Classification of Urban Land Use and Land Cover with K-Nearest Neighbour Classifier in the City of Cape Town, South Africa – Cape Flats Case Study.

Bosiu E. Lefulebe [1], Adriaan Van der Walt [1,3,*], and Sifiso Xulu [2]

1 Department of Geography, University of the Free State, Bloemfontein 9301, South Africa;

2 School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4000, South Africa
bosiulefulebe@gmail.com (B.E.L.); xulus4@ukzn.ac.za (S.X.)

3 Afromontane Research Unit (ARU), Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein 9301, South Africa

***** Correspondence: vanderwalta@ufs.ac.za; Tel.: +27-051-401-9653

KEY WORDS: machine learning, PlanetScope, K-nearest neighbour, urban land use and land cover

ABSTRACT:

The rapid growth of cities, owing to rural-urban migration and high birth rate has resulted in encroachment of the land use and land cover (LULC) in the Cape Flats, situated north of Cape Town, Western Cape, South Africa. PlanetScope imagery and K-Nearest Neighbour (KNN) classifier are used to map and monitor the following LULC classes within the Cape Flats; waterbodies, trees, built-up, and vegetation, we further do a time series map between 2016 and 2021.
The results showed that LULC changes between 2016 to 2021 are as follows: a decrease of 2.1% in vegetation, 0.4% in water bodies, and 11% in trees. Built-up areas, on the other hand, showed a significant increase of 13.6% over five years. The LULC changes in the Cape Flats were mainly triggered by an increase in built-up areas due to household construction to accommodate the increased population resulting from rural-to-urban migration and high birth rate. Classification accuracy from 2016 and 2021 was as follows; overall accuracy of 98.31% and 0.97 kappa coefficient in 2016, while overall accuracy 96.54% and kappa coefficient 0.95 in 2021. A combination of machine learning and high-resolution imagery showcased that high classification results can be achieved in monitoring subtle LULC changes. We recommended that all relevant stakeholders, including government officials and municipalities, formulate and adopt policies to protect against LULC degradation.

1. MANUSCRIPT

1.1 Introduction

The Cape Flats fall under Cape Town suburbs in the Western Province of South Africa. Cape Flats are within a low laying area and are surrounded by Northern Suburbs towards the North, Helderberg towards the East, Southern Suburbs towards the West, Peninsula on the South west, and the Atlantic Ocean to the South as indicated in figure 1-1 below. Cape Flats are characterised by winter floods which normally take place in July and August, owing to its low-laying topography surrounded by a high-laying terrain. Like most informal settlements and townships, Cape Flats have a limited drainage system, and most of them have been destroyed by land infilling to cater for accommodation (Lefulebe et al., 2015).
Historically Cape Flats suburbs were used for military training, farmers used them for vegetable farming nevertheless, all LULC activities disappeared because of the urbanisation that took place. The suburb is also associated with a high crime rate and is further classified as one of the most dangerous places in South Africa. South African National Defence Force has been deployed in Cape Flats Suburbs to combat crime, in past years.
The city of Cape Town's economy is built on five strategic areas (City of Cape Town, 2013);
a) Competitive city by applying and adapting regulatory changes.
b) Infrastructure, basic services including transport and information and communications technology.
c) Inclusive growth, applying work and skills programmes to enhance growth.
d) Trade and sector development, are maximised to the greatest advantage of the residents.
e) Environmental sustainability; ensuring long-term growth that is also sustainable.

The above-mentioned opportunities attract migration to the city, even though the biggest threat is rural-urban migration. According to South African Gateway (2020), Western Cape province migration statistics is 175 831 for outgoing migrants, and 485 560 incoming migrants, thereby resulting in a net migration of 309 729 2016 and 2021.

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-967-2023

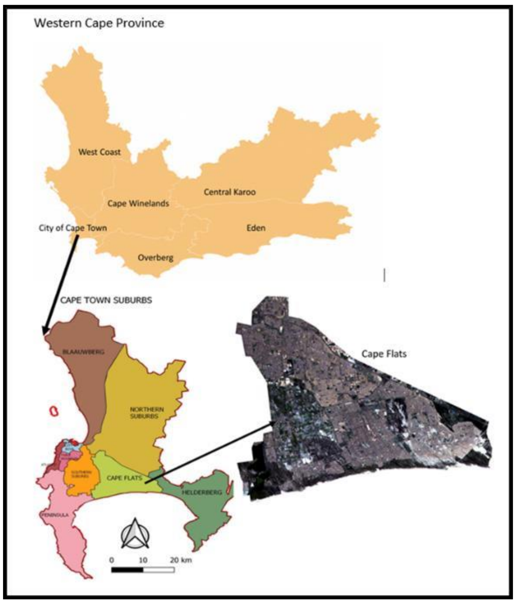

Figure 1-1 Location of the Cape Flats in the City of Cape Town, South Africa.

We chose the Cape Flats as a study area because it is a developing suburb, situated in the Western Cape province which is characterised by a high in-migration number between 2016 and 2021 and high birth rate, therefore it was crucial to understand the human footprint there for a period of over five years. It is further important to understand how rural-to-urban migration and high birth rate affects LULC in a township suburb. This study aims to use remote sensing data, PlanetScope imagery and KNN classifier to help elucidate the impact of rural-urban migration on LULC.

1.2 Literature

The theoretical basis of this study is predicated on the following literature; rural-urban migration, Urbanisation, land use and land cover, KNN classifier and change detection.

1.2.1 Urbanisation

Urbanisation is a process whereby cities become larger. The main reason for urbanisation is rural-urban migration and high birth rate (Mahbubur Rahman *et al*., 2018; Bodo, 2019; Njwambe *et al*., 2019). There are numerous reasons that accelerate rural-urban migration, the most important include economic reasons, better job opportunities and a better quality of life (Ruiz *et al*., 2013; Lefulebe *et al*., 2015; Njwambe *et al*., 2019). Other reasons in the area of social factors include better standards of households and services, such as good education and better health care, while in the area of the environment, people hope for better living conditions, such as a lower incidence of natural disasters. Rural areas currently have vastly poor housing conditions, a high rate of unemployment and low wages, and finally lack adequate sanitation and deficiency of clean drinking water (Lefulebe *et al*., 2015; Ochuko, 2016).

Rapid urbanisation is characterised by an increased number of residential buildings, industrial, commercial and transportation routes, however there is a drastic decrease of vacant land, while in other cases land use is changed. A typical example would be agricultural land being transformed to residential, this affects food security as less land is available to produce food (Rana and Marwasta, 2015).

Rural-urban migration is taking place at an alarming rate. It is expected that 2.5 billion people worldwide will migrate to cities in the next thirty years, and this is expected to occur mostly in Africa and Asia (Bhatta, 2010). The appreciation of these trends therefore become prime. According to South African Gateway (2020), the migration statistics for the Western Cape Province are 175 831 out-migrating and 485 560 in-migrating, resulting in a net migration of 309 729 in 2016 and 2021, it therefore becomes crucial to understand the human footprint in these areas. It is also important to understand how rural-urban migration affects the LULC in an urban suburb.

1.2.2 Land use and land cover (LULC)

LULC changes are accelerated by several factors and activities taking place on the earth's surface. These factors include livestock farming, forest management and harvesting and agriculture (Owoeye and Ibitoye, 2016). Due to the nature of developing countries, it is expected that rapid changes will occur in LULC as a result of the growth of cities resulting from rural-urban migration and high birth rate, as compared to developed countries. The Cape Flats are subject to this change as they are in South Africa which is classified as a developing country. The Western Cape Province is distinctly subject to high population growth, a high rate of household construction, a high rate of rural-urban migration, high birth rate and infrastructure development (Maree and Van Weele, 2013).

LULC terms are often confused, and their meaning is different. Land use is defined as how humans use land, both in terms of economic and cultural activities (Di Gregorio, and Jansen, 2000; NOAA, 2022), this includes residential, agricultural, mining, etc. Land cover specifies the physical land type, an illustration of the amount of land region covered by forests, agriculture, wetlands, waterbodies and other lands (Di Gregorio, and Jansen, 2000; NOAA, 2022). Additionally, different types of land cover are used differently, for instance; forests are used to produce wood while agricultural land is used to produce food. Land cover is determined by remotely sensed data such as satellite images, acquired at different periods. Comparison and analysis of land cover maps assist in understanding changes that occurred in the period in question, therefore informed decisions can be made. There are several stakeholders interested in land cover maps, a typical example is Coastal managers, who use such maps to understand the impacts of nature and human use on the landscape. The impact of rural-to-urban migration and high birth rate on LULC is the rapid growth of informal settlements, growth of built-up areas, and a decrease of vacant land, wetlands and the occupation of flood-prone areas. (Mahbubur Rahman et al., 2018).

Remote sensing data and GIS have been applied in LULC related studies. Mahbubur Rahman et al. (2018) conducted a study in Dhaka city, Bangladesh, whereby the authors explored the impact of rural-to-urban migration on LULC. Remote sensing and GIS were used in this study from 2006, 2010 and 2016. Informal settlements were digitised from Google high-resolution images. The major finding from this study was that built-up areas had increased, with informal settlements increasing rapidly, and vegetated and bare land areas decreasing as a result of rural-urban migration.

Another study conducted by Chouhan and Kannan (2019), evaluated the impact of urbanisation on LU patterns and the environment in Ajmer city, Rajasthan. Satellite images captured at different intervals between 2000 and 2017 were used to analyse and map land use patterns, GIS software used for change detection was ERDAS and ArcGIS. The authors applied the supervised classification to the images to analyse LU change. The findings from this study were that a large area of agricultural land was lost consequently from a high demand to construct residential houses, and market places. The inflow of water into the lakes changed as some of the households were constructed on the hill slopes, and forest areas were lost owing to mining activities, thus changing the LU pattern.

1.2.3 Supervised classification (K-Nearest Neighbour machine learning Classifier)

Supervised classification is dependent on human guidance to produce results, sample pixels in an image are selected by the user, they represent specific classes which may include trees, grass, water, bare land etc. A polygon is used to select classes. Training samples are vital as they determine overall classification accuracy. The user further guides the software to make use of created trained sample pixels as reference for the entire classification of other pixels in the image (Madariya *et al*., 2022). Three major identified steps in supervised classification are; training samples, extraction of signature and lastly image classification. User knowledge is crucial when determining input classes. Bonds are also set to determine the similarity on which pixels should be grouped together by the user. The following factors are taken into consideration when setting bonds; training area spectral characteristics, increment based on brightness, which is determined by reflection from image spectral bands. The number of output classes are also determined by the user.

Supervised classification technique requires the user to have prior knowledge of land cover in the study area. This is essential because hyperspectral data generated by pixels is required to train the classification algorithm before the model is run.

The KNN is a non-parametric and supervised algorithm, which is used to solve regression and classification problems (Atkeson et al., 1997, Han and Kamber, 2000, Guo et al., 2003). In principle, KNN assumes that similar features exist nearby. The classifier predicts classes by calculating the nearest distance between test and training data (Guo et al., 2003).

1.2.4 Land use and land cover change detection

Change Detection is defined as a process of recognizing differences on the earth's surface over a certain period, using remote sensing data such as satellite images acquired over different periods to monitor changes in LULC (Théau, 2008).

Updated information regarding the earth's surface has increased tremendously as it forms the basis of several applications such as LULC change monitoring, and resources monitoring at local, regional and global scales (Hussain et al., 2013).

Remote sensing Satellite data has been a backbone to change detection studies, and there have been huge improvements in the quality of satellite resolution in previous years, from low to high-resolution images, therefore improving the quality of change detection results.

A lot of research has been conducted regarding LULC change detection, however, very few studies focused on the impact of rural-urban migration on LULC, the following section illustrates some of the studies conducted. Abijith and Saravanan (2021) assessed LULC change detection and prediction using remote sensing data (Landsat images) at CA Markov, in the northern coastal districts of Tamil Nadu, India. LULC changes from 2009 to 2019 were conducted using Google Earth Engine (GEE), Terrset and GIS tools. The results indicated that the waterbodies were decreasing, while built-up areas increased. It was noted further that bare land and vegetation were found to be under stress as the majority of areas they covered were changing into buildings. Kappa coefficient of 87% and overall accuracy of 89% were achieved to validate the map.

1.3 Approach and Methodology

This section outlines the methodology applied in mapping LULC using PlanetScope satellite imagery and KNN classifier machine learning method. To date, the hypothesis is that the KNN classifier can classify LULC and other features within the urban city with high accuracy. This study aims at mapping LULC and thereafter performing change detection between 2016 and 2021 within the Cape Flats in the Western Cape, Cape Town. The produced time series map will help relevant stakeholders to understand the human footprint in Cape Flats township, resulting from rural-urban migration.

Image classification that was carried out in this study belongs to the supervised classification group, and major steps in supervised classification were followed in the methodology. Figure 1-2, is an overview of the methodology used in this study, it is summarised into the following six phases, image acquisition, pre-processing, training samples, image classification, results validation, and analysis.

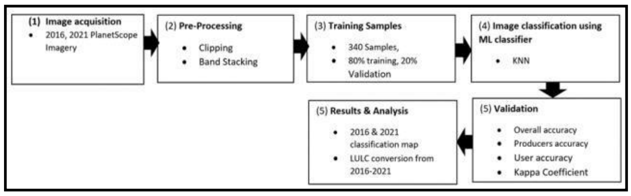

Figure 1-2 Methodology flowchart.

1.3.1 Preparation for data collection and pre-processing

The following section explains how data used in this study was collected, it further elaborates on the type of data and information obtained. Data collected includes Raster and Vector data from different organisations, the City of Cape Town, Planet Labs, and Vector Stack. Collected data includes satellite images and the demarcation of Suburbs within the City of Cape Town. Table 1-1 below is a list of data sets used in this research.

Data Set	Type	Supplier
City of Cape Town Boundary	Vector	City of Cape Town
Cape Town Suburbs Map	Raster	Vector Stack
PlanetScope Satellite Image	Raster	Planet Lab Inc.

Table 1-1 Datasets used in this research

a) City of Cape Town Boundary; was obtained as vector format shapefile from the Department of Information and Knowledge Management under the City of Cape Town. Information contained in this shapefile is demarcations of the entire city, which covers an area of two thousand four hundred and forty-five square meters (2445 km²) with an estimated population of four million seven hundred and ten thousand (4710000) in 2021. This data was useful in creating study area maps.

b) Cape Town Suburbs map; the City of Cape Town Metropolitan Municipality consists of nine suburbs namely Blaauwberg, Northern Suburbs, Helderberg, Cape Flats, Peninsula, Southern Suburbs, Atlantic Seaboard, Table Mountain National Park, and City Bowl. The Cape Town Suburbs map was downloaded from the Vector Stock website in a raster format. This

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-967-2023 | © Author(s) 2023. CC BY 4.0 License.

map was necessary for aiding with the extraction of study area limits of Cape Flats. The area covered by Cape Flats is two hundred and four square kilometres (204 km²). It is worth noting that the Cape Town Suburbs map was not georeferenced, therefore City of Cape Town Metropolitan Municipality Map was used to georeference City of Cape Town Metropolitan Municipality Suburbs by identifying similar vertices on both maps and applying spatial co-ordinates to geo-rectify. Once this phase was completed, it was followed by the digitization of study area boundaries on the Cape Flats. The boundaries were stored as a Shapefile, to enable easy data integration in a GIS system.

c) PlanetScope Image is a ready-to-use format, pre-processed for geometric, radiometric, and atmospheric corrections at Level 3B surface reflectance and orthorectified (Planet, 2020), it consists of the following channels Blue, Green, Red, and Near-Infrared. The image is sampled at 3 meters resolution (https://www.planet.com).

1.3.2 Satellite image processing

After the study data was acquired, it was necessary to pre-process and make data suitable for further analysis. Satellite images downloaded from planet labs extended beyond the study area, it was, therefore, necessary to clip the images to the study area polygon in QGIS software. Larger images take longer to process therefore it became vital to constrain the image to the study area only.

It was not necessary to pre-process the image, besides performing band stacking. Several satellite images covered the study area, and this necessitated the splitting of all imagery into four bands (R, G, B and NIR). The same bands from different images were merged without resampling as they all had the exact resolution, thus creating four-band composition imagery built from merged bands as separate layers.

1.3.3 Training samples

Supervised classification requires composite satellite imagery and trained samples. Therefore, based on the study's objective, four classes were created: (1) vegetation, (2) water bodies, (3) Built-up and (4) trees. These training samples are used to determine the overall areas of the four classes. A polygon shapefile was created, and different classes were randomly digitised throughout the study area from the satellite imagery. A unique ID number integer 64 was allocated to all digitised features, for example; all water bodies were assigned ID 2, while trees were assigned ID 4. Cape Flats had 340 samples. Large classification areas need more training samples when compared to small areas. Figure 1-3 below illustrates the attribute table of classes contained in the training layer. The image illustrates training polygons together with matching colours to their class, as indicated in the legend.

A consistent representation of training samples statistics over land cover is required to be collected (Congalton and Green 2009), Cape Flats consisted of 340 training samples made of 85 vegetation, 85 water bodies, 85 built-up and 85 trees.

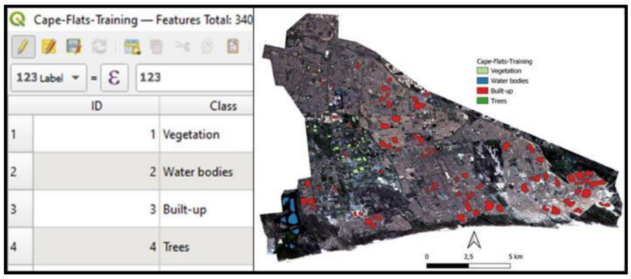

Figure 1-3 Cape Flats sample training layer

LULC classes identified are described as follows:

a) Vegetation consists of grass, crops, flowers and all plants that exist within the study area, excluding trees. When creating training samples under this class, samples were digitised for all vegetation members to ensure that all the vegetation is classified with good accuracy. The majority of vegetation is greener in the 2016 and 2021 images, indicating healthy status. Therefore, misclassification is expected to be very minimal should it be present.

b) Water bodies; include all accumulation of water on the surface of the land, and are categorised into two groups viz; artificial and natural water bodies. Manmade swimming pools and agricultural irrigation dams fall under manmade water bodies, while rivers streams, canals, ponds, wetlands and puddles fall under natural water bodies. All different waterbodies were digitised to avoid misclassification as each category reflected a different colour on the images, for instance, swimming pools reflected bluer on the images while wetlands reflected dark green.

c) Built-up class consists of buildings within the study area and all urban features that could not be categorised in any of the other classes, such as roads, pavements, tennis courts, and soil. The majority of buildings in the study area are households, shopping centres, hospitals, schools and many more.

d) Trees belong under vegetation (plants); however, they are put in a separate class as one of the objectives of this study is to quantify tree canopy coverage. Trees are characterized by the following features, stem, height and branches forming a canopy. A collection of densely populated trees forms a forest, and, forests growing in urban cities fall under urban forests. In this study the following urban forests exist; peri-urban forests and woodlands, city parks and urban forests, pocket parks and gardens with trees, trees on streets or in public squares, and other green spaces with trees.

1.3.4 Selecting classifier

KNN classifier was used in this study after literature consultation, based on the merits that it performed exceptionally well in other urban LULC studies. The classifier was applied in a supervised classification process to map LULC in the Cape Flats. The parameters employed to optimize the model are described below. Orfeo (OTB) software was used for the application and modelling of LULC maps, as a running platform for the classifier.

1.3.5 OTB supervised image classification procedure

OTB software supervised process is illustrated in Figure 1-4 below. All workflow steps including inputs and outputs are also explained covering every step. The process is simplified into five major steps which are; loading the raster, computation of image statistics, Training the image classifier, and creating image classification (www.orfeo-toolbox.org, accessed 10 March 2022).

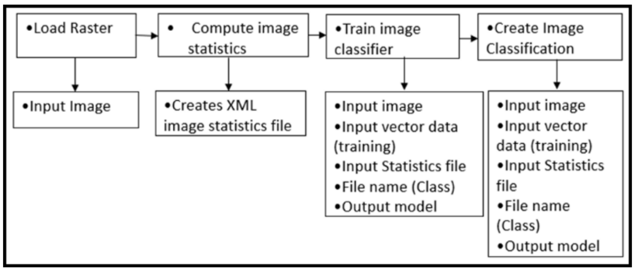

Figure 1-4 OTB Supervised image classification workflow

a) Load raster performs by selecting the image to be classified and also creating a training sample within the area of interest. In our case, four classes are; water bodies, built-up, vegetation and trees. A vector polygon shapefile layer is therefore created out of all classes, by digitizing samples of the image. Each polygon is assigned a unique ID that will be used in training the classifier.

b) In compute image statistics of the image to be classified, the input is the image and the output is an XML file. This file contains the variance and means of each feature. It further helps to understand the number of samples available in each of the four classes. Information produced is the number of samples per class and geometry. Only vector samples are supported in this application, in this case, polygons. The global mean and standard deviation are also computed from the image. The output statistics file with XML extension is then used in train image classifier and it serves the purpose of normalizing samples before learning. The computers Random-Access Memory (RAM) is critical as less memory terminates the process operation. The limitation when computing image statistics is that input images should have the same band order as other images.

c) At the stage of train image classifier; the image classifier is being trained. The following are used as inputs; image, vector data list (training samples layer), image statistics file (XML), number of training and validation sample ratio, and the classifier to use for training. The validation ratio uses all training layer geometry and therefore same fields used in training will be included. Outputs are estimated model files and the confusion matrix.

d) When using image classification; input files are; image, model file, statistics file. The output is a classified image. Image classification is performed by rendering the model file, as created by the train image classier application. Only image pixels with corresponding mask values above 0 value are classified. Pixels not classified are by default assigned 0 value on the output image. The classified image represents land use as per set classes.

1.3.6 Classification optimisation parameters

Normally optimization parameters for image training and classification are fine-tuned to produce the best possible results with classifiers. However, Immitzer et al. (2016), Trisasongko et al. (2017), and Luca et al. (2019), reported that optimization default values used in OTB for both processes of training and image classification produce highly accurate results. Therefore, default parameters were adopted in this study.

1.3.7 Validation and accuracy analysis

A confusion matrix (CM) is a common approach used to assess the quality and accuracy of classification maps. Additionally, the following accuracy assessment measures are obtained from CM, Overall accuracy (OA), Producers accuracy PA), Users accuracy (UA) and Kappa Coefficient (k)

a) Confusion matrix; is employed in this study as another measure of map validation, it determines the detection probability of land use classes (Pacheco et al., 2021). It is further seen as a technique of summarizing the performance of the classifier algorithm. The fundamental backbone of the confusion matrix is summarizing the total number of correct and incorrect predictions, including count values, broken down to each class. The confusion matrix thus gives a holistic view of errors being made by the classifier.

b) Overall accuracy is based on correctly classified pixels over a total sum of pixels and therefore it is represented as a ratio (equation 1). Post computation of the confusion matrix, the total sum of entries at the diagonal was divided by a total number of validation samples to obtain the OA.

$$OA = \frac{Number\ of\ correct\ predictions}{\text{Total number of Predictions}} \quad (1)$$

c) Producers accuracy; is represented as a ratio of correctly classified pixels of a certain class against reference pixels of that certain class (equation 2).

$$PA = \frac{Number\ correctly\ identified\ in\ a\ given\ map\ class}{\text{Number actually in that reference class}} \quad (2)$$

d) User accuracy; Correctly classified pixels of a given class to all pixels classified in this class, presented as ratio. The number correctly identified in a given map class divided by the number claimed to be in that map class, equals UA (equation 3).

$$UA = \frac{Number\ correctly\ identified\ in\ a\ given\ map\ class}{\text{Number claimed to be in that map class}} \quad (3)$$

e) Kappa coefficient is an accuracy basis to measure correspondence between datasets, Kappa value of 0 or less indicates that classification results are not usable. The ideal Kappa value range is 1 or close. The k was calculated from the confusion matrix (equation 4).

$$k = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+1} \times x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} \times x_{+i})} \quad (4)$$

where r = the number of rows and columns in the error matrix, xii = the number of observations in row i and column i, xi+ = the marginal sum of row i, x+i = the marginal sum of column i, and N = the total number of observations.

1.4 Results

Cape Flats main LULC classes were classified using KNN classifier based on PlanetScope image of 2016 and 2021. The impact of urbanisation that took place in the Cape Flats between 2016 and 2021 was then informed by change detection, to assist in understanding how different LULC classes have been affected by urbanisation.

1.4.1 Land use and land cover classification

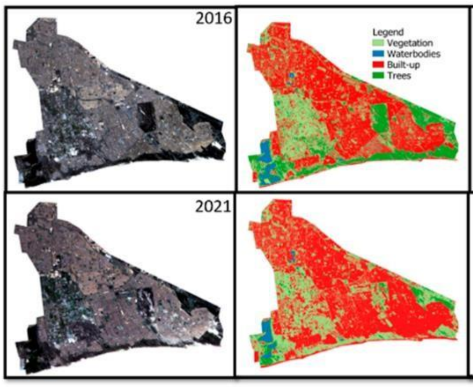

Figure 1-5 The Cape Flats, true PlanetScope imagery and K-nearest neighbour (KNN) classifier, 2016 top pannel and 2021 bottom panel

2016			2021	
Class	Percentage %	Area [km^2]	Percentage %	Area [km^2]
Vegetation	33.3	67256.4	31.2	62953.1
Waterbodies	2.8	5608.9	2.4	4746.38
Built-up	48.5	97871.8	62.1	125321
Trees	15.5	31209.8	4.4	8926.4

Table 1-2 Cape Flats representation of urban land use and land cover (LULC) conversion from 2016 to 2021

	> Reference					
V_Classified	Vegetation	Waterbodies	Built-up	Trees	Area	Weight
Vegetation	0.1687	0.0000	0.0022	0.0011	67256379	0.1721
Waterbodies	0.0000	0.4976	0.0000	0.0001	194494059	0.4976
Built-up	0.0012	0.0000	0.2486	0.0006	97871778	0.2504
Trees	0.0012	0.0008	0.0002	0.0777	31209804	0.0799
Total	0.1711	0.4984	0.2510	0.0795	390832020	
Area	66885699	194781469	98082545	31082307	390832020	
SE	0.0001	0.0000	0.0001	0.0001		
SE area	56488	18649	54396	44249		
95% CI area	110716	36551	106617	86728		
PA [%]	98.5634	99.8379	99.0443	97.6950		
UA [%]	98.0202	99.9855	99.2576	97.2959		
Kappa hat	0.9761	0.9997	0.9901	0.9706		
Overall accuracy [%] = 99.2502						
Kappa hat classification = 0.9885						

Area unit = m^2, SE = standard error, CI = confidence interval, PA = producer's accuracy, UA = user's accuracy

Table 1-3 The Cape Flats Area Based Error Matrix 2016

The OA of 99.25% and 0.99 kappa hat were achieved in 2016 classification.

	> Reference					
V_Classified	Vegetation	Waterbodies	Built-up	Trees	Area	Weight
Vegetation	0.2980	0.0026	0.0095	0.0017	62953137	0.3117
Waterbodies	0.0000	0.0234	0.0000	0.0000	4746384	0.0235
Built-up	0.0000	0.0001	0.6205	0.0000	125320941	0.6206
Trees	0.0007	0.0004	0.0007	0.0424	8926353	0.0442
Total	0.2987	0.0265	0.6307	0.0441	201946815	
Area	60322804	5358491	127364967	8900553	201946815	
SE	0.0002	0.0001	0.0001	0.0001		
SE area	32264	14880	27240	15075		
95% CI area	63237	29164	53390	29548		
PA [%]	99.7615	88.3499	98.3805	96.1059		
UA [%]	95.5933	99.7438	99.9851	95.8281		
Kappa hat	0.9372	0.9974	0.9996	0.9564		
Overall accuracy [%] = 98.4266						
Kappa hat classification = 0.9693						

Area unit = m^2, SE = standard error, CI = confidence interval, PA = producer's accuracy, UA = user's accuracy

Table 1-4 Cape Flats Area Based Error Matrix 2021

The OA of 98.43% and 0.97 kappa hat were achieved in 2021 classification.

1.4.2 Urban land use and cover change analysis 2016 to 2021

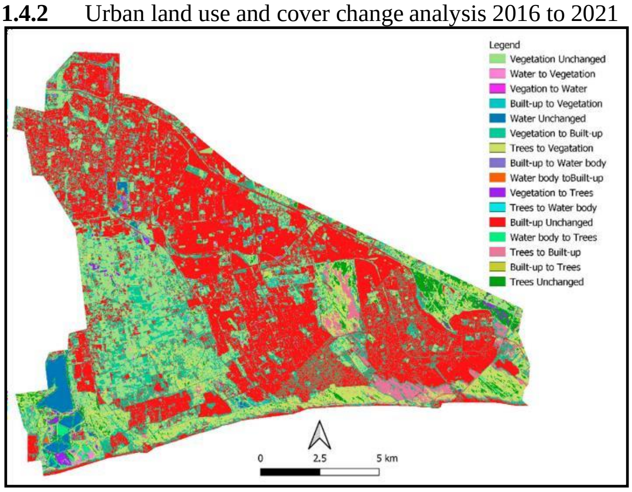

Figure 1-6 Cape Flats spatial representation of urban land use and land cover (LULC) conversion from 2016 to 2021

LULC has changed as indicated in figure 1-6 and table 1-4. Vegetation decreased by 2.1% which translates to an area of 4303.2 km2, waterbodies decreased by 0.4% translating to an area of 862.5 km2, built-up increased by 13.6% translating to an area of 27449.2 km2, and trees decreased by 11% translating to an area of 22283.5 km2. While Figure 6-7 below is a graphical comparison representation of LULC in Cape Flats between 2016 and 2021. It is visible that there is a decrease in vegetation during 2021, waterbodies fairly remain the same between 2016 and 2021, built-up areas increase significantly in 2021, and there is a significant decrease in trees.

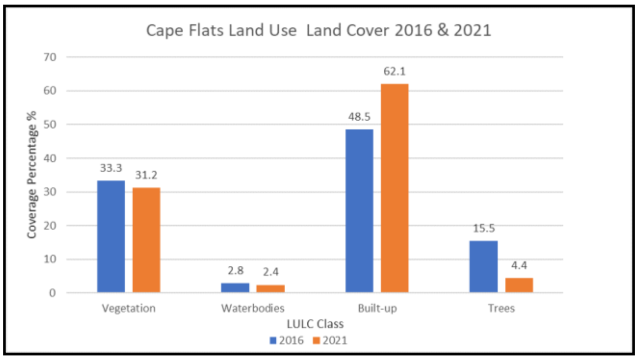

Figure 1-7 urban land use and land cover (LULC) conversion between 2016 and 2021 over the Cape Flats.

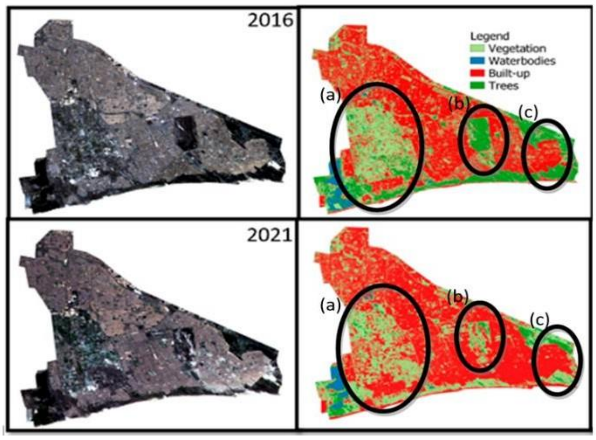

Figure 1-8 Major LULC changes between 2016 to 2021

1.5 Discussion

This study identified rural-urban migration and high birth rate as a major contributing factor behind urbanisation taking place around the Cape Flats. As a result of a high number of people migrating from rural to urban areas and high birth rates, there is a high demand of housing shelter, which currently takes form as unlawful occupation of land for shelter purpose known as informal settlements, and low-cost housing subsidised by government known as Reconstruction and Development Programme (RDP). When rapid urbanisation takes place, it is expected that most vacant land shall be utilized for housing purposes, therefore this informs LULC in the Cape Flats. Lefulebe et al. (2015) pointed out a crucial element that most people involved in rural-urban migration are unemployed and therefore migrate to cities to get jobs, hence an accelerated increase in informal settlements as renting is also expensive but to a lesser degree in comparison to planned suburbs.

Change detection analysis from 2016 and 2021 indicating a five years period is described as follows; the is an evident increase in built-up class from 2016 to 2021 of 13.6% and total LULC being 27449.2 km^2, waterbodies have slightly increased and is statistically ignorable with 0.4%, constituting 862.5 km^2.
Figure 1-8 below on the top right panel indicates the status of LULC in 2016. The major corresponding areas which underwent changes are indicated in the bottom right panel of 2021 with corresponding letters (a, b, c). Circle (a) in 2016 was occupied mainly by vegetation and less tree canopy, however, in 2021 some of the vegetation and trees were lost to built-up classes. While circle (b) in 2016 was covered mainly by trees, in 2021 some trees were also replaced by built-up areas. Lastly, circle (c) in 2021 shows trees that were within the circle in 2016, and were replaced by built-up areas. Constructions of new buildings have been responsible for the majority of LULC changes as seen in figure 1-7. This finding agrees with the following studies (Mahbubur Rahman et al., 2018; Chouchan and Kannan, 2019; Abijith and Saravanan,2021)

This study also showcased how remote sensing can be applied to monitor subtle changes resulting from urbanisation, using PlanetScope high resolution satellite image and machine learning strategies when specifically using a KNN classifier. Results validation from both 2016 and 2021 confusion matrix showed very high reliability of the maps.
Accuracy and validation computation results from the study indicated that the reliability of the maps is high and therefore can be used in different applications not only limiting them to amongst others town planning, Lefulebe *et al*., (2015) also indicated the usefulness of maps in relation to urban infrastructure, location of essential services such as water and toilets etc.

1.6 Conclusions and outlook

This study proved that urbanisation is a real threat to LULC in cities, with 15.8% of vegetation and trees 8%, lost within five years period. Additionally, the population of the Cape Flats significantly increased between 2016 and 2021, owing to rural-urban migration and high birth rate (Mahbubur Rahman *et al*., 2018; Bodo, 2019; Njwambe *et al*., 2019). Remote sensing data (satellite images) and machine learning algorithm showcased the possibility of producing high-accuracy LULC maps that can be used in decision-making. Further, the accuracy of maps produced was higher as compared to traditional cartography and land surveying techniques used in map production, and this renders them cost-effective as they were delivered within a reasonably short period.

The middle-class Cape Flats are characterised by potential future development projects, which include construction of households. It is noted that an increase in urban population is instrumental towards unplanned growth of cities, sustainability is negatively impacted by unplanned urban growth (Sun *et al*., 2013). Negative sustainability factors include; increased surface runoff, reduced air quality, increase in urban temperature, decrease in quantity of life (Merbitz *et al*., 2012; Wang *et al*., 2013).

Future studies can extend the outlined methodology to similar urban problems and in a larger context such as the whole of western cape. It would also be important to understand the impact on LULC in rural areas owing to rural to urban migration.

ACKNOWLEDGEMENTS

We gratefully acknowledge the City of Cape Town for providing aerial photographs and the Planet Labs for PlanetScope Imagery used in this study. Thanks to the Society of South African Geographers for the support. We also thank the anonymous reviewers for the constructive feedback that improved this manuscript.

REFERENCES

Abijith, D. and Saravanan, S., 2021. Assessment of land use and land cover change detection and prediction using remote sensing and CA Markov in the northern coastal districts of Tamil Nadu, India. Environmental Science and Pollution Research, pp.1-13.

Atkeson, C.G., Moore, A.W. and Schaal, S., 1997. Locally weighted learning. Lazy Learning, pp.11-73.

Bhatta, B., 2010. Analysis of urban growth and sprawl from Remote Sensing Data. Advances in Geographic Information Science. https://doi.org/10.1007/978-3-642-05299-6.

Chouhan, B.P. and Kannan, M., 2019. Impacts of urbanisation on land use pattern and environment: A case study of Ajmer City, Rajasthan. Asian Review of Social Sciences, 8(1), pp 87-91

Congalton, R.G. and Green, K., 2019. Assessing the accuracy of remotely sensed data: principles and practices. Third Edition, ISBN: 9780429052729. DOI:10.1201/9780429052729

Di Gregorio, A. and Jansen, L.J.M., 2000. Land cover classification system: Classification concepts and user manual. Food and Agriculture Organization of the United Nations. Rome, 53, pp.179.

Guo, G., Wang, H., Bell, D., Bi, Y. and Greer, K., 2003. KNN model-based approach in classification. OTM Confederated International Conferences. Springer, Berlin, Heidelberg, pp. 986-996.

Han, J., and Kamber, M. (2000). Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco.

Hussain, M., Chen, D., Cheng, A., Wei, H. and Stanley, D., 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. ISPRS Journal of Photogrammetry and Remote Sensing, 80, pp.91-106.

Immitzer, M., Vuolo, F. and Atzberger, C., 2016. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. Remote sensing, 8(3), p.166.

Lefulebe, B., Musungu, K. and Motala, S., 2015. Exploring the potential for geographical knowledge systems in upgrading informal settlements in Cape Town. South African Journal of Geomatics, 4(3), pp.285-298.

Madariya, G. and Sharma, S., 2022. A comparative study on supervised and unsupervised techniques of land use and land cover classification. The Pharma Innovation Journal, 11(3), pp. 517-521

Mahbubur Rahman, M., Hassan, M.S., Md. Bahauddin, K., Khondoker Ratul, A. and Hossain Bhuiyan, M.A., 2018. Exploring the impact of rural–urban migration on urban land use and land cover: a case of Dhaka city, Bangladesh. Migration and Development, 7(2), pp.222-239.

Merbitz, H., Buttstädt, M., Michael, S., Dott, W. and Schneider, C., 2012. GIS-based identification of spatial variables enhancing heat and poor air quality in urban areas. Applied Geography, 33, pp.94-106.

Njwambe, A., Cocks, M. and Vetter, S., 2019. Ekhayeni: Rural–urban migration, belonging and landscapes of home in South Africa. Journal of Southern African Studies, 45(2), pp.413-431.

NOAA. Historical Maps and Charts audio podcast. National Ocean Service website, https://oceanservice.noaa.gov/podcast/july17/nop08-historical-maps-charts.html, accessed on 08/10/2022.

Ochuko, E., 2016. Consequences of rural-urban migration on the source region of Ughievwen Clan Delta State Nigeria. European Journal of Geography ,7(3), pp. 41 – 56.

Owoeye, J.O., Ibitoye, O.A., 2016. Analysis of Akure urban land use change detection from remote imagery perspective. Urban Studies Research, pp. 1-9. doi:10.1155/2016/4673019

Pacheco, A.D.P., Junior, J.A.D.S., Ruiz-Armenteros, A.M. and Henriques, R.F.F., 2021. Assessment of k-nearest neighbor and random forest classifiers for mapping forest fire areas in central Portugal using Landsat-8, Sentinel-2, and Terra imagery. Remote Sensing, 13(7), p.1345.

Planet_Imagery Product Specification [WWW Document]. Available online: https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf (accessed on 19 March 2022).

Rana, S. and Marwasta, D., 2015. Urbanisation trends in developing countries: Comparative study of Yogyakarta City and Kathmandu Valley. Journal of Natural Resources and Development, 5, pp.29-36.

Ruiz, I. and Vargas-Silva, C., 2013. The economics of forced migration. The Journal of Development Studies, 49(6), pp.772-784.

South Africa Info. https://southafrica-info.com/ accessed 20 April 2022

Sun, C., Wu, Z.F., Lv, Z.Q., Yao, N. and Wei, J.B., 2013. Quantifying different types of urban growth and the change dynamic in Guangzhou using multi-temporal remote sensing data. International Journal of Applied Earth Observation and Geoinformation, 21, pp. 409-417.

Théau, J.,2008. Change detection. Encyclopedia of GIS, pp77-84

Van Weele, G.; Maree, G. 2013. State of environment outlook report for the Western Cape Province: Introductory matter; Western Cape Government Environmental Affairs Development Planning: Cape Town, South Africa pp 49.

Wang, S.W., Gebru, B.M., Lamchin, M., Kayastha, R.B. and Lee, W.K., 2020. Land use and land cover change detection and prediction in the Kathmandu district of Nepal using remote sensing and GIS. Sustainability, 12(9), p.3925.

www.orfeo-toolbox.org, accessed 10 March 2022