

CAPABILITY OF RULE BASED CLASSIFICATION TECHNIQUE IN IDENTIFYING SPRAWL AREAS: A CASE STUDY OF AREAS AROUND ELDORET TOWN, KENYA

ODHIAMBO S¹., MWASI B²., NGETICH J.³

University of Eldoret, Kenya.

Email: sodhiambo53@yahoo.com

Abstract

Use of satellite images to map urban land uses has been successful to varying degrees since the launch of medium resolution sensors producing images of 30 m spatial resolution. With this resolution, the extent of urban settlements can be detected. However, details of urban land use classes cannot be identified from 30m resolution images. With higher resolution images such as 10m Sentinel 2, most urban land use classes can be identified with fairly high classification accuracy using pixel-based classification techniques. Land use details, and hence classification accuracy can be improved using object-based image analysis (OBIA) techniques with the high resolution images. These techniques combine spectral, textural and spatial information to distinguish objects related to information classes. This study used rule based classification OBIA algorithms to accurately map urban land use classes and isolate the emerging sprawl settlements around Eldoret Town. Specifically, the study applied five variations of vegetation indices to extract land use/cover data from high resolution Sentinel-2 images of 2020 to identify urban sprawl areas. The analysis identified eleven urban sprawl areas with an overall classification accuracy of 91.67% and Kappa coefficient of 0.90. The findings confirm that use of rule based classification technique in LULC classifications gives high classification accuracy results.

Keywords: Rule Based Image Analysis; Urban Sprawl Patterns

Introduction

The world's rapid rate of urbanization is continuously increasing urbanization has been defined as the conversion of natural spaces to built-up areas for residential, commercial, and industrial land uses. It is important to note that the growth of urban areas is not uniform worldwide. In the UN Urbanization Prospects report 2021, 30% (225.3 million) of the World's population was living in urban areas in 1950, 56.61% (4.46 billion) in 2021, and it is expected to increase to 68% (6.68 billion) by 2050. It is projected that 90% of the projected growth of the World's urban population between 2021 and 2050 will occur in Asia and Africa (Zhang et al, 2014).

The rapid population growth calls for provision of more housing, schools, transportation network and utilities resulting in rapid but skewed urbanization driving change in land use/cover (LULC) patterns. One of the impacts of urbanization is increase of impervious surfaces resulting in more water runoff and hence water pollution and flooding (Wilson et al. 2003) as well as increase in urban temperature. Uncontrolled increase of impervious surfaces on lands that were formally agricultural lands, forests, grasslands, water bodies and wetlands coupled with population growth results in scarcity of food, environmental pollutions, destruction of ecological structure, and unemployment (Maktav and Erbek 2005). Therefore, a technique that can extract data from built-up areas more efficiently is urgently required to provide the data to urban planners to control auto-expansion of built-up areas.

Satellite remote sensing has continuously provided data with different spectral and spatial resolutions ranging from course (1km), low (80m), medium (30m), high (10-15m) and ultra-high (below 10m). High and ultra-high spatial resolution images have provided solutions for mapping and monitoring urban growth (Zhang et al 2014). Consequently, several image classification techniques have been used to extract built-up areas from satellite data. Results of these classifications vary depending on the satellite data used and the classification technique used. For example, the study by Bhatta et al. (2007) demonstrated that LISS-IV images of 5.8m pixel size, though able to show the sprawl patterns, also suffer 15 to 20 percent overall accuracy when used for the classification of cityscapes by pixels and mixed classes. However, rule based classification techniques have proven effective for extracting land uses in terms of their spectral values, shape, texture, morphology among others.

This study uses rule based image classification technique to isolate built-up areas in a mixed land use environment in order to identify the sprawl areas from high resolution Sentinel-2 satellite image of 2020 around Eldoret town.

Materials and methods

Study Area

The study area is areas around a gazetted Eldoret Municipality boundary in Uasin Gishu County, Kenya, and covers 58 sub-locations. It is bounded by Latitudes 00°52' 00"N and 00° 18' 00"N and Longitudes 34° 51 '00"E and 35° 31' 00"E covering approximately 1973km². Most location of the study area, that is, western part (Turbo area), eastern part

(Ainabkoi and Moiben areas), northern part (Ziwa area) and southern (Kapseret and Cheptiret areas) receive an average rainfall of between 625 mm to 1,560 mm, with two distinct peaks between March and June, August and September. Dry periods occur between November and February. Temperatures range between 7°C and 29°C with dominant soil types being Orthic Ferralsol (Fo) and Humic Nitosols (Hn). Generally, these conditions are favorable for livestock keeping, crop and fish farming. The interpolated population figures of the study area (58 sub-locations) for the years 2000, 2016 and 2020, shows a steady population increase from 437,049 to 774, 201 and 870, 271 respectively. Figure 3 shows the study area map.

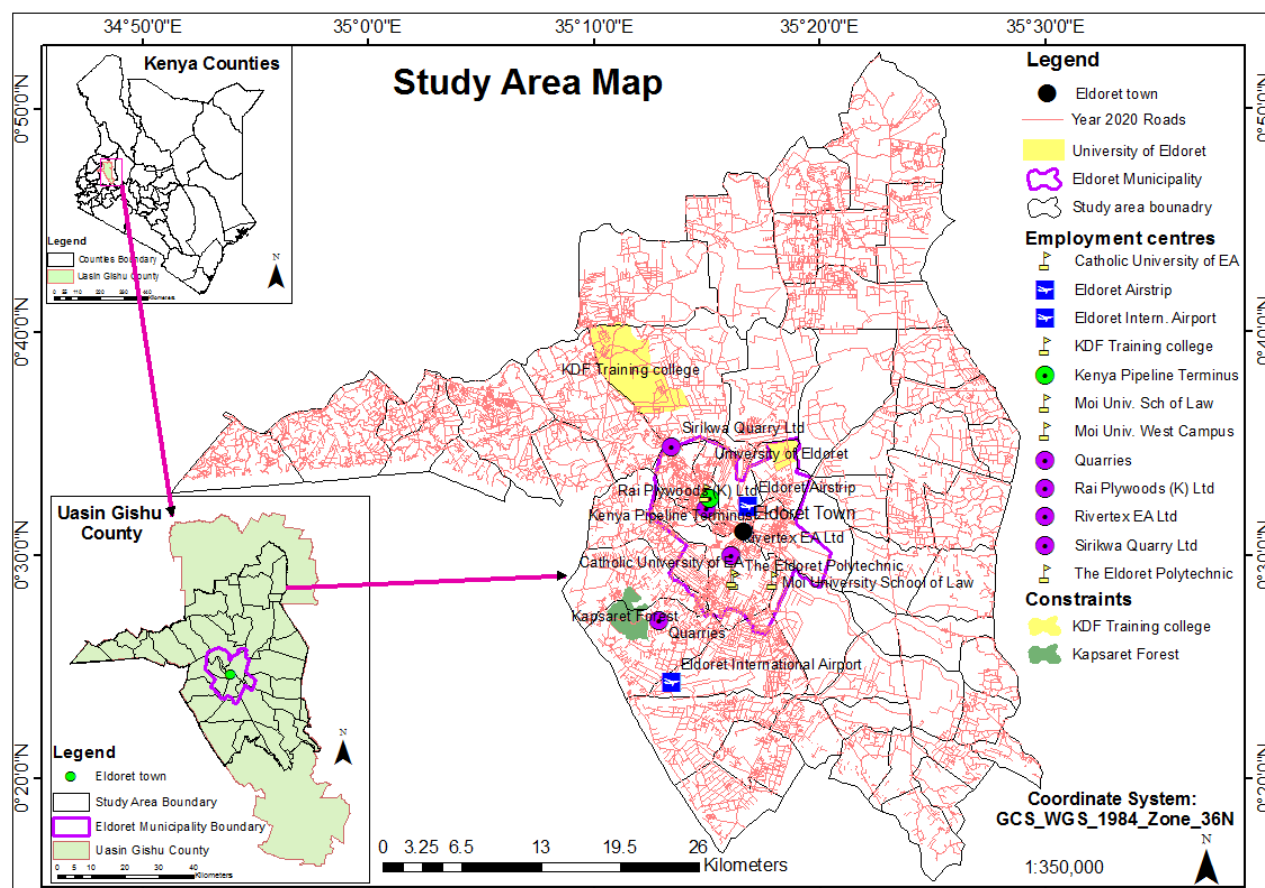


Fig. 3: Location of the study area

Data and Data Processing

Data for this study was obtained from both primary and secondary sources (Table 1).

Table 1: Characteristics of Data Used

S/No	Type of data used	Scale/Resolution	WRS_path/raw, Granules/Tiles	Years
1.	Sentinel-2 image	10m	T36NYF-100×100km	12 th Dec. 2020
2.	Sub-locations shapefile			2019

Primary Data

The data obtained from primary sources included; land use categories information obtained by classifying Sentinel-2 image for 12th December 2020 obtained from United States Geological Survey (USGS) website (USGS, <https://earthexplorer.usgs.gov/>).

Secondary Data

Secondary data used was sub-locations data covering the study area from Kenya National Bureau of Statistics (KNBS, 2019).

Methods

Figure 4 below describes the steps followed in obtaining urban sprawl areas for the year 2020.

Figure 4: Urban sprawl data extraction procedure

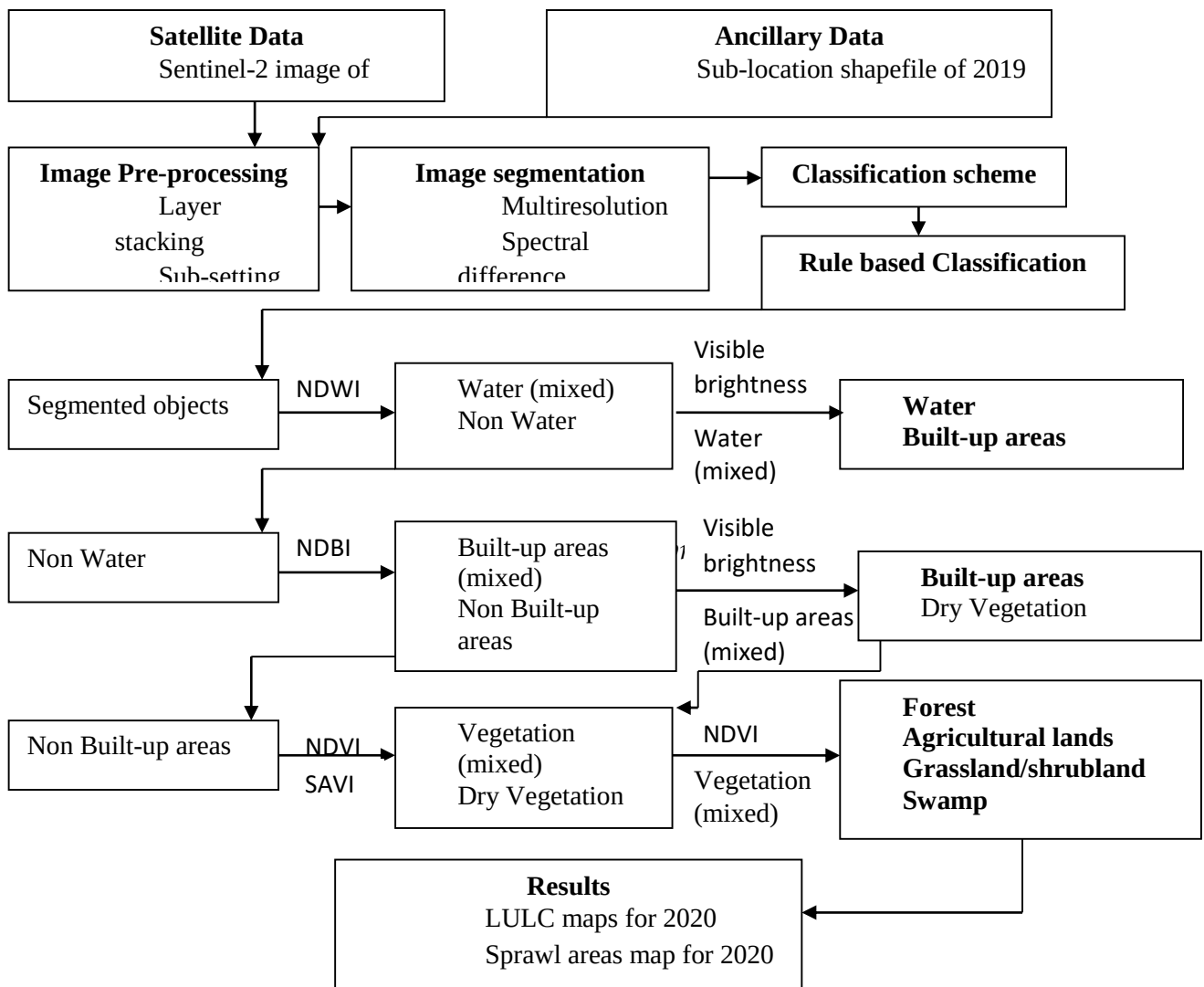


Image Pre-processing

Four bands were layer stacked for Sentinel-2 images of 2020 i.e. band 2 (blue), band 3 (green), band 4 (red) and band 8 (Near Infra-Red) all at 10 m spatial resolution. The composite image was then clipped using a sub-location shapefile to cover the study area.

Image Segmentation

Image segmentation was done in two steps namely; multiresolution and spectral difference segmentation algorithm.

In Multiresolution segmentation, the 'Scale parameter' was set to 35 since it's the restricting parameter that stops the object from getting too heterogeneous or is the average size of the object. Since there is no rule for 'Scale Parameter', trial and error was used and value 35 was found to be the best value for 'Scale Parameter'. 'Shape' (geometric form of the object) was set to weight of 0.3 since it defines the shape criterion to be used when segmenting the image. The higher its value, the lower the influence of colour on the segmented process hence care was taken not to put higher value to avoid distorting spectral information. Finally, 'Compactness' was set to value 0.6. The higher its value, the more compact the image object may be and the NIR value was set to 2.

Spectral Difference segmentation was performed next in order to merge already segmented objects in multiresolution segmentation with the same spectral values together during the segmentation process. 'Maximum Spectral Difference' value was set to 15 and the NIR band to 3.

Land Cover Classification Scheme

Due to heterogeneous nature of the study area, a list of six land cover classes was identified during a reconnaissance survey of the study area considering their exhaustiveness to accommodate all land cover features (Table 2).

Table 2: Major Land Cover Classes in the Study Area

S/No	Name	Description
1.	Built-up area	Human constructed structures, buildings, roads and other impervious surfaces
2.	Water	Rivers, ponds and other water bodies
3.	Forest	Both man-made and natural
4.	Agricultural lands	Both cultivated and non-cultivated
5.	Grassland/Shrubland	Both natural and planted
6.	Swamps	Both permanent and seasonal

The six land cover classes were inserted into the image processing software under class hierarchy.

Rule-Based Classification of 2020 image

Rule-based classification was done using eCognition developer. Five indices were used i.e. normalized difference vegetation index (NDVI), normalized difference water index (NDWI), normalized difference built-up index (NDBI), Visible brightness and soil adjusted vegetation index (SAVI).

First, NDWI $(\text{Green}-\text{NIR})/(\text{Green}+\text{NIR})$ was applied on the segmented image and threshold NDWI values set between 0.5 to 1. The software assigns all objects in the image meeting the threshold into water class or else any object outside the threshold values is assigned to non-water objects. Since built-up areas also have high NDWI value (McFeeters, 1996), visible brightness $(\text{Red}+\text{Green}+\text{Blue})/3$ was applied on water class to isolate built-up areas from it since built-up areas have higher values of visible brightness value than water features.

Secondly, NDBI $(\text{SWIR}-\text{NIR})/(\text{SWIR}+\text{NIR})$ was applied on non-water objects since they have higher reflectance value of SWIR than NIR and threshold NDBI values set between -1 to 1. The software assigns all non-water objects in the image meeting the threshold into built-up areas class or else it assigns them to non-built-up areas. Since drier vegetation also possesses higher NDBI value (Zha et al., 2003), visible brightness $(\text{Red}+\text{Green}+\text{Blue})/3$ was applied because built-up areas have higher values of visible brightness value than drier vegetation features.

Third, NDVI $(\text{NIR}-\text{Red})/(\text{NIR}+\text{Red})$ was applied on the non-built-up area objects and threshold NDVI values set between -1 to 1. All objects meeting the threshold were assigned into vegetation class and together with drier vegetation isolated by NDBI above. SAVI $(\text{NIR}-\text{Red})/(1+k)/(\text{NIR}+\text{Red}+k)$ was applied to extract vegetation less than 15% cover mostly within Eldoret town since it has high reflectance in the NIR band. 'K' is the correction factor that ranges from 0-1 for very high vegetation density to very low density respectively. A correction factor of 0.6 was used. More threshold values were set to ungroup vegetation into forest, agricultural lands, grassland/shrubland and swamps.

After grouping the objects into the six LULC classes, classification was executed, and the raster classified image exported to ArcMap for post classification and generation of statistics.

2.3.5 Accuracy Assessment

Accuracy of classification results was done by creating a confusion matrix. The process produces four metrics namely, the user's accuracy, the producer's accuracy, the overall accuracy with and Kappa statistic (Congalton, 1991a). The producer's accuracy gives the percentage of correctly classified ground truth sites for each class. The user's accuracy gives the proportion of correctly classified sites in the classified image for each class while the overall accuracy is a combination of the two accuracy measures. The Kappa statistic expresses the probability that the values presented in the error matrix are significantly different from those from random samples of equal size (Benjamin, 2004). Sample points were randomly selected as reference data from original Sentinel-2 and google earth images of 2020. Thirty test pixels for each class were considered to be the best sample for assessing

accuracy (Zhao, 2013). The test pixel was then overlaid on the classified image in order to generate a confusion matrix table. From the confusion matrix table, the four matrices were calculated using the formulae below;

$$1) \text{ Overall accuracy} = \frac{\text{Total No. of correctly classified pixels (diagonal)}}{\text{Total No. of Reference pixels}} * 100$$

$$2) \text{ User accuracy} = \frac{\text{No. of correctly classified pixels in each category}}{\text{Total No. of classified pixels in that category (total row)}} * 100$$

$$3) \text{ Producer accuracy} = \frac{\text{No. of correctly classified pixels in each category}}{\text{Total No. of Reference pixels in that category (total column)}} * 100$$

$$4) \text{ Kappa coefficient (T)} = \frac{(TS * TCS) - \sum (\text{column total} * \text{row total})}{TS^2 - \sum (\text{column total} * \text{row total})} * 100$$

Where; TS is Total samples; TCS is Total correctly classified samples

Finally, the LULC map for 2020 was then used to identify and map sprawl areas, and designed in ArcMap. Sprawl areas map 2020 was then finally generated.

Results and Discussions

LULC Maps for 2020

Fig 5 shows the LULC map for the year 2020.

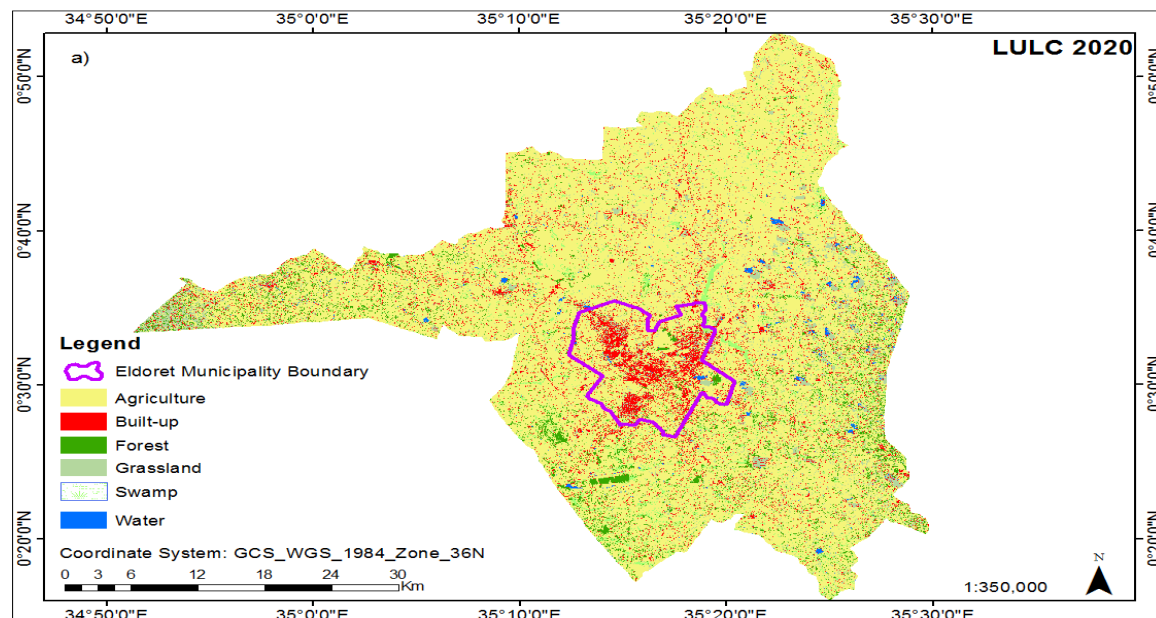


Figure 6: LULC Map of the study Area in Year 2020

Classification techniques by Rule-Based using the five indices (NDVI, NDWI, NDBI, SAVI and visible brightness) discriminated the six land cover classes i.e. built-up areas, agriculture/farmlands, water, grassland/shrubland, swamps and forest by use of their spectral values. Similar findings were reported by Mwakapuja et al (2013) in their study

of usage of indexes for extraction of built-up areas and vegetation features from Landsat TM images, a case study of Dar Es Salaam and Kisarawe Peri-Urban areas in Tanzania.

Table 3: Summary of Land Use/Cover Classification Statistics for 2020 (area in km²)

LULC Type	2020	
	Area (km ²)	(%)
Agriculture/Farmland	1524.82	77.29
Built up areas	138.912	7.04
Forest	106.757	5.41
Grassland/shrubland	123.475	6.26
Swamp	67.3586	3.41
Water	11.4492	0.58
Total	1972.769	100

3. 2 Classification accuracies for 2020

Table 4: User's and Producer's Accuracy Results by Rule Based Algorithm

LULC Type	2020	
	Producer's (%)	User's (%)
Agriculture/Farmland	87.10	90.00
Built-up Areas	96.55	93.33
Forest	100	90.00
Grassland	81.25	86.67
Swamp	90.32	93.33
Water	96.67	96.67

Table 5: Classification Accuracy Assessment for 2020 by Rule Based Classification Algorithm

Image	Overall Accuracy (%)	KappaCoefficient (%)
2020	91.67	90

90%) and hence can be used in extracting land covers in areas with similar characteristics since the LULC classes of particular interest for this study namely built-up area, forest, water, grassland/ or shrub land, swamp and agricultural or farmlands were isolated efficiently and accurately from Sentinel-2 images of 2020 by use of indices.

Disclosure statement: No potential conflict of interest was reported by author(s)

References

Anderson et al (1976). A Land Use and Land Cover Classification System for Use with Remote Sensor Data. Geological Survey Professional Paper No. 964, U.S. Government Printing Office, Washington DC, 28.

Bhatta R., Vaithyanathan, S., Singh N. P., Verma, D. L., (2007). Effect of feeding complete diets containing graded levels of *Prosopis cineraria* leaves on feed intake, nutrient utilization and rumen fermentation in lambs and kids. *Small Rum. Res.*, 67 (1): 75-83

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, 37(1):35-46.

Kenya National Bureau of Statistics (KNBS 2019). Kenya Population and Housing Census Reports

Maktav D., Erbek, F. S. (2005). Analysis of urban growth using multitemporal satellite data in Istanbul, Turkey. *International Journal of Remote Sensing*, 26(4): 797-810.

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, 17(7):1425-1432.

Mwakapuja F., Liwa E., Kashaigili J. (2013). Usage of indices for extraction of built-up areas and vegetation features from Landsat TM image: A case of Dar es Salaam and Kisarawe peri-urban areas, Tanzania, 3(7): 273-283

Mwasi B. (2004). Landscape change dynamics in semi-arid part of Baringo district, Kenya based on Landsat TM data and GIS analysis

Wilson M. (2003). Discovery Listening: Improving Perceptual Processing. *ELT Journal*, 57, 335-343.<http://dx.doi.org/10.1093/elt/57.4.335>

Zha, Y., Gao, J., Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International journal of remote sensing*, 24(3): 583-594.

Zhang J., Li P., Wang, J. (2014). Urban built-up area extraction from Landsat TM/ETM+ images using spectral information and multivariate texture. *Remote Sensing*, 6(8):7339-7359.

Zhao, P. (2013). The impact of urban sprawl on social segregation in Beijing and a limited role for spatial planning. *Tijdschrift voor economische en sociale geografie*, 104(5): 571-587.