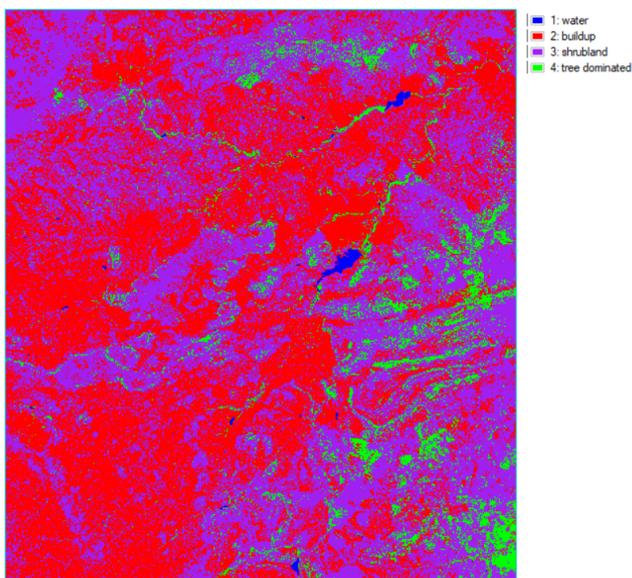


Figure 9: NDVI-NDBI Features used in the classification (MLA)

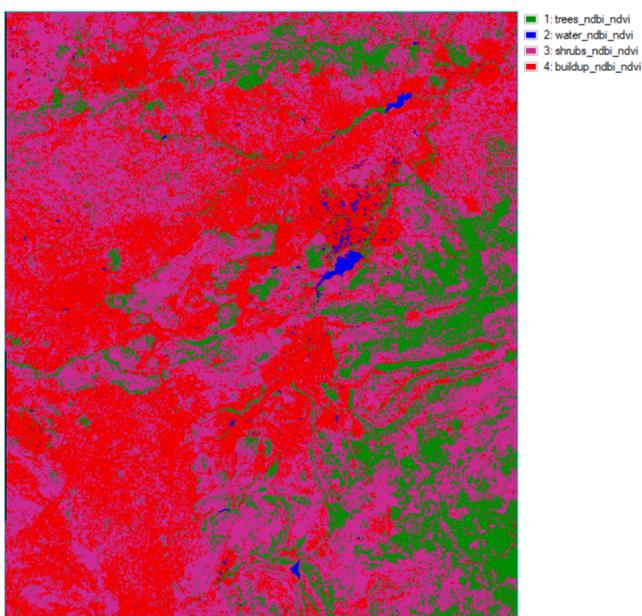


Figure 10: NDVI-NDBI Features used in the classification (ANNs)

The spectral separability of the LCLU is shown in Figure 11 below, generally the ndbi and the ndvi values are distinct for all the classes, the values for the trees is marginally higher than that of shrubs which means these two classes may be confused for each other. The standard deviations for these features are higher for the build-up class which indicates mixed pixels. As expected, the values of these features are negative for the water class but again the signal variability is large possibly due to water quality (algae, siltation, and wastewater) as these water bodies are manmade [14].

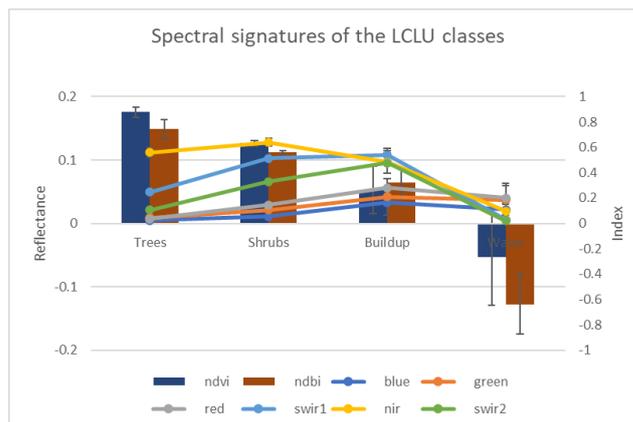


Figure 11: Spectral signatures of the LCLU classes

The overall accuracy and kappa coefficient shown in Table 4 below were highest (99.18%, 0.99; 99.38%, 0.99 and 99.38%, 0.99) respectively for the MLA and ANNs using all the eight (8) features (ALL) and using NN. Using three (3) features, the overall accuracy was slightly higher for the RGB features (98.76%) than the NNN features (98.56%) but same kappa coefficient for the MLA classification. For the ANNs classification, NNN features performed better (96.29%) than the RGB features (93.81%). Using NN gave very good results (99.38%, 0.99) and (96.91%, 0.95) for both MLA and ANN, respectively. The MLA classifier used both the mean and the variance of the features' values while the ANNs only used the mean values. The more the features used in classification, the higher the classification accuracies for both classifiers [15].

Table 4: Overall Accuracy and Kappa Coefficient

Features scheme	No of features	Matrix	Classification scheme	
			MLA	ANNs
ALL	8	OA	99.18%	99.38%
		KC	0.99	0.99
RGB	3	OA	98.76%	93.81%
		KC	0.98	0.89
NNN	3	OA	98.56%	96.29%
		KC	0.98	0.94
NN	2	OA	99.38%	96.91%
		KC	0.99	0.95

Table 5 below shows the Producer Accuracy and User Accuracy for the two classification schemes using eight and three spectral features and ratios of the land use land cover classes. Again, using all the eight spectra features gave the highest producer accuracies and user accuracies for both classification schemes. In all the classifications performed, the trees LCLU class has a 100% user accuracy followed by the water LCLU class (99.45%); buildup LCLU class (95.78); and finally, shrubs LCLU class (92.91%). The MLA classification user accuracy is 100% for trees

and water LCLU classes, 99.29% for shrubs LCLU classes and 95.53% for the buildup LCLU class. By studying the original satellite image, it is very clear that the build-up LCLU class has the most mixed pixels which might not adequately be resolved by the variance of the reflectance values.

Table 5: Producer and User Accuracies

Features scheme	No of features	Classification	LCLU	Matrix			
				PA (%)	UA (%)		
ALL	8	MLA	Trees	100	100		
			Water	100	100		
			Buildup	100	96.61		
			Shrubs	91.30	100		
		ANNs	Trees	93.02	100		
			Water	100	100		
			Buildup	100	100		
			shrubs	100	93.88		
		RGB	3	MLA	Trees	90.70	100
					Water	99.65	100
					Buildup	99.12	95.76
					Shrubs	100	97.87
ANNs	Trees			58.14	100		
	Water			100	97.72		
	Buildup			89.47	100		
	shrubs			100	65.71		
NNN	3			MLA	Trees	88.37	100
					Water	100	100
					Buildup	100	94.21
					Shrubs	95.65	100
		ANNs	Trees	72.09	100		
			Water	100	98.95		
			Buildup	97.37	88.10		
			shrubs	93.48	100		
		NN	2	MLA	Trees	100	100
					Water	100	100
					Buildup	100	97.44
					shrubs	93.48	100
ANNs	Trees			100	100		
	Water			100	96.25		
	Buildup			90.35	96.26		

shrubs 91.30 100

CONCLUSIONS

In all classifications carried out, the overall accuracy and kappa statistic varied from 93.81% and 0.89 to 99.38% and 0.99, respectively. The highest classification accuracies were obtained by either using all eight features or two features (indices only) for both classification schemes which affirms the importance of Normalized Difference Vegetation Index (NDVI) and Normalized Difference Build-up Index (NDBI) in LCLU mapping. These two indices are robust enough to be used to detect shrubs, trees, water, and build-up in a satellite image further, the ANNs classifier is also robust enough to be used for this classification. Although the MLA classifier used both the mean values and variance of the features, the ANNs classifier only used the mean values of the features. It is very clear that the build-up LCLU class has the most mixed pixels which might not adequately be resolved by the variance of the reflectance values.

REFERENCES

1. Statistics Botswana. Botswana Demographic Survey Report 2017
2. Ahmad, A., & Quegan, S. Analysis of Maximum Likelihood Classification on Multispectral Data. *Appl Geograph.* 2012; 6(129): 6425-6436.
3. Anderson, J, Hardy, E, Roach, JT, & Witmer, REA land use and land cover classification system for use with remote sensor data. In U.S. Geological Survey Professional Paper, 1976; 964
4. Arvidson, T, Goward, S, Gasch, J, & Williams, D. Landsat-7 long-term acquisition plan: development and validation. *Photogramm Eng Remote Sensing*, 2006; 72(10): 1137-1146.
5. Barrett, AB, Duivenvoorden, S, Salakpi, EE, Muthoka, J. M., Mwangi, J., Oliver, S., & Rowhani, P. Forecasting vegetation condition for drought early warning systems in pastoral communities in Kenya. *Remote Sensing of Environment*, May, 2020.
6. Chavez, P. S. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, 1998; 24(3): 459-479.
7. Collins, J. B., & Woodcock, C. E. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. *Remote Sensing of Environment*, 1996; 56:66-77.
8. Daldegan, G. A., Roberts, D. A., & Ribeiro, F. de F. Spectral mixture analysis in Google Earth Engine to model and delineate fire scars over a large extent and a long time-series in a rainforest-savanna transition zone. *Remote Sensing of Environment*, 2019.
9. Darkoh, M. B. K. The nature, causes and consequences of desertification in the drylands of Africa. *Land Degradation Dev*, 1998;1:1-20.
10. Dematte, J. A. M., Huete, A. R., Ferreira Jr., L. G., Nanni, M. R., Alves, M. C., & Fiorio, P. R. Methodology for Bare Soil Detection and Discrimination by Landsat TM Image. *The Open Remote Sensing Journal*, 2009; 2(1): 24-35.
11. Du, Y., Teillet, P. M., & Cihlar, J. Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection. *Remote Sensing of Environment*, 2002; 82(1): 123-134.
12. Fischer, I., Hennecke, F., Bannes, C., & Zell, A. JavaNNS: Java Neural Network Simulator. User Manual, Version 1.1 [Internet]. University of Tübingen; 2002.

13. Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., & Sibley, A., et al. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 2010; 114: 168-182.
14. Gessner, U., Machwitz, M., Conrad, C., & Dech, S. Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles. *Remote Sensing of Environment*, 2013; 129: 90-102.
15. Hansen, M. C., Egorov, A., Potapov, P. V., Stehman, S. V., Tyukavina, A., Turubanova, S. A., Roy, D. P., Goetz, S. J., Loveland, T. R., Ju, J., Kommareddy, A., Kovalsky, V., & Forsyth, C., Bents, T. Monitoring conterminous United States (CONUS) land cover change with Web-Enabled Landsat Data (WELD). *Remote Sensing of Environment*, 2014; 140: 466-484