Utilization of mid-resolution imagery and remote sensing techniques to characterize urban growth over the Philippi Horticultural Area

by

Mr Darron Isaacs

Thesis submitted in fulfilment of the requirements of the degree:

MTech: Cartography

Department of Civil Engineering and Geomatics
in the Faculty of Engineering and the Built Environment

at the Cape Peninsula University of Technology

Supervisor: Mr. K. Musungu (CPUT)
Co-supervisor: Co-supervisor: Dr P Kumar (CPUT)

Bellville

February 2023

Declaration

I declare that this research dissertation is my own unaided work. It is being submitted for the BTech Degree at Cape Peninsula University of Technology, Cape Town. It has not been submitted before for any degree or examination in any other University.

_______Darron Leslie Jack Isaacs____
(Signature)

Signed in Cape Town this 27th___ day of __February____ 2023

Abstract

The Philippi Horticultural Area (PHA) is a peri-urban area with a long history of food production dating back to the mid-1800s. The total area of the PHA comprises over 3000 hectares, of which 1200 hectares are suitable for food production. However, farming within the PHA has been affected by increased development in the area. Thus, the need for a spatial understanding of urban growth in the area is imperative. This study aims to utilize satellite imagery combined with remote sensing to identify vegetation and urban indices and to analyse land use through change detection mapping. Part of the process was to determine the change detection through the classification of satellite images and the insertion of vegetation indices in order to produce accurate land cover changes on a map. Limitations to the study were restriction to sensors being used, the years available of images for the study area. Certain population statistics were not available for the years in question and only three classifiers could be used. The study was done between the years 2015 and 2021.

This study used image classification and PlanetScope digital images data which were essential in the application that was used in this study, namely machine learning (ML). The need to provide analysis and decision-making for the PHA has been a challenge, but currently, remote sensing has globally been applied with the use of current advanced satellite systems and sensors. In terms of classifying satellite imagery, three machine learning techniques have been applied to this research, namely Random Forest (RF), K Nearest Neighbour (K-NN), and Support Vector Machines (SVM). The best-performing classifier was used to classify the images into six classes namely urban fabric, water, vegetation (agricultural and natural), and bare ground (sand and bare ground). Change detection was done on images of consecutive years by displaying differences over time and then by mapping the trajectory of urban growth.

The main findings in this study was the growth in urban fabric. Urban fabric started at 30% during 2015 and 2016, it increased by 2% during 2016 and 2017, and a further 3%% between 2017 and 2018, there was a slight decrease by 1% between 2018 and 2019, it increased slightly by 1% in 2019 and 2020 and further increased by another 1% between 2020 and 2021. In contrast, farming area started off at 10% for 2015 and 2016, it increased by 9% during 2016 and 2017, reduced by 9% again during 2017 and 2018, there was another increase by 7% between 2018 and 2019, a slight decrease of 2% between 2019 and 2020 and further decreased by a 3% between 2020 and 2021.

The vegetation indices resulted in the following being found with the overall classification accuracy showing improvement with the inclusion of indices for each year. In 2015 the accuracy was found to increase by .04%, in 2016 the accuracy improved by 6.6%, in 2017 the accuracy improved by 27.5%, in 2018 the accuracy increased by 10.8%, in 2019 the accuracy increased by 1.6%, in 2020 the accuracy improved by 18.6% and in 2021 the accuracy increased by 7.8%.

The accuracy results for two of the three classifiers—Random Forest (RF) and Support Machine Vector—were comparable. Despite the great accuracy of the two classifiers, Random Forest (RF) consistently outperformed Support Vector Machines. The findings indicate that high accuracy classifications for mapping the agricultural and urban edge are possible using both the SVM and RF classifiers. It is clear that the study region influences how well the classifiers function.

Acknowledgements

I would like to thank the Cape Peninsula University of Technology;

I would like to thank my supervisors Mr Kevin Musungu and Assoc. Prof. Pallav Kumar.

I would like to thank The Department of Environmental Affairs and Development Planning (Western Cape Government) and the Geomatics Development Fund for sponsorship.

I would also like to thank God for giving me the strength and ability to conduct this research.

Table of Contents

Page

List of Figures

Page

Body

Appendices

List of Tables

Page

Body

Chapter 1 Introduction

In this chapter the Philippi Horticultural Area will be introduced as well as the research problem, question and as well as the objectives and outcomes.

1.1 Background and Motivation

Philippi is one of the biggest suburbs situated within the Cape Metropolitan area in the Western Cape province of South Africa. It forms a part of the city municipal area and contains many farms which are sparsely populated compared to the remainder of the town. The area of Philippi is additionally known to grow about 80% of Cape Town's vegetables within the farmland and is slated for several large developments including the 'Philippi Mini-City' (Indego, 2018). However, a changing urban edge and rezoning of land from agricultural to mixed use in the Philippi Horticultural Area (PHA) has resulted in less land being used for agricultural activity. This has resulted in illegal dumping, conflicting land use, winter flooding, and safety and security concerns for the remaining farmland, which is under severe threat. Natural vegetation is being replaced with impermeable concrete, and there is encroachment of commercial, residential and informal land use which reduces the available recharge area and increases the danger of aquifer contamination.

Figure 1.1 Map of study area within the Western Cape. (Source: GeoEye)

1.1.1 Details of the research area

The Philippi Horticultural Area (PHA), a peri-urban area with a long history of food production going back to the middle of the nineteenth century, is situated in the Cape Flats area of the City of Cape Town region. The Khoi and San were formerly a nomadic people who used the region to graze livestock and hunt for food. In 1833, while it was still known as 'Die Duine' (The Dunes), the first public records of local citizens were made. When German settlers arrived at Philippi in Cape Town in three waves between 1860 and 1883, they quickly earned a reputation for being able to cultivate vegetables in the sand-filled Cape Flats. The region was mostly utilized for grazing until the 1970s, when there were a few farms in the area. In the neighbouring areas of Langa, Gugulethu, Browns Farm, Samora Machel, and Crossroads, economic migrants from the former Ciskei and Transkei Bantustans in the Eastern Cape began settling when apartheid laws were implemented in the late 1970s and early 1980s. The 1980s saw a rise in opposition to apartheid, and Philippi increasingly became a haven from the political upheaval in the areas mentioned above (Setplan 2017).

At least half of Cape Town's vegetables are produced here, despite the city being encircled by residential suburbs that include townships, unincorporated areas, and historic neighbourhoods. Over 3 000 hectares of land are included in the region, of which 1 200 hectares are ideal for food production. More than 50 different horticultural crops are grown by smallholder and commercial farmers in the PHA. Nearly 100 000 tonnes of fresh vegetables are reportedly cultivated yearly in the PHA, with a significant percentage of that product finding it into Cape Town's food chain. It is a component of an interconnected network of economic systems that benefit each other and the region as a whole, adding to their mutual economic integrity and sustaining about 4 000 jobs inside the PHA (Spain, 1984). In the time of apartheid, it was reserved for coloured people, who made up 70.5% of the population, with white people and black Africans making up the remaining population. The remaining core of the PHA consists of about 1 884 hectares of agricultural land inside the Cape Flats District of the City of Cape Town (CoCT). With a total size of 3 168.65 ha, the larger PHA region has a wide variety of formal and informal land uses, including industrial and residential, which forms so-called buffer areas around the core PHA. There are nine informal communities in the larger PHA region, just one of which is situated in the core. The horticulture is unique and sites like sand mining, and mining of silica sand should be preserved. The main PHA footprint has, however, been reduced as a result of a series of planning choices made after 1988 (Indego, 2018).

1.2 Research problem

The Philippi Horticultural Area was impacted by the area's rising urbanization. Urbanization in the Western Cape was accompanied by an increase in urban sprawl in all socioeconomic classes. Urban expansion into the periphery of cities and towns typically involves unplanned, low-density growth. Urban sprawl in the Western Cape was partially a result of poor spatial planning (within the historical context) and land use decision-making, which were both factors in the rapid population increase into urban areas, with smaller household sizes leading to more households. The upshot of urban spread has been a less than optimal use of resources, particularly where this sprawl engulfed productive agricultural area on the city's borders. The need for housing and urban growth for the community is a need that is currently being viewed by the City of Cape Town. Housing is definitely a challenge in certain parts of the province, but protection of food sources also needs to be monitored as it plays a big role in supplying vegetation for the city.

Therefore, a further spatial understanding of urban expansion in the area is urgently needed. The goal of the study is to determine how remote sensing may be utilized to obtain more ways to monitor the growth of the urban areas within the PHA.

1.3 Research Question

The proposed research questions are as follows:

1) How can remote sensing technologies be used to map urban growth in the Philippi Horticultural Area?
2) What is the trend in urban growth in the Philippi Horticultural Area?

1.4 Objectives and outcomes

The aim of this work was to map the urban growth within the Philippi Horticultural Area using remote sensing. To achieve this, the following objectives were met:

1) To identify appropriate indices to facilitate the mapping of urban growth.
2) To classify land cover in the study area over different periods.
3) To detect changes in the PHA using remote sensing.

The outcomes were as follows:

1) The identification of key vegetation and urban indices.
2) The creation of multitemporal land cover maps.
3) Change detection maps over a seven-year period.

1.5 Delineation

The limitations of the study are:

1) The data used in this study was limited to the period 2015 to 2021.
2) It is constrained to the use of PlanetScope satellites.
3) Only three machine learning algorithms, namely Random Forest, K-Nearest Neighbour and Support Machine Vector were used in this study.
4) No up to date third party statistical information data from STATS SA was used in this study.

1.6 Assumptions

The assumption was that urban growth has sprawled over the PHA area and it has affected the agricultural sector.

1.7 Methodology

The process that will be followed for this study is to download mid-resolution satellite imagery and compute it through a GIS/remote sensing software package. This software will process various images and classify them accordingly. Pre-processed multispectral data was downloaded from PlanetScope for the dates April 2015, 2016, 2017, 2018, 2019, 2020 and 2021. To further improve the accuracy levels 18 vegetation indices were used and processed in order to improve the accuracy levels of the classifications these number of indices were created from the data based on previous literature. Six classes of samples were created for the following namely, urban fabric, water, vegetation, natural vegetation, bare ground and bare ground – sand. In addition, the Boruta feature selection was used to reduce the number of features that are statistically least relevant. The Dzetsaka Plugin in QuantumGIS software was utilised to perform pixel-based classifications using random forest, support vector and k-nearest neighbour classifiers. Post classification accuracy assessment was done in the Semi-automatic plugin. It was repeated for each classification between 2015 and 2021. Change detection was done by differencing consecutive classifications using the Semi-automatic plug-in tool. once this has been done the land cover tool was used to create the change detection maps for the various years. These additions provided an accurate assessment which helped improved the change detection maps to provide a visual representation of the land cover change over the area of interest. See Figure 2.1 for the summary of the methodology. A graphical summary of the methodology is presented below.

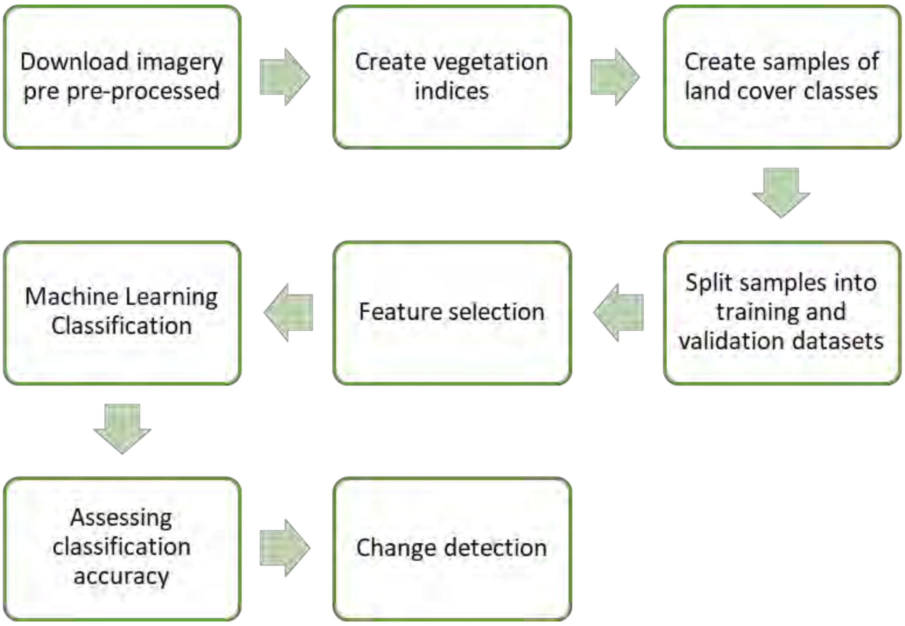


Figure 1.2 Flow chart of the methodology process take for this research.

1.8 Analytical studies

An analysis of mid-resolution satellite imagery was done over the Philippi Horticultural area to determine change in urban growth over a 7-year period. The results determined the amount of growth that occurred and the visual changes over the given period using Planet scope data.

The research methods utilised resulted in the data being collected and the spatial data from various custodians related to the focus area. This data was processed through a GIS desktop application called Quantum GIS running on the Windows 10 operating system. The satellite data collected was processed through this application and analysed using a change detection method.

Each image will be processed and analyzed through the software. This analysis will provide a solid foundation with the data that will be used. This type of study will speak to how raster data can be transformed to be a mechanism that can produce results and to provide analytic assessment to the outcome.

1.9 Organisation of dissertation

Chapter 1 introduces the study area. Chapter 2 discusses the literature review and theory including urbanization, classification and change detection methods used previously. Chapter 3 describes the study area and the research methodology. Chapter 4 presents the results of the research. Chapter 5 discusses the findings of the research. Chapter 6 presents the conclusion and recommendations in relation to the findings.

Chapter 2 Literature review and theory

2.1 Introduction

This chapter reports on various studies that have utilised remote sensing in mapping urbanization.

2.2 Urbanization

Urbanization (or urbanisation) refers to the changes in the size, density, and land use of cities. Urbanization encompasses several factors such as industrialization, population growth, population mobility and population density as well as the growth or decrease of cities (Vlahov and Galea, 2002).

Between the 1950s and 2014, the percentage of people living in urban areas worldwide rose from 30% to 54%. The worldwide urban population is expected to increase by 2.5 billion people over the next few decades, mostly in Asia and Africa. Urban land cover could increase by 1.2 million km^2 by 2030, roughly tripling the global urban land area since 2000 (Seto et al., 2012). The rate of expansion within urban areas requires an intense degree of environmental sustainability to maintain the areas and communities living in them (Fernandes, 2002; Weber and Puissant, 2003; Sudmeier-Rieux et al., 2015; Li et al., 2019).

Several scholars from across the world have identified the negative impact that urbanisation of agricultural areas has on food security (Deng et al., 2006; Gomes et al., 2019; Zhou et al., 2019). In developing countries, the scale and speed of urbanization can result in unsustainable settlements that create pressure on ecosystems (Pham et al., 2015). The direct pressures caused by urbanisation are usually localised around cities themselves (Bugnot et al., 2019; Liu et al., 2019). A direct loss of land is created through construction of buildings and infrastructure which results in the loss of the ecosystem services that facilitate food production (Cui et al., 2019; García-Nieto et al., 2018; Wang et al., 2019; Wen et al., 2019). Studies have shown that urbanization in agricultural areas can lead to a decline in both animal breeding and the cultivation of crops (Pham, Kappas & Faust, 2021). Some studies have also shown that more land elsewhere may be required than the urbanised agricultural land to yield the same amount of harvest because of differences in soil properties and climate (Andrade et al., 2022). Thus, it is important to monitor changes to land cover in areas that are traditionally used for agriculture (Guan et al., 2019). The next section highlights the use of earth observation technologies for monitoring land cover.

2.3 Land use and land cover mapping using remote sensing

Many vegetation monitoring research and applications, such as wetland mapping, plant health and species mapping, and even mapping urban forests, have made use of remote sensing (Kaplan et al., 2018; Sharifi et al., 2020; Singh et al., 2020; Stubbings et al., 2019; Sanesi, et al., 2019; Lefulebe, Van der Walt & Xulu, 2022). Remote sensing offers the instruments needed for long-term monitoring of the transition between farmland and urbanized regions. It is important to define land use and land cover:

‘Land use is defined as a series of operations on land, carried out by humans, with the intention to obtain products and/or benefits through using land resources’ (Coffey et al., 2013).

‘Land cover is defined as the vegetation (natural or planted) or man-made constructions (buildings, etc.) which occur on the earth surface. Water, ice, bare rock, sand, and similar surfaces also count as land cover’ (Coffey et al., 2013).

The increasing availability of open image processing software and open data from earth observation (EO) satellites combined with toolboxes have popularised land use and land cover (LULC)-focussed research. Several machine learning algorithms have been studied and implemented to classify different features on the ground (Lefulebe, Van der Walt & Xulu, 2022). There is a growing need to monitor changes in intra-urban structures due to urbanization in cities of developing countries. Remote sensing has proven to be a versatile tool in achieving this, although the spatial resolution has a major influence on the efficiency of the mapping process (Kuffer et al., 2011). Medium resolution satellite dataset products such as Sentinel and Landsat have frequently been utilized for regional level analysis but the advent of constellations of CubeSats such as PlantScope have extended the capacity of remote sensing-based studies.

In recent years, machine learning has also improved the capacity for remote sensing datasets to map intra-urban dynamics such as population density, immigration, infrastructure and business development (Wieland & Pittore, 2014; Magidi & Ahmed, 2019). For instance, studies have focussed on mapping informal settlements (Kohli, 2012; Hacker et al., 2013), land surface temperature (Guha et al., 2019) and urban land use change (Wu et al., 2021).

2.4 Vegetation Indices

Vegetation Indices (VIs) have been used for crop type identification and agriculture land cover classification. Spectral information of the original bands and vegetation indices can provide additional information for detailed analysis which can both be utilized for crop classification. It is possible to detect different cultivating patterns since multitemporal vegetation indices enhance analysis of crop growing pattern and different crop classes. Vegetation indices have also been widely used for crop monitoring studies because they can serve as powerful and simple indicators of stress, biophysical attributes, and crop maturity.

Many studies have proven to increase accuracy levels in various scenarios by adding vegetation indices (reference?). Vegetation indices improve the spectral information and raise the separability of the classes of interest, which in turn impact on classification accuracy. Vegetation indices are used to differentiate between each land use and land cover, reduce intra-class variation, and increase inter-class variability. In addition, some crops' spectral features are enhanced while others are suppressed by the various combinations of vegetation indices (Kuzucu & Balcik 2017).

The vegetation indices will assist in improving the accuracy levels of the classifications of the images.

Below is the table displaying all the indices used within the study:

Table 2.1 Vegetation Indices (VIs) that were derived from multispectral images in this study.

Vegetation Indices	Formula	References
Enhanced Vegetation Index 2 (EV12)	2.5*(N-R)/(1+N+(2.4*R)	Huete et al. (1999)
Normalized Difference Vegetation Index (NDVI)	(N-R)/(N+R)	Kriegler et al. (1969)
Green Normalized Difference Vegetation Index (GNDVI)	(N-G)/(N+G)	Gitelson et al. (1996)
Red Green Vegetation Index (RG)	R-(G-R)	Coops et al. (2006)
Green Ratio Vegetation Index (GRVI)	N/G	Sripada et al. (2006)
Transformed NDVI (TNDVI)	sqrt[0.5+(N-R)/(N+R)]	Bannari et al. (2002)
Normalized Green Red Difference Index (NGRDI)	(G-R)/(G+R)	Bannari et al. (1995)
Normalized Difference Water Index (NDWI)	(G-N)/(G+N)	Gao (1996)
Modified Simple Ratio (MSR)	(N/R-1)/ sqrt(N/R+1)	Chen (1996)
Green Soil Adjusted Vegetation Index (GSAVI)	1.5*(N-G)/(N+G+0.5)	Broge and Leblanc (2001)
Log Red (LogR)	logR	Vogelmann et al., (1993)
Ratio Vegetation Index (RVI)	R/NIR	Colwell et al. (1974)
Vegetation Index Number (VIN)	NIR/R	Gao (1996)
Transformed Vegetation Index (TVI)	Sqrt [NDVI + 0.5]	Perry and Lautenschlager (1984)
Differenced Vegetation Index (DVI)	(NIR-R)	Clevers (1986)
Normalized Difference Green Index (NDGI)	(G-R)/(G+R)	Rouse et al., (1974)
Redness Index (RI)	(R-G)/(R+G)	Lyon et al., (1998)
Green Chlorophyll Index (GCI)	(NIR/G)-1	Huete et al., (1994)
Modified Soil Adjustment Vegetation Index (MSAVI)	2 * NIR + 1 – sqrt ((2 * NIR + 1)2 – 8 * (NIR - R))) / 2	Huete et al., (1994)
Renormalized Difference Vegetation Index (RDVI)	(R-R)/(R+R)0.5	Chen (2006)

In urban area studies, remote sensing based on mid-resolution multi-spectral data has proven to be an effective tool. Previous papers have presented methods for mapping urban developments in terms of impervious surface expansion based on spectral index ratios. T. Accurate information of urban land-cover is crucial for urban management since it has a direct impact on runoff prevention, urban vegetation planning, monitoring and enhancing air quality, and even mitigating the consequences of climate change. In order to better understand urban area ecosystems and thereby improve the environment and human life quality in urbanized areas, reliable urban land-cover mapping combined with vegetation indices can provide crucial information with very accurate results (Villa, 2012).

2.5 Feature selection

As a dimensionality reduction strategy, feature selection seeks to pick a small subset of the pertinent characteristics from the initial ones by omitting features that are unnecessary and redundant. Feature selection eliminates redundant and irrelevant data. It typically results in greater learning accuracy, lower cost of computing performance and easier model interpretation. Generally speaking, insignificant qualities that cannot be utilized to distinguish between numerous classes being monitored are grouped together (unsupervised). Since unnecessary features throw the learning system off balance, impair memory, and result in inefficient computation, deleting them may actually help develop a better model. This can speed up computation, increase learning accuracy, and provide a deeper comprehension of the learning model or data. Wrapper-based and filter-based approaches are the two main categories of feature selection techniques (Rodrigues et al., 2014). While filter-based approaches evaluate the value of individual features to pick features through ordering, wrapper-based approaches use the classifier's performance as an assessment criterion for optimizing the feature subset. Recent studies focused on filter-based techniques but used feature selection in object-based classification (Ma et al., 2015). Certain studies utilized evolutionary algorithms to implement feature selection for object-based classification used by van Coilie, Verbeke and De Wulf (2007), whereas Laliberte, Browning and Rango (2012) used the GINI index as the splitting criteria to rank object features.

One of the popular methods for feature selection is Boruta (Kursa and Rudnicki, 2010). In Boruta, the first step involves creating a randomized version of the features to be assessed called shadow features. The original features are only determined to be important if they improve the classification better than the best shadow feature. Szul, Sylwester and Krzysztof (2021) uased to select data for the prediction of energy consumption for building heating. Agjee et al. (2016) used Boruta and Recursive Feature Elimination to improve the classification accuracy when mapping the effect of weevils as a water hyacinth biocontrol measure. Arjasakusuma, Swahyu, Kusuma and Phinn (2020) used Boruta and other feature selection algorithms to select the most important features for estimating forest height using hyperspectral data, airborne light detection and lidar height metrics as separate features and when combined. It is evident that Boruta is a viable method for feature selection in remote sensing data.

This process assists in focusing on the selected features and eliminates redundancy. This method will prove the assurance that the outcomes speak directly to the intended classes.

2.6 Image classification

Image classification of remote sensing images is the process of grouping all pixels in a satellite image or aerial image to extract land cover features (Andrade et al., 2022; Wenjing and Wang 2020).

It is evident from prevailing literature that scholars have employed a variety of classification techniques depending on the scene being classified. These techniques can be either pixel-based or object-based (Franklin & Ahmed, 2017; Pande-Chhetri, et al., 2017; Puliti, Talbot, & Astrup, 2018). The pixel-based approaches can further be categorised as supervised or unsupervised, although some studies have used a hybrid approach (Chang et al., 2020; Goldblatt et al., 2018; Louargant et al., 2018). Due to the availability of high-quality data, temporal frequency, and broad coverage compared to the conventional data collection methodologies, the use of remote sensing in several agro-environmental applications is attractive (Li et al., 2020).

Supervised classification usually starts with the creation of training samples. Samples are based on either ground-truth knowledge or visual image interpretation, which has been identified and defined by an expert. They usually consist of a point-location with an associated class label corresponding to the feature of interest. A training dataset can be used to train a classifier on any kind of input feature space and image type and is defined as a basic structure (Samaniego., L & Schulz., K. 2009).

The state-of-the-art in machine learning is represented through classification algorithms and is based on different concepts such as tree-based, nearest neighbor and function-based. The following section describes three commonly used classifiers. Images need to be classified in order for a change detection process to take place. These classifiers will be introduced as the first phase to getting the process started. This will assist in other tools to be introduced as well as provide an accurate outcome for this study. A more detailed discussion of classification can be found in Aldoski et al. (2013).

2.6.1 K Nearest Neighbour (KNN)

The k-nearest neighbour algorithm (KNN) is a supervised learning classification process that utilises proximity to a known point to predict or classify the class of each individual data point that needs to be identified. KNN's partial mandate is to establish locations that can be identified near to each other. The k training samples nearest to the element in the resource space are made known through the use of KNN algorithms. These datasets act as the main KNN algorithm adjustment parameters and also have a strong impact on spatial prediction. The KNN classification algorithm is frequently utilized in remote sensing with abundant data applications (Aldoski et al., 2013).

KNN is a non-parametric machine learning algorithm (MLA) that does not assume anything regarding the main data set. In order to classify a pixel within a KNN is determined by its known class ID (da Penha Pacheco et al., 2021). These classification systems become questionable when common statistical assumptions are present, such as those that underlie linear regression models, which are also used along with nonlinear, heterogeneous, and noisy data. The similarity indices – which are used by KNN to recognize new cases, and which are from the contributing instances in the dataset – are also often referred to as distance functions. Cases are categorized when utilizing neighbour class voting and these usually create the best-case scenario with the highest similarity indexes. The ideal number of neighbours (K) are established by the metrics used for regression and classification. For ongoing variables, Euclidean distance is the most well-known distance metric. For discrete variables, however, the overlap metric also known as the Hamming distance is frequently used. Additionally, used as metrics are correlation coefficients like the Pearson and Spearman correlation, like correlation coefficients are additions utilised as metrics (De Winter et al., 2016). A variable K value can be used depending on the dataset which was selected. An introduced empirical rule of thumb which makes parameter adjustment problematic for various applications is that K is equal to the square root of the number of samples (Shahabi et al., 2020).

Samaniego and Schulz (2009) used KNN to classify agricultural land cover at test sites in the Parthe-catchment in Leipzig, Germany. The study used Landsat 5 and Landsat 7 and reported overall classification accuracies of between 70% and 98% with higher accuracies noted for multitemporal datasets. Sun and Huang (2010) used KNN in an adaptive algorithm to overcome the limitation of the traditional k-nearest neighbour algorithm (KNN) which usually identifies the same number of nearest neighbours for each test example, while Dudani (1967) used KNN rule to weight distance function for the purpose of assigning a class to an unclassified sample.

The figure below indicates how classes are grouped together to make a decision:

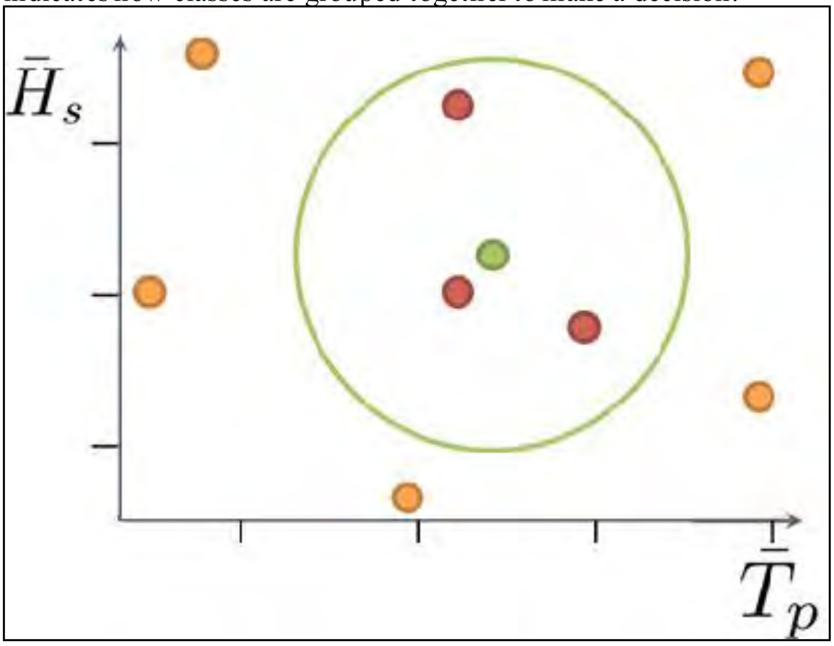


Figure 2.1 K-Nearest Neighbour examples. (Source: Chailan 2015)

In Figure 2.1 is shown how KNN utilises proximity to a known point to predict the same colour circles to represent the point to be classified. The red points are the closest 3 points that will be used to determine its class.

2.6.2 Random Forest (RF)

Random forest is a set of decision trees initially displayed by Leo Breiman and Adele Cutler (2012), also known as random trees. RF uses a string of simple decisions to assign class labels based on the results of sequential tests. (Berhane et al., 2018). The leaves, which are composed of a set of decision sequences where tests are applied at the nodes of the trees, represent the class labels of the branches of a decision tree. In random forest, the input feature vector is classified with every tree in the forest where the last forecast is based on a greatest voting system. The trees are trained with the same parameters with different sets of training instances. For each training set, the same number of vectors as in the original set are randomly selected with replacements and are determined by using a bootstrap procedure on the original training dataset (Breiman 2001). A random subset of the variables is used at each node of the trained trees in order to find the best split for all the nodes and trees by a training parameter. The size of subsets generated at each node is fixed and none of the trees that are built are pruned. The out of bag error for each tree is estimated from vectors that were discarded during the training phase of the individual tree classifiers by sampling with replacement.

For parameters, the only requirements by Random Forest are number of trees which are required to produce a complete forest which then consists of the number of randomly selected predictor variables, the quantity of predictor variables to be chosen at random. In general, out-of-bag errors decrease as the number of trees increase. When the number of trees is greater than a predetermined threshold, the error is convergent according to the Law of Large Numbers (Da Penha Pacheco et al, 2021). It is always of great value to determine out-of-bag errors against the number of trees, despite the forest having sufficient tress or not. If there is low number of randomly selected predictor variables it will display a weak prediction for the future, but there is little correlation between them, which can lessen the generalization error. In Random Forest, the number of randomly selected predictor variables can be determined through the utilization of the square root of the input variable calculation or a third of the count. Random Forest is a cost-effective method to compute where there is an indifference to the settings utilized to determine it and its outliers (Georganos et al., 2021). Furthermore, choosing the right parameters is simple compared to individual decision trees, over-fitting is less of a problem, and there is no need to prune the trees, which is a laborious task (Feng et al., 2015).

Lefulebe, Van der Walt & Xulu (2022) utilised random forest to classify and detect LULC changes as part of a study to map urban forests over the City of Cape Town between 2016 and 2021. The study reported an overall accuracy of 94.8% and a kappa accuracy of 0.92. An assessment of the effectiveness of a random forest classifier for land cover classification was done (Rodriguez-Galiano et al., 2012). Another study was done for Prediction with Confidence Based on a Random Forest Classifier by Devetyarov and Nouretdinov (2010) to represent formal predictor for a new flexible framework that outputs region predictions with a guaranteed error rate

The Figure below shows how Random Forest filters down to it's decision:

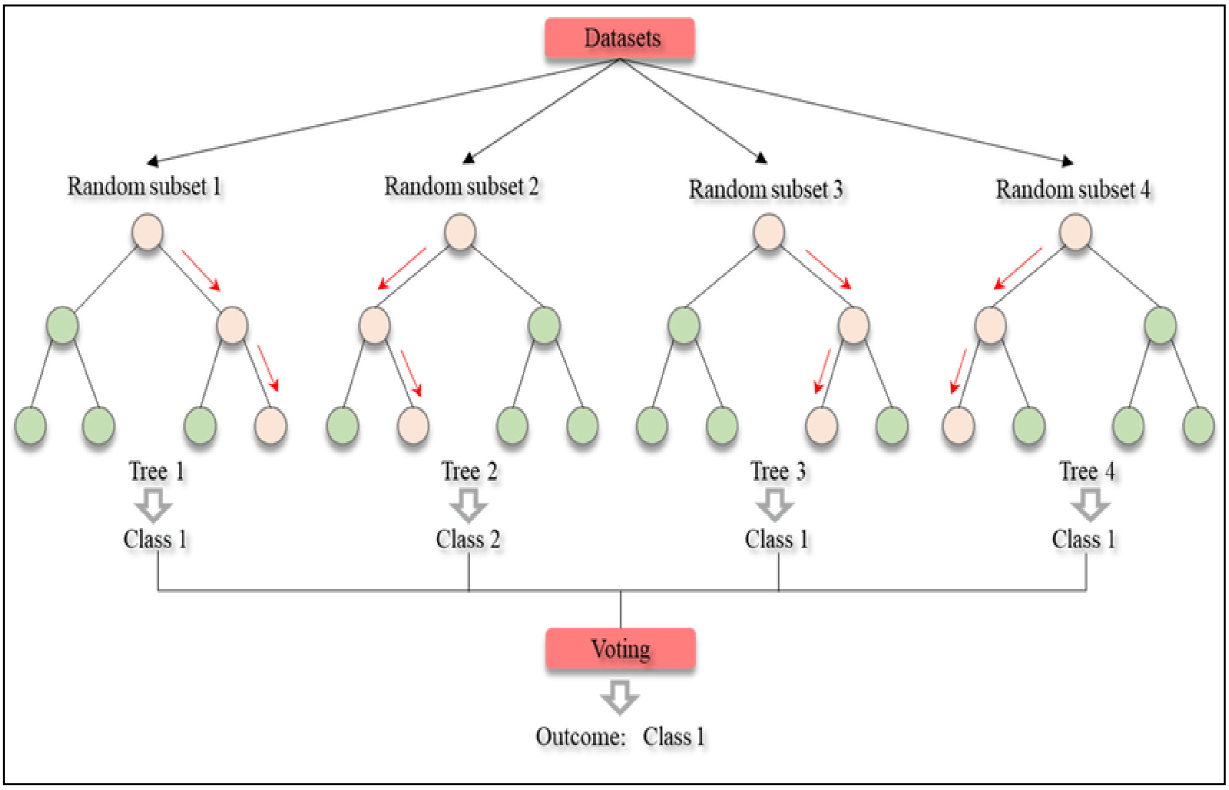


Figure 2.2 Random Forest examples. (Source: Shih et al., 2019)

In Figure 2.2 uses a string of simple decisions to assign class labels based on the results of sequential tests. The class labels of the branches of a decision tree are represented by the leaves, which are made up of a collection of decision sequences where tests are used at the tree's nodes. In random forest, each tree in the forest classifies the input feature vector, and the most recent forecast is based on a largest voting mechanism.

2.6.3 Support Vector Machine (SVM)

The support vector machine (SVM) was initially developed by Vladimir Vapnik and is a classifier originating from statistical learning theory. SVM separates any two classes of interest by identifying an optimal linear separating hyperplane (Figure 2.3). A kernel function is utilised to forecast non-linearly separable classes from the initial feature gap to an elevated dimensional space and this is where the non-linearly dividing classes cannot be distinguished by a linear hyperplane (Lee et al., 2012). SVM takes into consideration the maximum margin concept, which fully extends the gap between the dividing hyperplane and the nearest feature vectors to select the optimal separating hyperplane between two classes (Cervantes et al., 2020). When the location of other feature vectors does not compromise the hyperplane, it means that these feature vectors are called ‘support vectors’. This particular hyperplane maximizes the ability of the SVM in order to envision the correct class of previously unseen selected samples. Furthermore, a gentle margin buffer which allows some data points to infiltrate the separation through the hyperplane without disturbing the end result is presented in SVM with outliers in the data. Therefore, the gentle margin parameter determines an exchange between the hyperplane and the size of the margin (de Castro Paes et al., 2022). It is important to consider for the normalization of the feature variables by their covariance or average variance as the SVM classifier is reliant on a distance measure. All-in-one classifiers are a widely used and simplified generalization method to train multiple samples and to apply SVM to multi-class problems. According to a standard ten-fold cross-validation method during the training phase of the classifier, optimal SVM parameters are selected. Decreasing the risk of over-fitting is used less often than a k-fold cross-validation (Jie et al., 2015).

Fan et al., (2018) has done a comparison study between SVM and Extreme Gradient Boosting in order to predict daily global solar radiation by means of temperature and precipitation in humid subtropical climates. A review on landslide susceptibility mapping using SVM and compared to four other methods which was done by Huang and Zhao (2018), and the results would determine that SVM is the best classifier out of the rest. Another study was done to apply SVM machine learning models for forecasting solar and wind energy resources and using the model for fast and accurate results (Zendehboudi et al., 2018) which would prove a viable model to use. Figure 2.3 below shows the relationship between classes and the hyperplane.

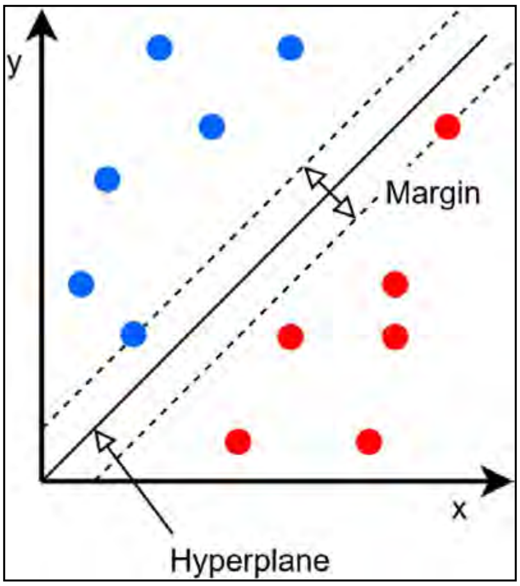


Figure 2.3 Support Vector Machine (Source: Bhandari et al., 2021)

The illustration above demonstrates how SVM may discover the best linear separating hyperplane between any two classes of interest. From the initial feature gap to an elevated dimensional space, where the non-linearly splitting classes cannot be separated by a linear hyperplane, non-linearly separable classes are predicted using a kernel function. In order to choose the best separating hyperplane between two classes, SVM takes into account the maximum margin concept, which fully expands the distance between the dividing hyperplane and the closest feature vectors.

2.7 Change Detection

The surface of the earth changes continuously due to the natural phenomena or human activities. The process of identifying the changes which have occurred over time on the earth surface is called change detection. Change detection of earth's surface is carried out effectively in the field of remote sensing using various techniques (Mishra et al., 2017). Common change detection methods include change vector analysis (Basak and Haque, 2017); principal components analysis (Zhu, 2017); and comparison of land cover classifications (Viana et al., 2019). Further techniques include the use of images in a deeply supervised image differencing network for change detection in high resolution bi-temporal remote sensing images by Zang et al. (2020); in land use land cover change detection, and monitoring of urban growth using remote sensing and GIS techniques (Das and Angadi , 2021); and vegetation index differencing (Wu et al.,2017). Change detection has also been done by combining automatic processing and visual interpretation (Mas et al., 2017).

There are numerous practical applications that utilise information derived from change detection in land-cover (LC) and land-use (LU) (also referred to as LULC) (Asokan et al., 2019). These include city planning, management of land resources, damage assessment, disaster monitoring as well as deforestation. It is necessary for the LULC data to be updated, given the fact that the LULC constitutes an important source of vital data for decision making. The framework of change detection, which makes use of a multi-temporal dataset, involves qualitative examination of the temporal repercussions of an event as well as quantifying the changes. There has been a lot of interest in research on LULC change detection. Change detection is an active topic of study with new approaches being created for improved detection outcomes (Ardila et al., 2012; Chen et al., 2013; Kim et al., 2013), many of which have previously been developed over the past forty years (Demir et al., 2012; Volpi et al., 2013).

Change detection mapping will provide the visual changes over the PHA and determine which feature class has grown the most against the agriculture sector of the area. CD does not only show visual changes but also statistical changes. e.g. acreage of changes, timeline of changes or which years had the most change.

2.8 Summary

This study utilised several recommended algorithms based on the preceding studies. These included the use of vegetation indices, feature selection and machine learning. The use of higher resolution imagery was used to determine change detection over a period of 7 years within the Philippi Horticultural Area. A previous study done by Musungu and Mkhize (2019) utilised images from the years 1990 to 2016 using maximum likehood algorithm, minimum distance and spectral mapping to classify images, whereas this study aims to use 7 years from 2015 to 2021, is classified using RF, KNN and SVM classifiers, as well as machine learning and feature selection. The results of this study shows an increase in urban fabric and an eventual decrease towards the final year for vegetation.

Chapter 3 Research methodology

This section introduces the study area as well as the methodology used in this study. The PHA is located Southeast of the City of Cape Town (Figure 3.1).

3.1 Study area

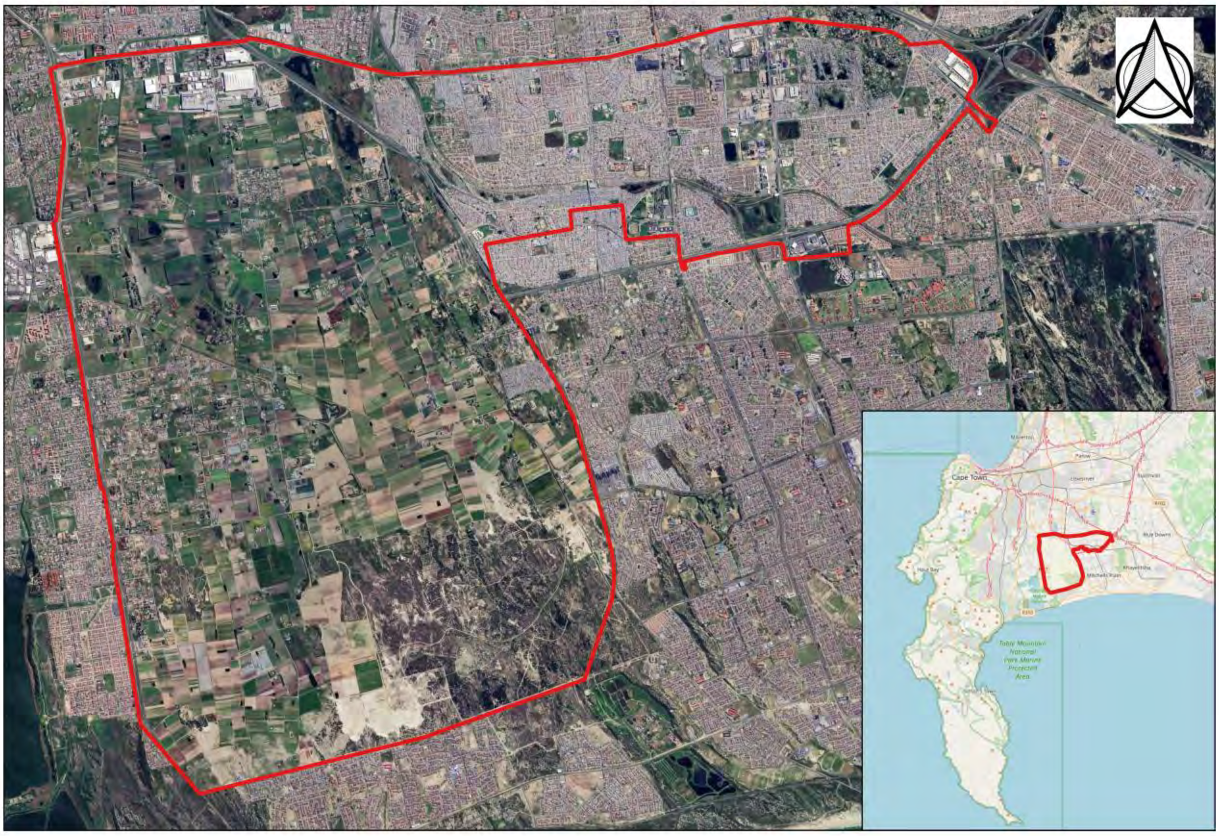


Figure 3.1 Locality map of the Philippi Horticultural Area (PHA). (Source: GeoEye)

The PHA study area is delineated by roads to the North, West and East. To the south by the urban township of Strandfontein and farms. The unique areas in the PHA are noted below:

1. The Highlands Estate (Area 2).

   This greater PHA area includes a broad range of both formal and informal land uses and comprises 3,168.65ha (excluding Highlands Estate) (Setplan, 2017). The dominant land use classes in the PHA are illustrated in Figure 3.2 as:

   a. A dominance of agriculture and smallholdings, with smallholdings being inclusive of lifestyle/ residential/ mixed-use smallholdings (e.g. Schaapkraal: Area 1) and small commercial farms;

   b. Occurrence of urban land uses (including residential, industrial, business, etc.) in Schaapkraal, Knole Park and Schaapkraal Estate (Area 1), the northern area abutting the Lansdowne Industrial Area (Area 3), Highlands residential area (Area Highlands) and west of Weltevreden Road (Area 4).

c. Area 5 is predominantly thicket and grassland-covered dunes together with mining.
d. Area 6 is predominantly agriculture and smallholdings, together with dune and grassland chapter.

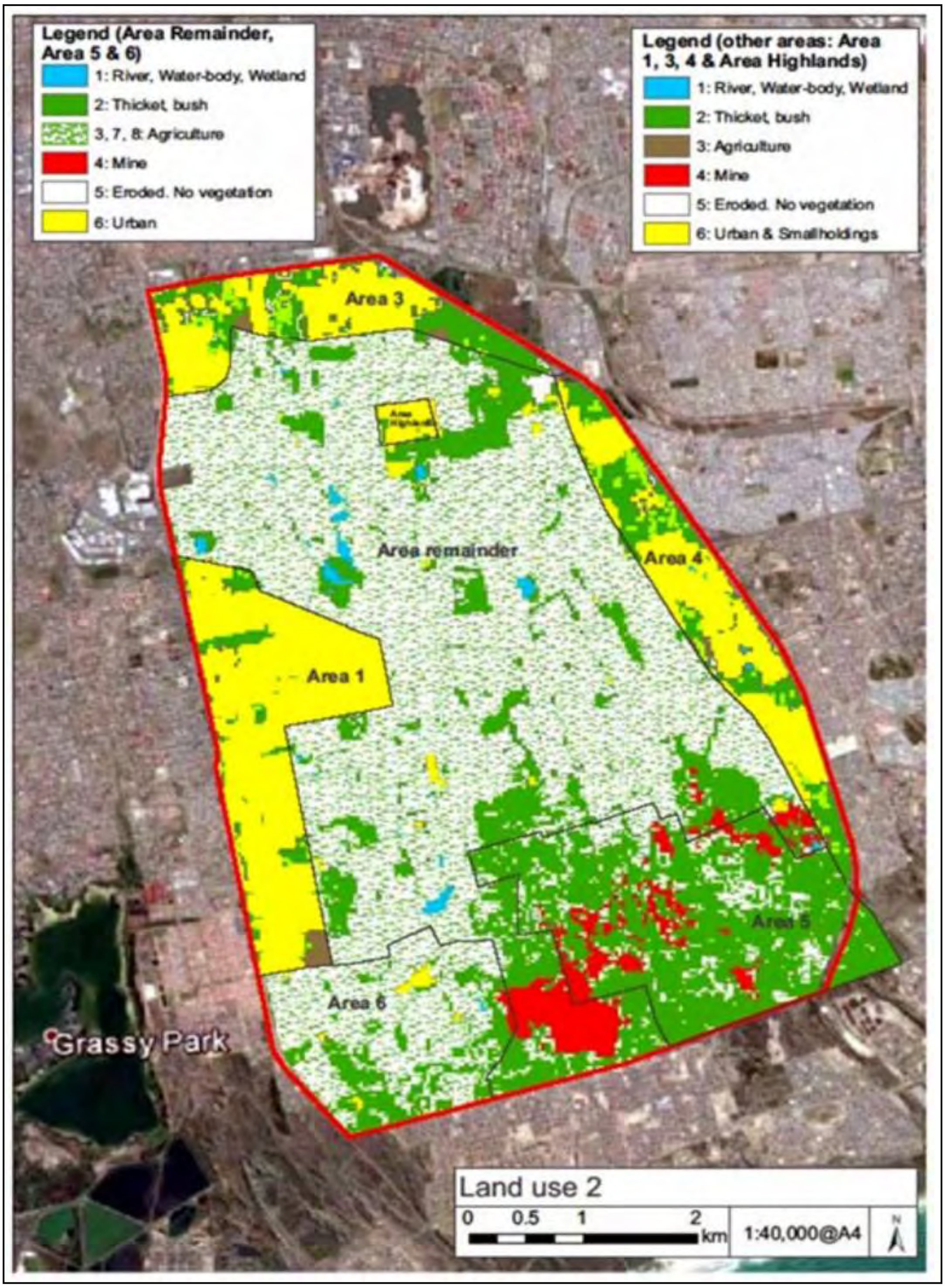


Figure 3.2 Map showing the predominant land use around the PHA prior to 1967 (City of Cape Town 2012).

Figure 3.3 The outer boundary represented by the red line indicates the reduction of the PHA core owing to a sequence of planning decisions (City of Cape Town 2012).

3.2 Data collection

The study used multispectral planet data. The three most commonly used satellites are Sentinel, Landsat and SPOT. The PlanetScope satellite produces good resolution of between 3.7 and 4 meters native resolution while Landsat has a moderate resolution of about 15 – 30 meters. The data was eventually downloaded from the Planet website. Each Planet Scope satellite is a CubeSat 3U form factor (10 cm by 10 cm by 30 cm). The complete Planet Scope constellation, which consists of approximately 130 satellites, is able to image the entire land surface of the Earth every day (equating to a daily collection capacity of 200 million km²/day). This satellite was chosen because it has a high resolution and it can be applied to applications like monitoring rapid changes in vegetation and land use and detailed vegetation mapping, which compliments this study quite well. (Planet (2021). Available from https://www.planet.com/ (Accessed: 01/10/2021). The data has already been pre-processed for atmospheric correction.

Table 3.1 below displays the description of the bands used as well as when the imagery was captured.

3.2.1 Data

The tables below indicate the satellite data used and its attributes. The PS2 instrument was utilised for this study. The differences between these instruments are the accessibility of the imagery and the enhancement of the telescopes over time.

Table 3.1 Constellation overview and sensor specification: PlanetScope.

Mission Characteristics	Sun-synchronous Orbit		
Instrument	PS2	PS2.SD	PSB.SD
Instrument description	Name: Dove Classic This equipment, which was constructed using a telescope known as "PS2," records red, green, blue, and near-infrared channels. It generates scene products that are roughly 25 x 11.5 square kilometers. From July 2014 to April 29, 2022, when the earliest imagery was accessible.	Name: Dove-R This instrument, which was constructed using the same "PS2" telescope but updated Bayer pattern and pass-band filters, records red, green, blue, and near infrared channels. It creates items called Scene that are roughly 25 x 23 square kilometers. The earliest imagery that was accessible was between March 2019 and April 22, 2022.	Name: SuperDove This device, which uses the same "PSB" telescope and filter response as PS2.SD, collects red, green, blue, near infrared, as well as a new red edge, green I, coastal blue, and yellow channel. It creates goods called Scene that are roughly 32.5 x 19.6 sq km. The earliest imagery that is now trackable is from mid-March 2020.
Orbit Altitude (reference)	475 km (~98° inclination)		
Max/Min Latitude Coverage	±81.5° (depending on season)		
Equator Crossing Time	9:30 - 11:30 am (local solar time)		
Sensor Type	Four-band frame Imager with a split-frame VIS+NIR filter	Four-band frame imager with butcher-block filter providing blue, green, red, and NIR stripes	Eight-band frame imager with butcher-block filter providing coastal blue, blue, green I, green II, yellow, red, red-edge, and NIR stripes
Spectral Bands	Blue: 455 - 515 nm Green: 500 - 590 nm Red: 590 - 670 nm	Blue: 464 - 517 nm Green: 547 - 585 nm Red: 650 - 682 nm	Blue: 465 - 515 nm Green: 547 - 585 nm Red: 650 - 680 nm Coastal Blue: 431 – 452

			nm Green I: 513 – 549 nm Red Edge: 697 – 713 nm Yellow: 600 – 620 nm
Band 4	NIR: 780 – 860 nm	NIR: 846 – 888 nm	NIR: 845 – 885 nm
Ground Sample Distance (nadir)	3.7 m (approximate)	3.7 m (approximate)	3.7 m (approximate)
Frame Size	24 km x 8 km (approximate)	24 km x 16 km (approximate)	32.5 km x 19.6 km (approximate)
Maximum Image Strip per orbit	20,000 km²	20,000 km²	20,000 km²

Table 3.2 Indicates the sensor, bands and resolution.

Date	Sensor	Bands Captured	Ground resolution (m)
12 April 2015	PlanetScope	Red, Green Blue and NIR	3.7
05 April 2016	PlanetScope	Red, Green Blue and NIR	3.7
24 April 2017	PlanetScope	Red, Green Blue and NIR	3.7
17 April 2018	PlanetScope	Red, Green Blue and NIR	3.7
19 April 2019	PlanetScope	Red, Green Blue and NIR	3.7
06 April 2020	PlanetScope	Red, Green Blue and NIR	3.7
29 April 2021	PlanetScope	Red, Green Blue and NIR	3.7

3.2.2 Research equipment

The equipment that was utilised was a desktop computer running Quantum GIS version 3.8 on a Windows 10 operating system.

3.3 Vegetation indices

This study assessed the performance of vegetation indices in improving classification. In total, 20 vegetation indices were created. The list of indices is shown under section 2.3 in Table 2.1. The indices were created in QGIS using the Raster calculator algorithm. The figure below shows a screenshot of the methodology. Vegetation indices were created using the raster calculator tool in QGIS. The corresponding bands were subtracted, added or ratioed deposing on the equations in table 3.4. The results were saved as GeoTiff files and added to the project for feature selection.

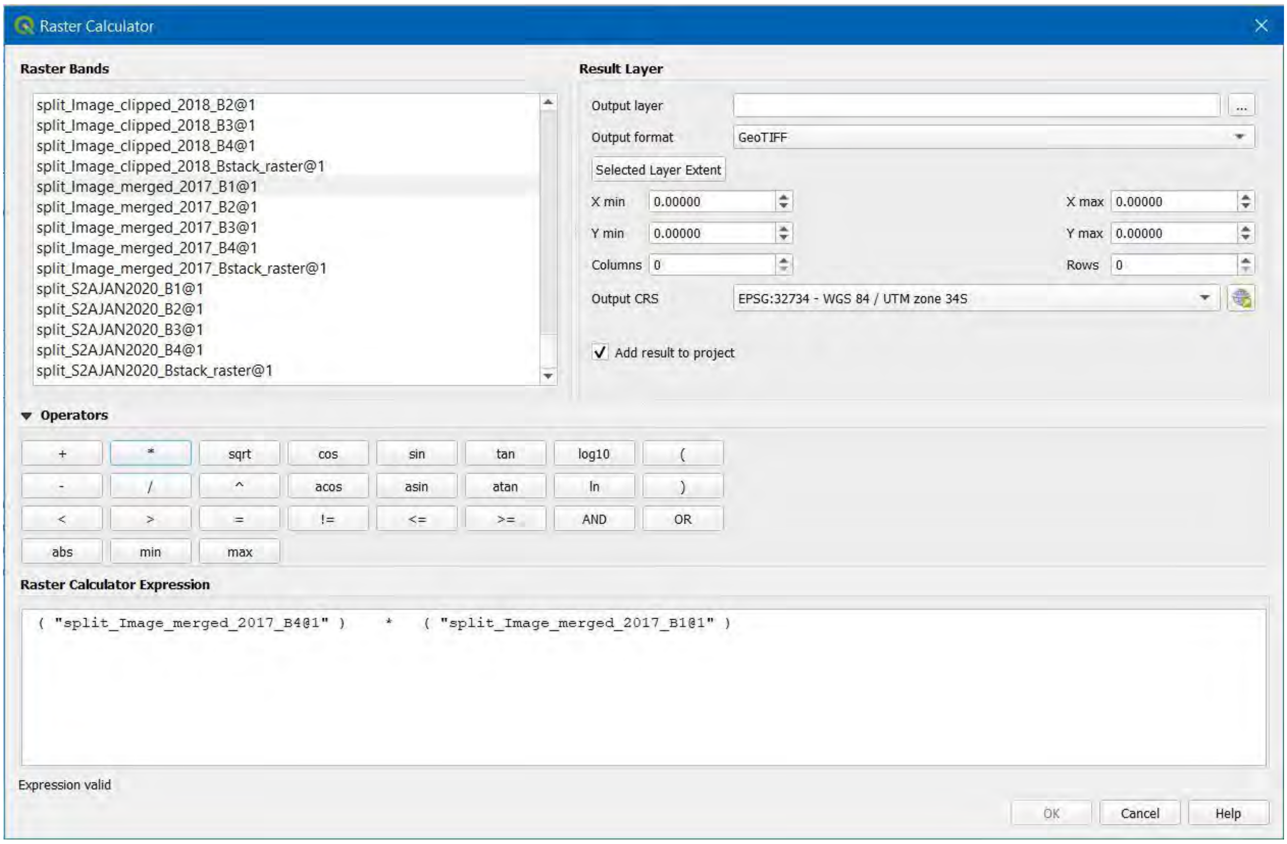


Figure 3.4 Raster calculation example of vegetation indices.

3.4 Feature selection

A shapefile was created and sample polygons of the classes of interest were created over the satellite data. There were 6 classes created, namely urban, water, vegetation, natural vegetation bare ground and bare ground-sand. The shapefile was split into two processing methods using the vector manipulation and training geoprocessing tools in QGIS. Seventy percent of the polygons were used as the training sample and 30% were retained as an independent validation sample. A total of 3 383 points were used as training points with a spacing of four meters in QGIS (See figure below). They were created using the 'random points in the polygon' algorithm. The choice of four meters was to ensure that the spacing between the points was more than the spatial resolution of the satellite data which was 3.7 meters.

As mentioned, twenty vegetation indices were created based on the list in table 2.2. The reflectance values of both the indices and the satellite bands were sampled by intersecting the 3 383 random points with the bands and indices. The data at each point consisted of corresponding reflectance values of bands and indices. The training and validation data were exported to spreadsheets with rows corresponding to the points and columns corresponding to reflectance values of bands and indices. A sample of the reflectance values is shown in Table 3.3.

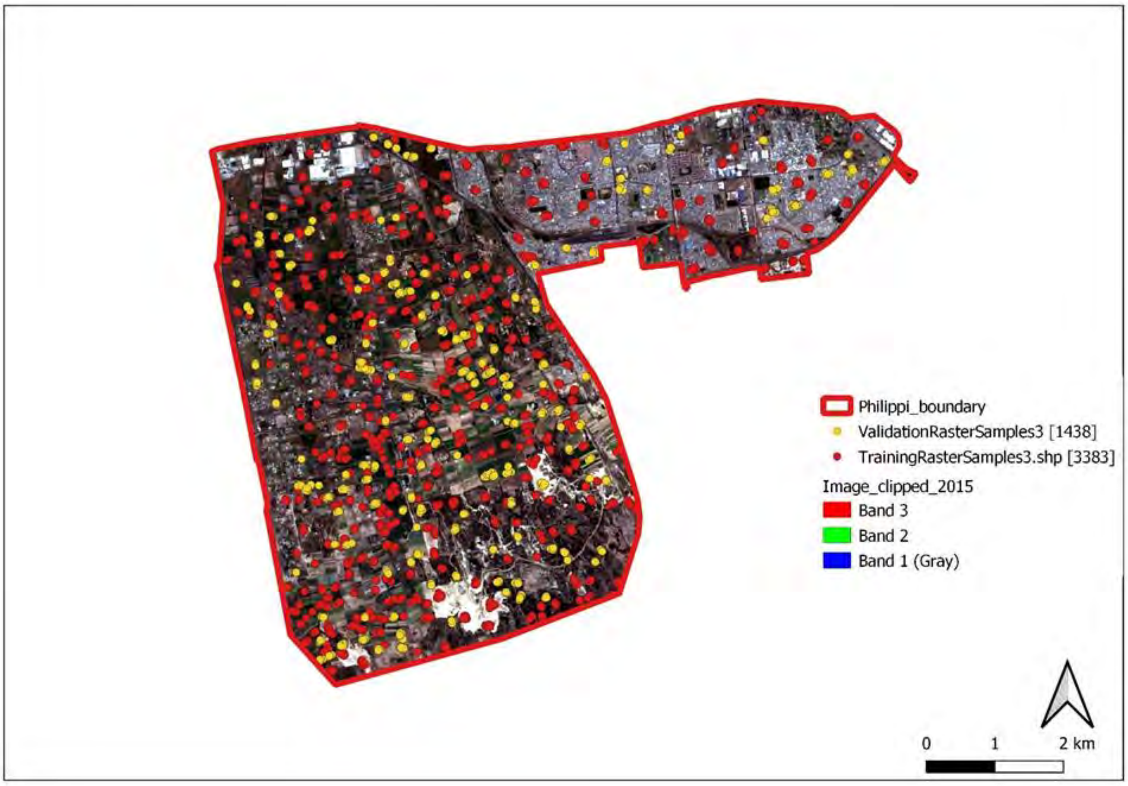


Figure 3.5: Map indicating training and validation samples done in the QGIS software.

Green	Red	NearIR	TVI_2015	RVI_2015	DVI_2015	VIN_2015	NDGI_201	RI_2015	RG_2015	NDVI_201!	GNDVI_20	EV12
0.1188	0.0789	0.4049	1.08344	0.19486	0.326	5.13181	0.20182	-0.20182	0.039	0.67383	0.54631	0.51121
0.1218	0.0831	0.4335	1.08549	0.1917	0.3504	5.21661	0.18887	-0.18887	0.0444	0.67828	0.56132	0.53646
0.1142	0.0749	0.433	1.09775	0.17298	0.3581	5.78104	0.20783	-0.20783	0.0356	0.70506	0.5826	0.5551
0.1173	0.0746	0.4368	1.09921	0.17079	0.3622	5.85523	0.22251	-0.22251	0.0319	0.70825	0.57661	0.56039
0.1215	0.0842	0.4287	1.08244	0.19641	0.3445	5.09145	0.18133	-0.18133	0.0469	0.67167	0.55834	0.52812
0.1249	0.0876	0.4303	1.07783	0.20358	0.3427	4.9121	0.17553	-0.17553	0.0503	0.66171	0.55007	0.52224
0.1142	0.0746	0.4125	1.09256	0.18085	0.3379	5.52949	0.20975	-0.20975	0.035	0.6937	0.56636	0.53078
0.1188	0.0789	0.4049	1.08344	0.19486	0.326	5.13181	0.20182	-0.20182	0.039	0.67383	0.54631	0.51121
0.1133	0.0704	0.4301	1.10394	0.16368	0.3597	6.10938	0.23353	-0.23353	0.0275	0.71868	0.583	0.56236
0.1189	0.0795	0.4029	1.08185	0.19732	0.3234	5.06792	0.19859	-0.19859	0.0401	0.6704	0.54427	0.50731
0.1119	0.0824	0.4922	1.10145	0.16741	0.4098	5.9733	0.15183	-0.15183	0.0529	0.71319	0.62953	0.60623
0.1127	0.0814	0.5073	1.1061	0.16046	0.4259	6.23219	0.16126	-0.16126	0.0501	0.72346	0.63645	0.62535
0.1147	0.083	0.4945	1.10116	0.16785	0.4115	5.95783	0.16034	-0.16034	0.0513	0.71255	0.62344	0.6074
0.113	0.0844	0.4932	1.09898	0.17113	0.4088	5.8436	0.14488	-0.14488	0.0558	0.70776	0.62719	0.60268
0.1127	0.0851	0.4673	1.09174	0.18211	0.3822	5.49119	0.13953	-0.13953	0.0575	0.69189	0.61138	0.57163

Table 3.2 Sampled reflectance values which was generated through QGIS software using the raster sampling tool.

The Feature selection was done in R Studio using the Boruta library. The Boruta package and corresponding packages was installed on the R Studio application as well as the libraries attached to process the application. The Excel spreadsheets which had the training and validation data were imported into the R studio application. The feature selection was then done using Boruta and converted. All 18 indices were assessed, and the top 4 indices were selected. The data was then plotted to display the outcome (refer to figure 4.2 for the results).

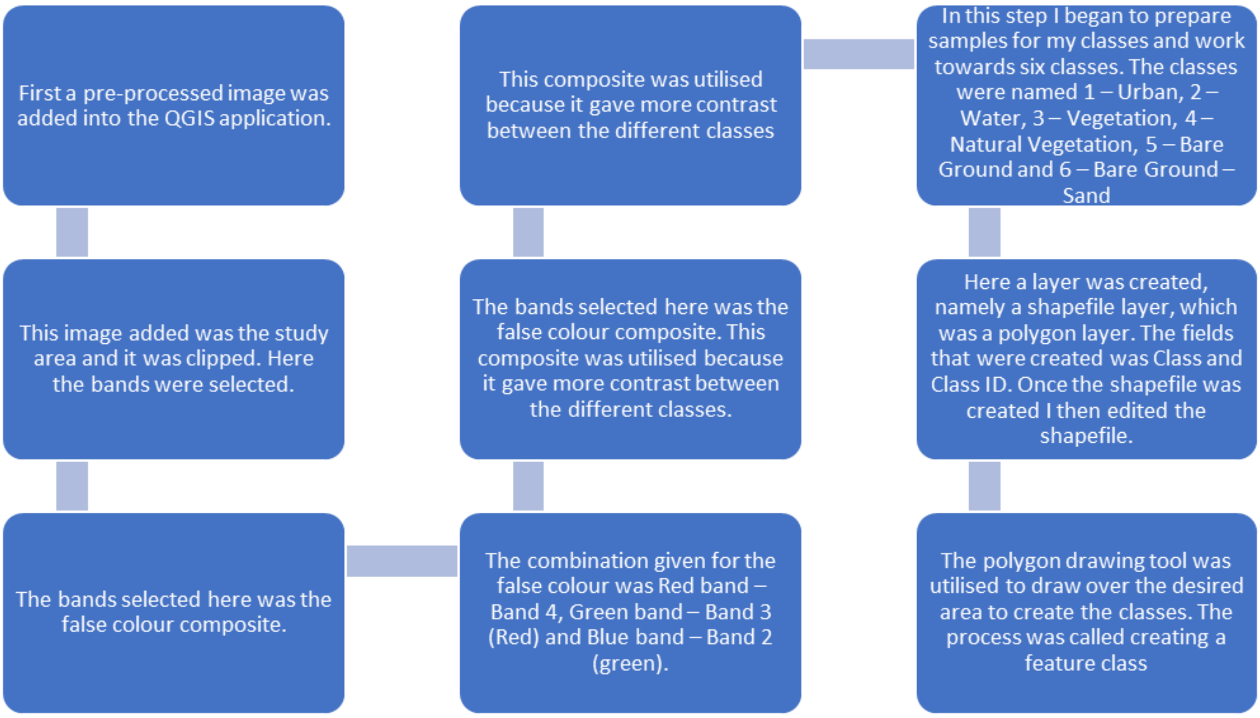
3.5 Classification

TSeveral remote sensing plugins, such as the semi-automatic categorization (Luca, 2020) and Dzetsaka (Karasiak, 2017) were utilised within Quantum GIS. These open-source plugins have been utilized in a number of urban studies, including those by Leroux et al. (2018), Sejati, Buchori, and Rudiarto (2019) and Sejati et al. (2020). These applications were utilized during the pre-processing, classification, and derivation of classification accuracy. In this study, the primary features to be classified were urban, water, natural vegetation, vegetation and sand or bare ground. These classes were predominantly present within the area and would speak well to the classification and end results being determined.

3.5.1 Classification imaging process

Pre-processing Satellite Imagery using QGIS and Semi-Automatic Classification Plugin

The flow chart below describes the process of the classification done.



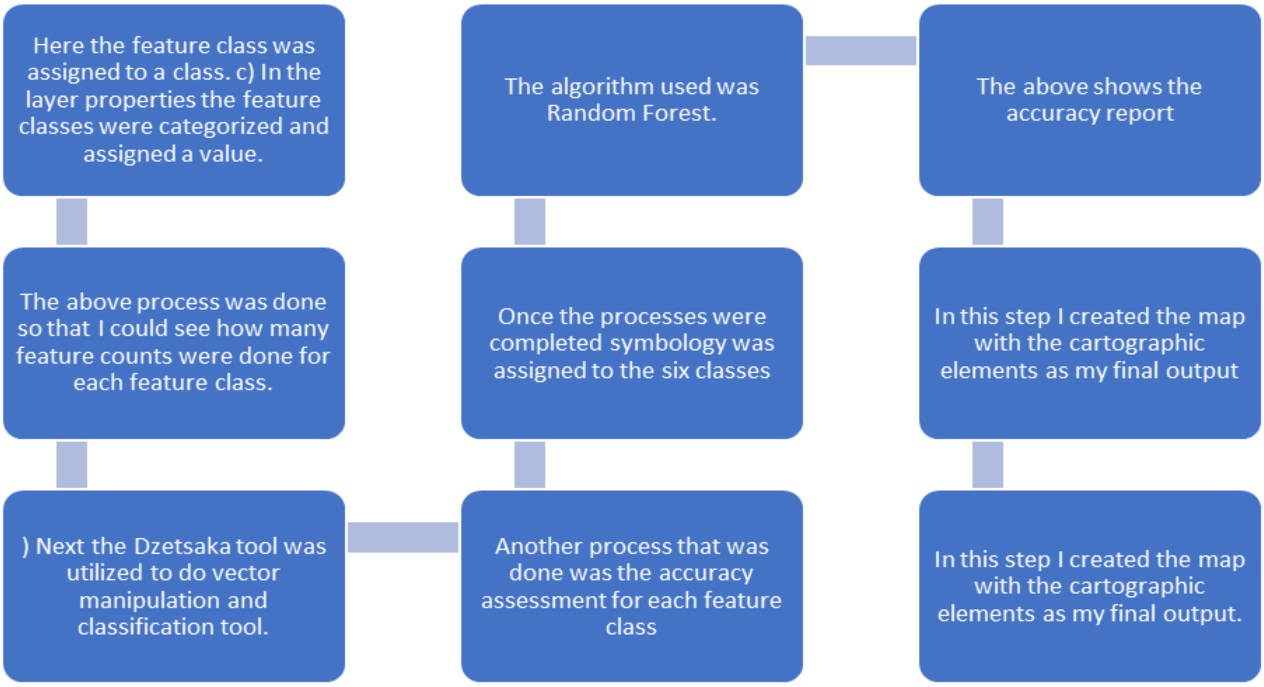


Figure 3.6: Flow chart indicating the classification process done on the images.

3.5.2 Selecting classifiers

Three classifiers were tested in this study. Only the data corresponding to the most important features was used in selecting a classifier. The assessment was done using the Caret library in R Studio. Classification models were built for Random Forest, SVM and KNN in R-Studio using the sampled reflectance data. The results of all 3 classifiers were compared to ascertain classification accuracy and kappa hat accuracy. (Refer to Appendix K). It was then concluded that Random Forest was the best classifier for this dataset due to it producing the highest accuracy levels of the three classifiers used (refer to figure 4.1 for the results).

3.5.3 Classification using machine learning

The classification process was done in QGIS software. First samples were created of 6 classes. In that process shapefiles were created, and each assigned with a feature name. These shapefiles were created over the PHA image. The shapefiles created were drawn over the specified feature on the image. Once there was a sufficient number of shapefiles created for the area the Dzetsaka classification tool was used to train and predict models for the feature classes. In the training algorithm the raster image was selected of the area as well as the classes created. A field was selected from the column of the shapefiles attribute table. Then the type of classification was selected from the three classifications selected for this study. Once that was done a pixel percentage was allocated. After all the inputs the training algorithm was run. After this the prediction model process was next. The raster for the area was selected again and the training model created was inserted into the process. Once this process had been completed the bands created were named and symbolized.

3.5.4 Accuracy assessment

After creating new raster outputs, the accuracy assessment took place in order to determine the accuracy of the classifications. This process was done for each year over a 7-year period. The semi-automatic classification plugin tool was used. In this process the classified image was used together with the training shapefile. From this process resulted a percentage to indicate the accuracy out of 100. Once the results were computed the raster calculation was done in conjunction with the vegetation indices to improve the accuracy levels of the classification.

3.5.5 Change Detection

The images that were classified together with the indices calculated were processed through the QGIS software utilising the land cover change tool. Two consecutive years were selected (e.g. 2015 and 2016) and run in the program. The program then processed the images, and the output created a percentage difference between each class. The output shows the change per class i.e., the regions that stayed the same and regions that changed from one class to another and corresponding acreages. Image differencing method was used for this study as in Viana et al., (2019) The outputs of the maps could distinguish between the different years and displaying a change in the area.

4. Results

In this chapter the results of the classifications and accuracies computed through the Quantum GIS software programme are presented.

4.1 Choosing classifiers

The figure below shows the comparison of the training accuracies of the classification models. There were three classifiers used in this research, namely Random Forest (RF), K-Nearest Neighbour (KNN) and Support Machine Vector (SVM). The figure below shows the results for training the classification models using random forest, k nearest neighbour and support vector machines respectively.

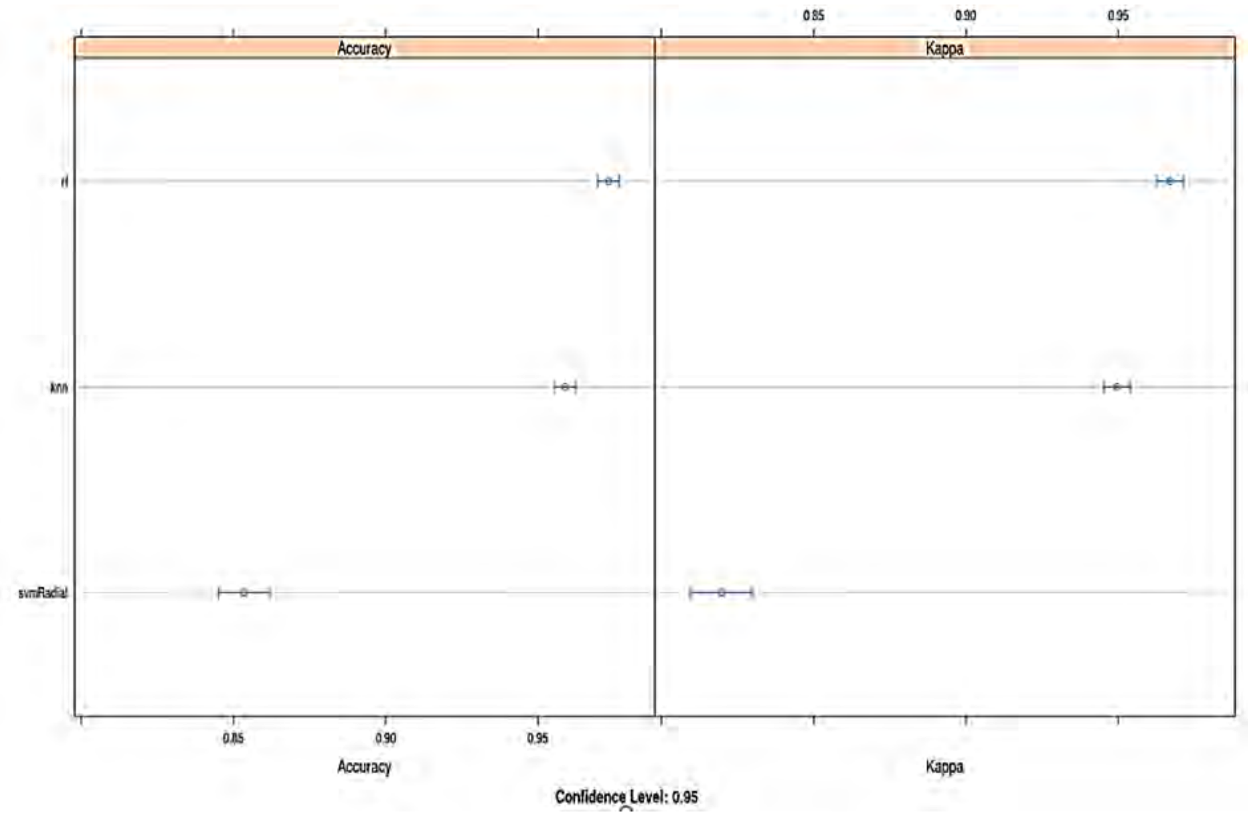


Figure 4.1 Classification model accuracy for all three classifiers used.

This figure indicates which classifier performed the best out of the three selected. On the left the figure indicates the overall accuracy and on the right the figure indicates the kappa accuracy Random Forest had the highest accuracies.

4.2 Identifying the best indices

The results in figure 4.2 below show the output from RStudio showing the relative feature importance as per the method described in section 3.6.2. There were 20 vegetation indices used in total to determine which ones were best suited to incorporate into the methodology to determine accuracy. The most relevant indices were MSAVI, RDVI, RG and GNDVI.

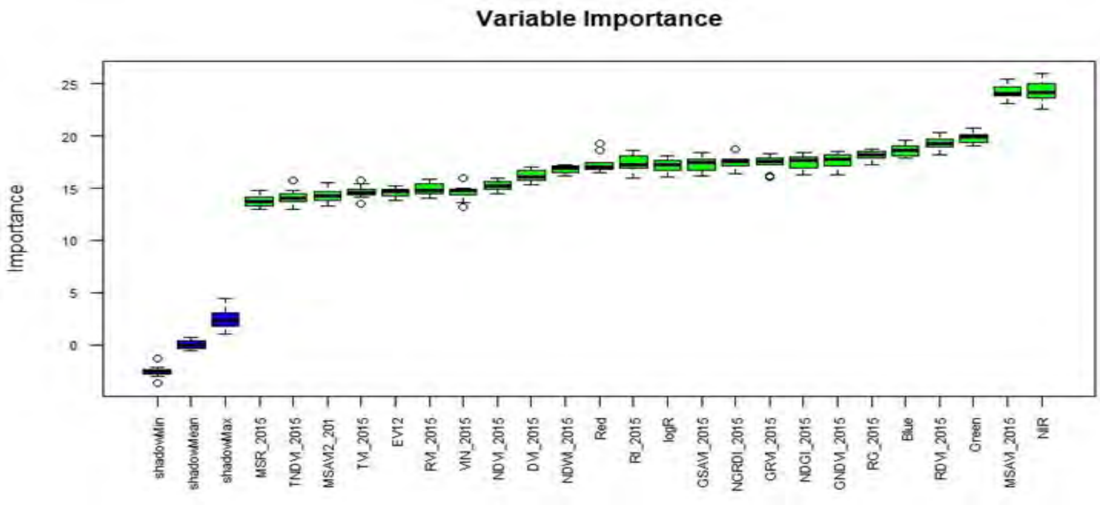


Figure 4.2 Results from Boruta showing feature importance. This figure displays which indices performed best when coding was applied.

4.3 Classification maps

4.3.1 Maps for 2015

The figures below show classification maps for the years 2015 – 2021.

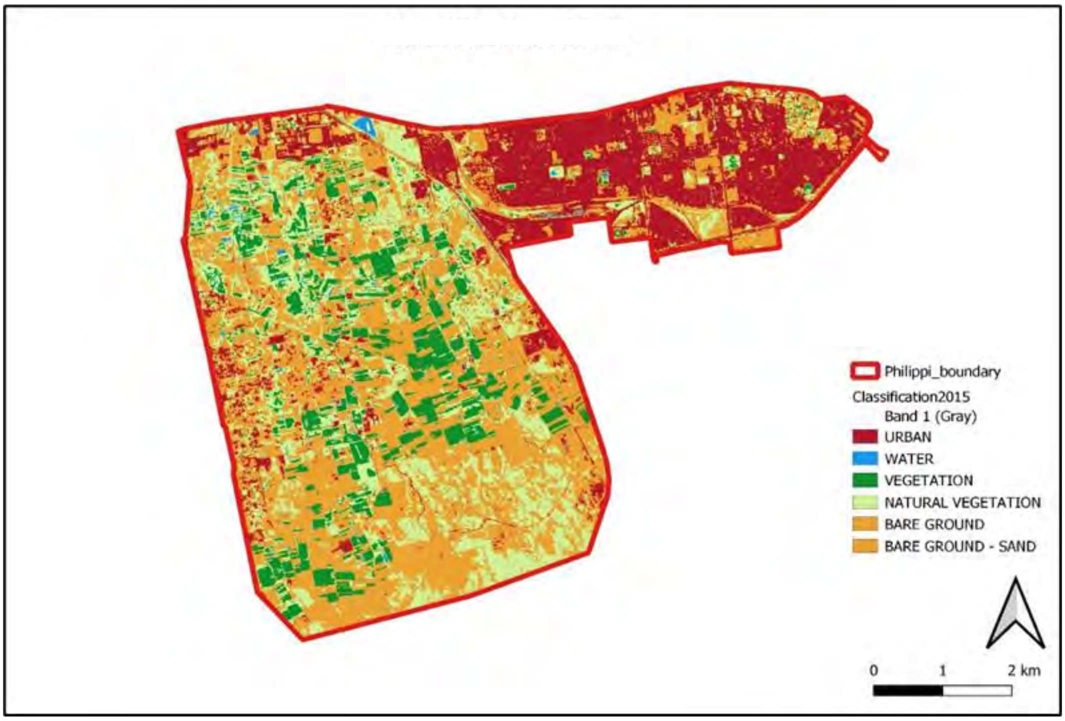


Figure 4.3. Map showing the classified image from 2015 for the layer stack of only original bands.

The classification is based on the layer stack including the most relevant vegetation indices to improve the classification outputs of the original process. s.

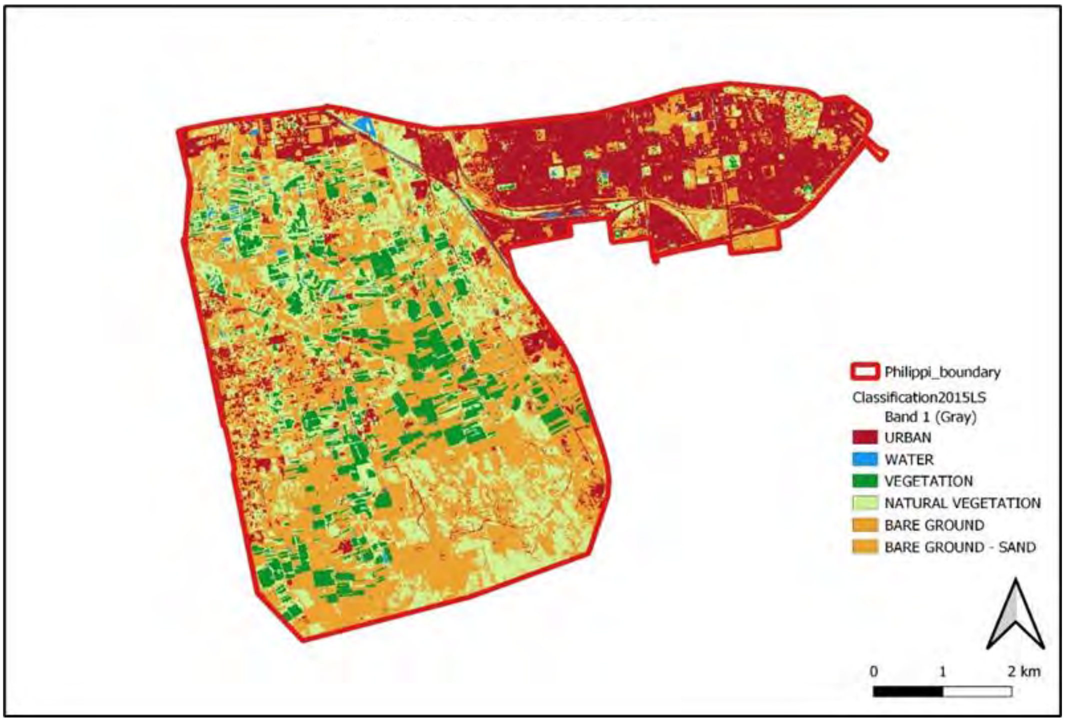


Figure 4.4. Map showing the classified image from 2015 for the layer stack of original bands and best indices.

The classification is based on the layer stack including the most relevant vegetation indices to improve the classification outputs of the original process.

4.3.2 Maps for 2016

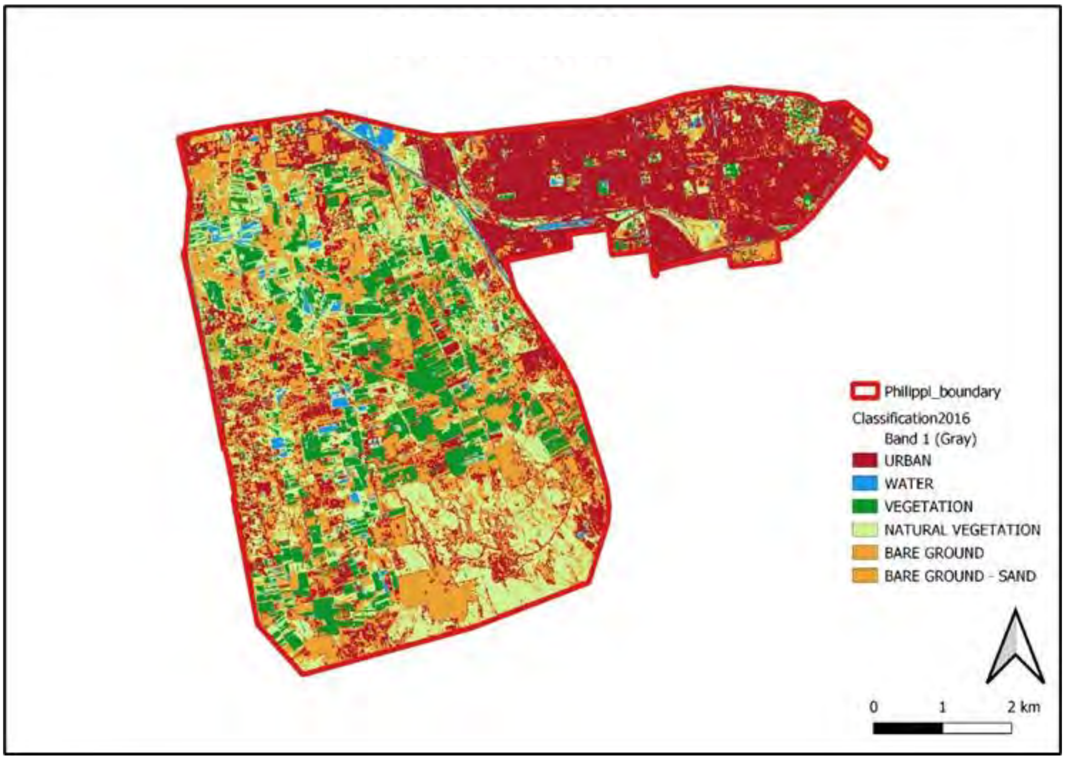


Figure 4.5. Map showing the classified image from 2016 for the layer stack of only original bands. These classifications were done without vegetation indices and indicated the difference in accuracy levels.

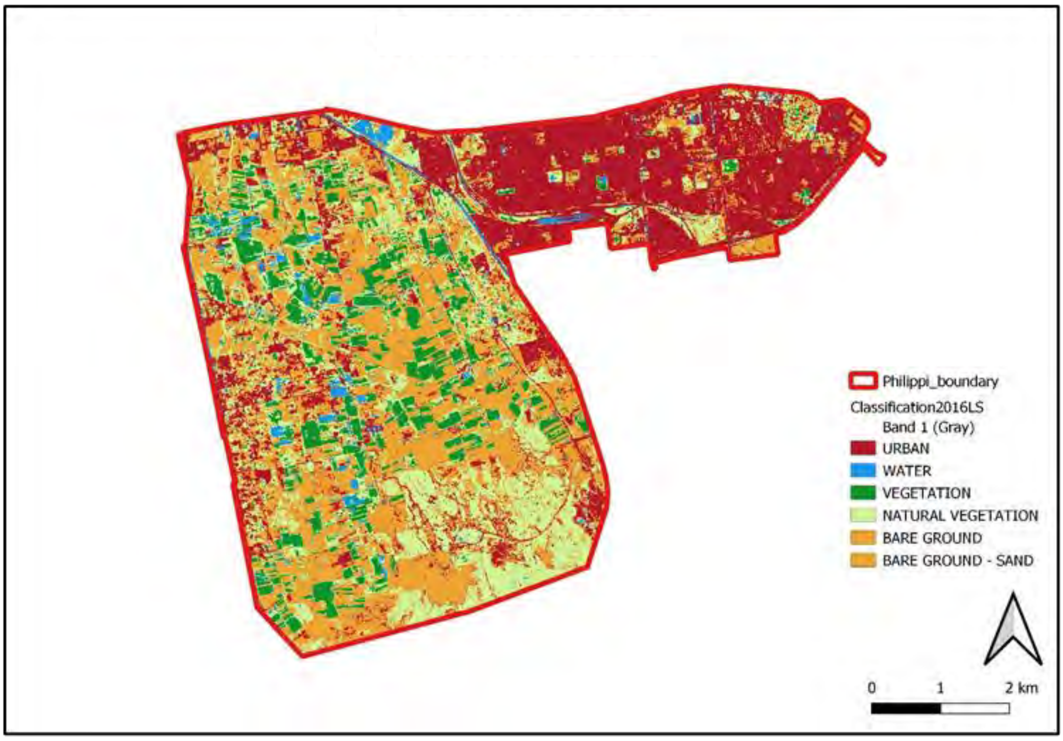


Figure 4.6. Map showing the classified image from 2016 for the layer stack of only original bands and best indices.

The classification is based on the layer stack including the most relevant vegetation indices to improve the classification outputs of the original process.

4.3.3 Maps for 2017

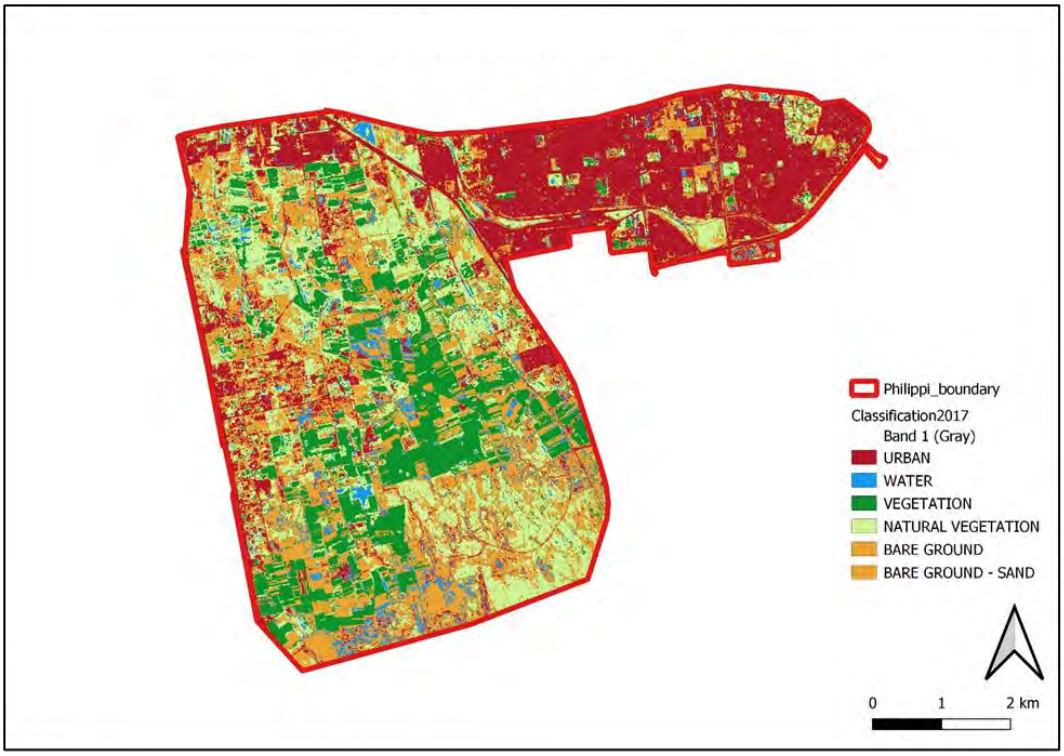


Figure 4.7. Map showing the classified image from 2017 for the layer stack of only original bands. These classifications were done without vegetation indices and indicated the difference in accuracy levels.

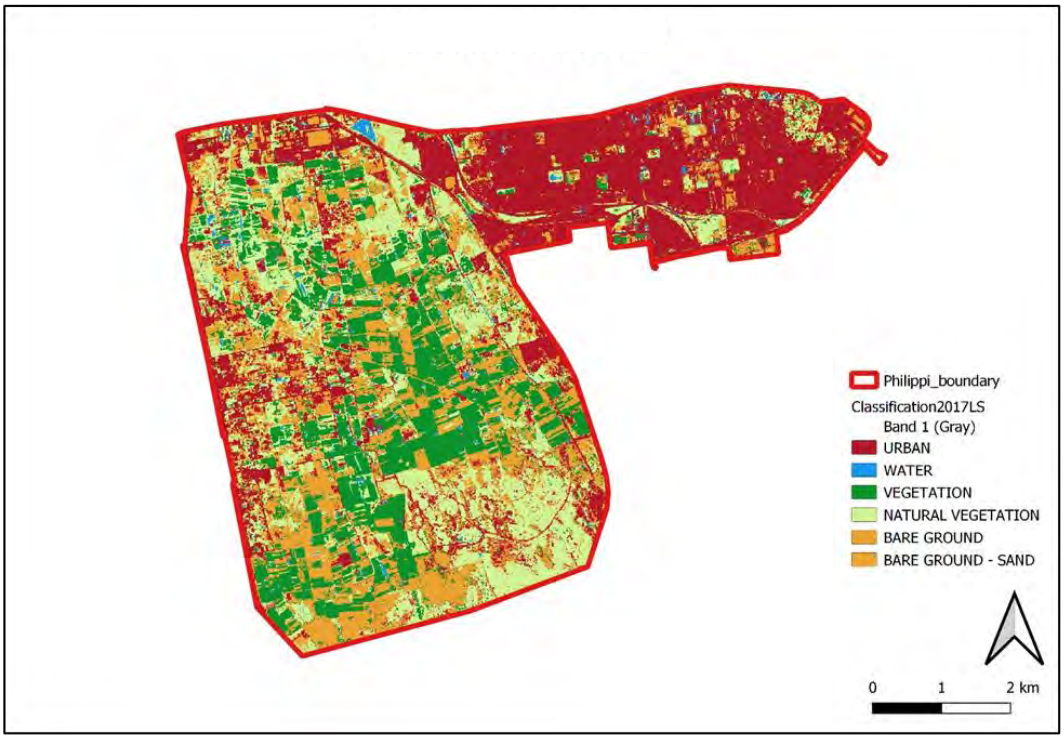


Figure 4.8. Map showing the classified image from 2017 for the layer stack of only original bands and best indices.

The classification is based on the layer stack including the most relevant vegetation indices to improve the classification outputs of the original process.

4.3.4 Maps for 2018

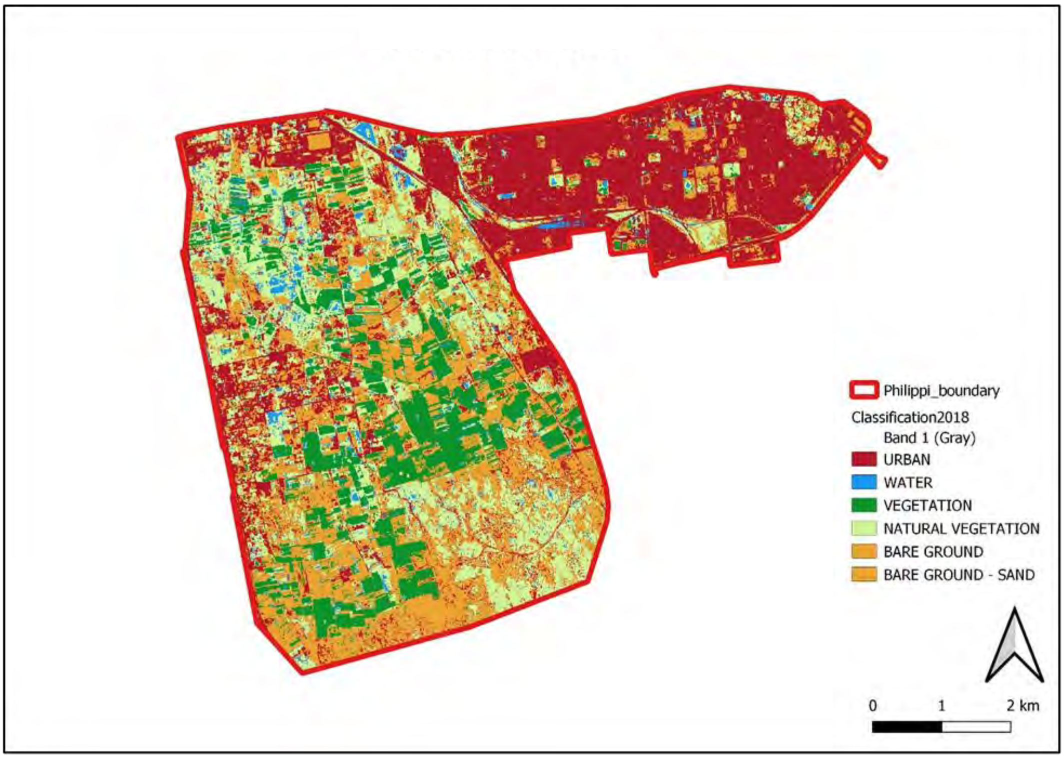


Figure 4.9. Map showing the classified image from 2018 for the layer stack of only original bands. These classifications were done without vegetation indices and indicated the difference in accuracy levels.

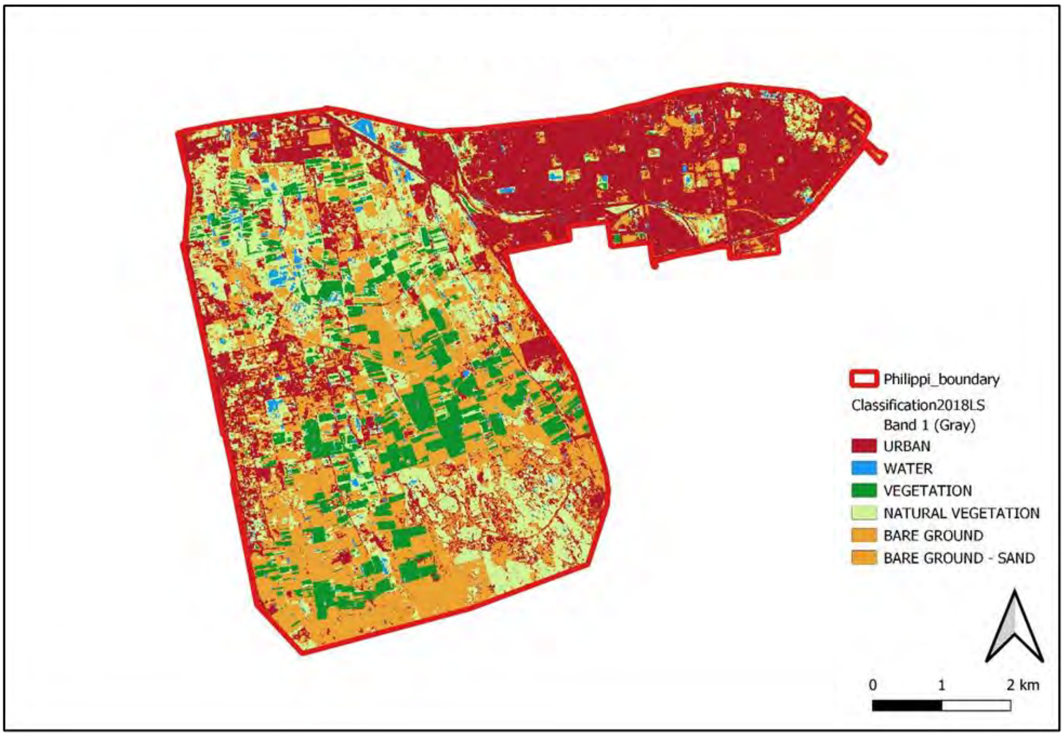


Figure 4.10. Map showing the classified image from 2018 for the layer stack of only original bands and best indices.

The classification is based on the layer stack including the most relevant vegetation indices to improve the classification outputs of the original process.

4.3.5 Maps for 2019

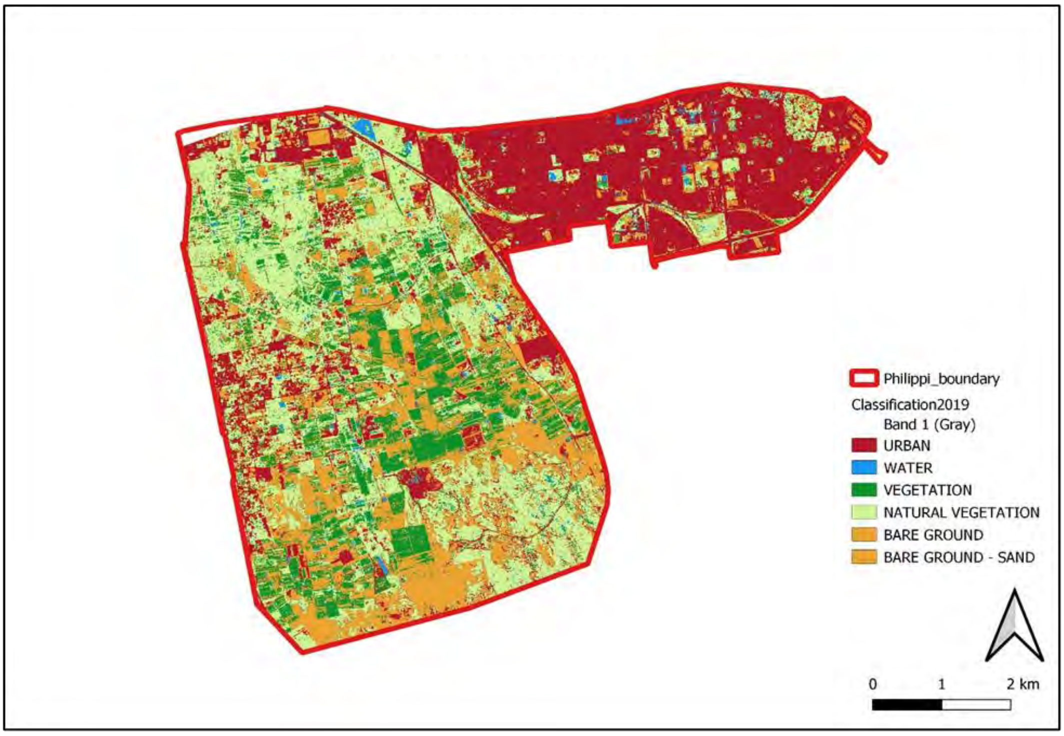


Figure 4.11. Map showing the classified image from 2019 for the layer stack of only original bands. These classifications were done without vegetation indices and indicated the difference in accuracy levels.

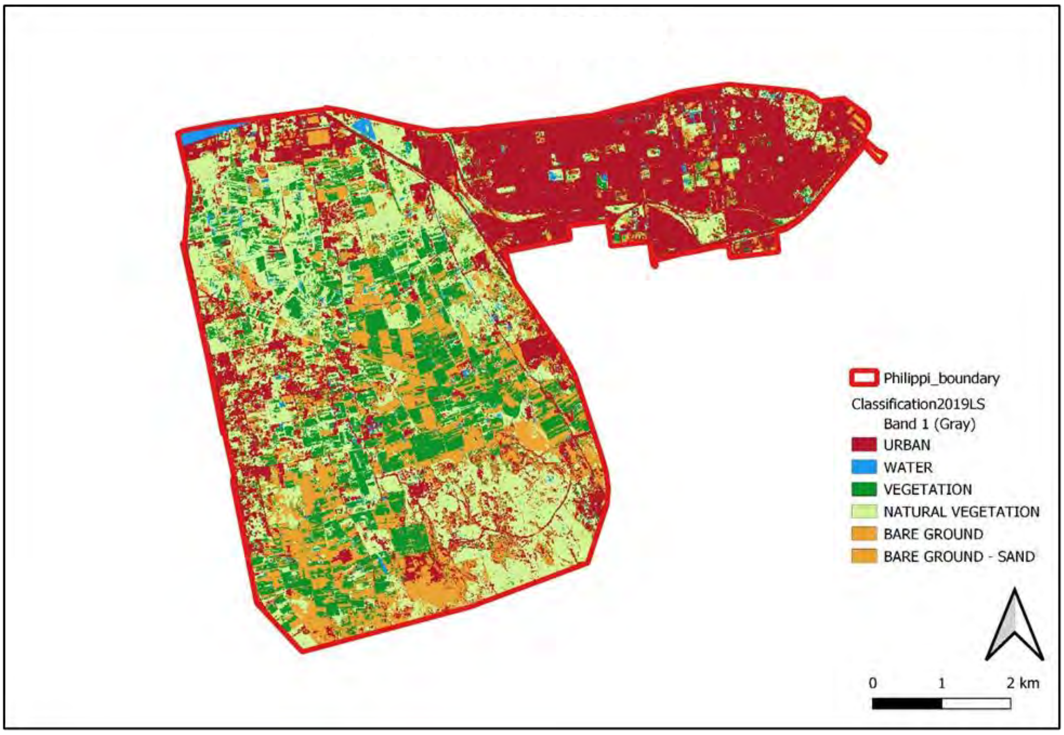


Figure 4.12. Map showing the classified image from 2019 for the layer stack of only original bands and best indices.

The classification is based on the layer stack including the most relevant vegetation indices to improve the classification outputs of the original process.

4.3.6 Maps for 2020

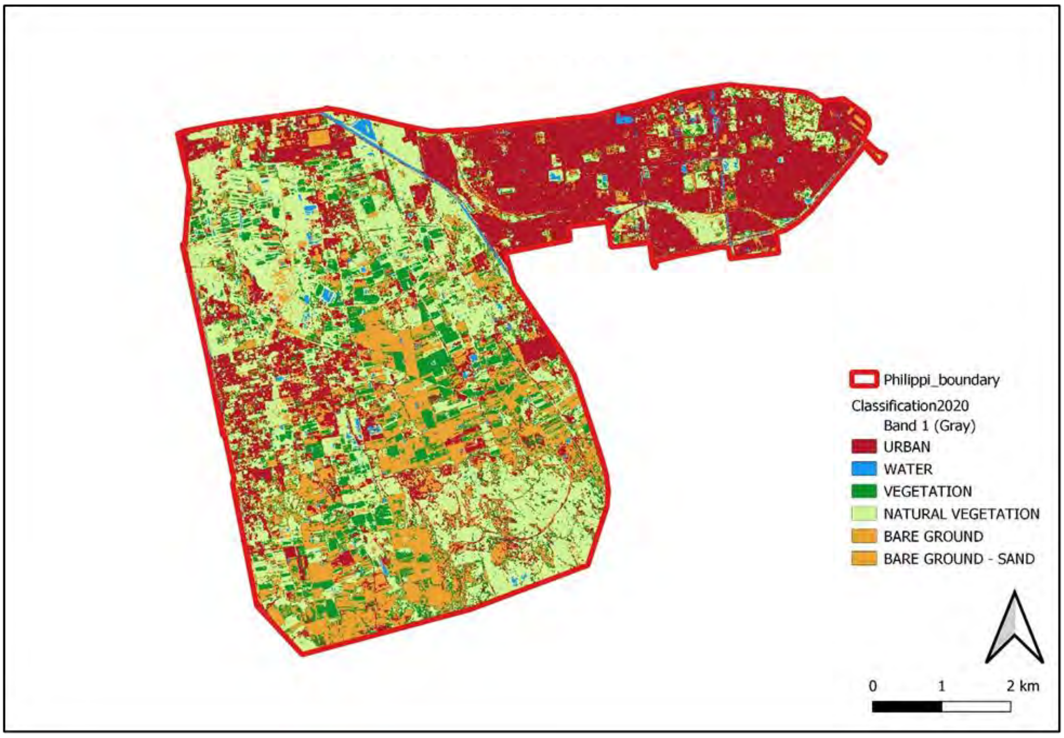


Figure 4.13. Map showing the classified image from 2020 for the layer stack of only original bands. These classifications were done without vegetation indices and indicated the difference in accuracy levels.

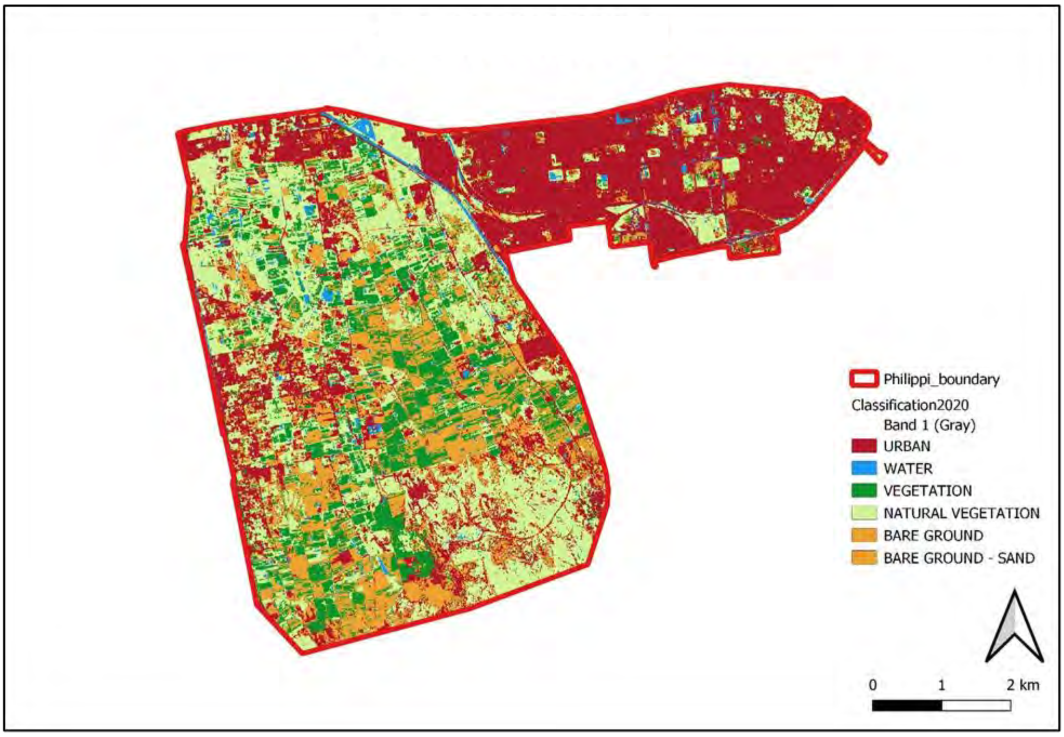


Figure 4.14. Map showing the classified image from 2020 for the layer stack of only original bands and best indices.

The classification is based on the layer stack including the most relevant vegetation indices to improve the classification outputs of the original process.

4.3.7 Maps for 2021

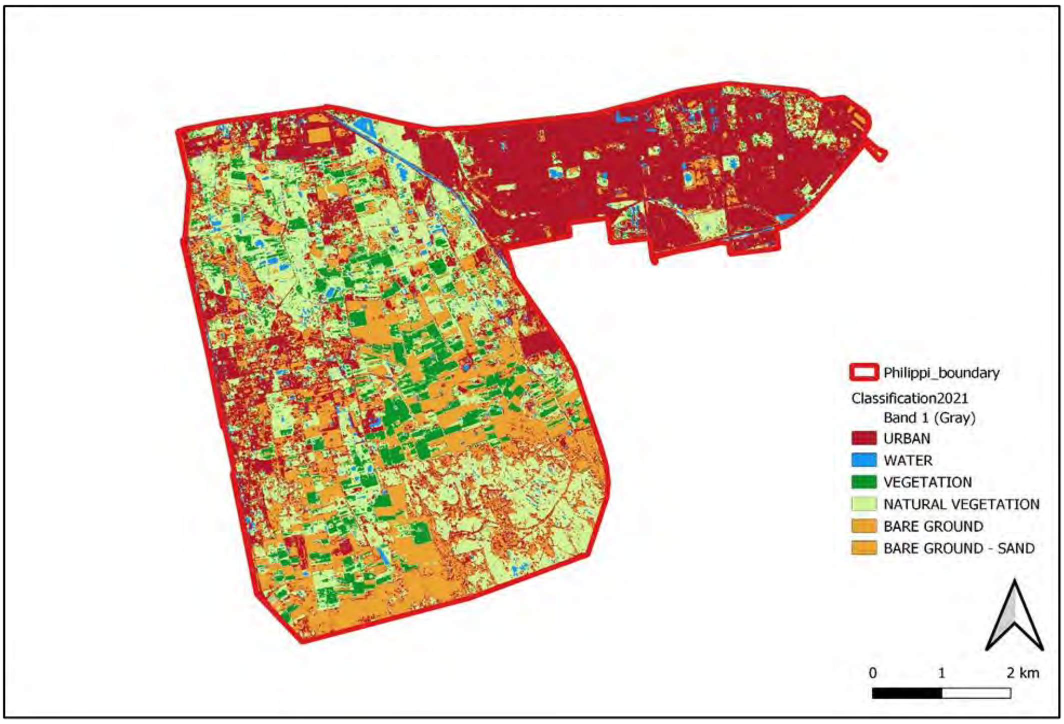


Figure 4.15. Map showing the classified image from 2021 for the layer stack of only original bands. These classifications were done without vegetation indices and indicated the difference in accuracy levels.

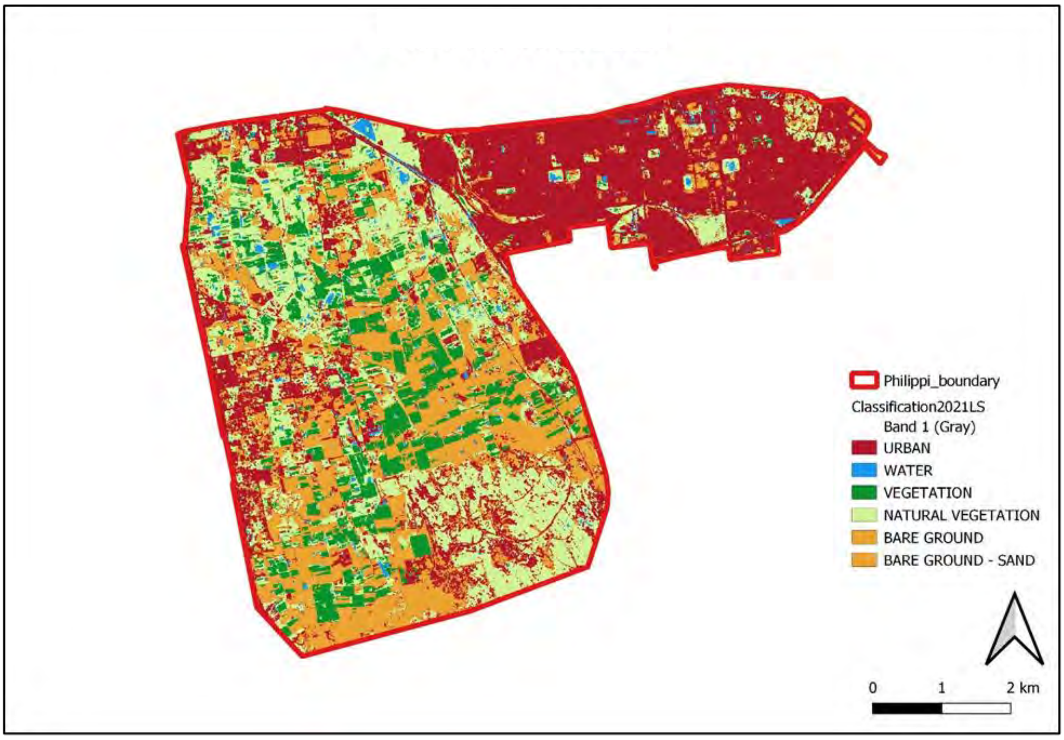


Figure 4.16. Map showing the classified image from 2021 for the layer stack of only original bands and best indices.

The classification is based on the layer stack including the most relevant vegetation indices to improve the classification outputs of the original process.

4.4 Classification statistics

The tables below show the error statistics for the study period.

4.4.1 Classification accuracies after using Random Forest

Table 4.1 Classification accuracies Random Forest

Random Forest		1	2	3	4	5	6
2015	PA [%]	96.1165	99.6686	100	97.2307	97.342	100
	UA [%]	97.7492	100	96.875	97.2477	96.5517	100
	Kappa hat	0.9745	1	0.967	0.9688	0.9558	1
2016	PA [%]	98.512	99.4562	88.938	71.3964	81.8696	96.729
	UA [%]	92.2734	94.0092	95.2904	81.6203	92.1212	97.7707
	Kappa hat	0.9075	0.8908	0.949	0.7871	0.9092	0.9768
2017	PA [%]	86.993	98.9514	44.5345	74.2191	44.4366	71.8397
	UA [%]	91.8968	55.5124	91.2546	82.4096	73.9511	79.9084
	Kappa hat	0.9032	0.3896	0.893	0.799	0.6575	0.7952
2018	PA [%]	77.7961	99.1431	79.53	80.5706	76.8712	99.628
	UA [%]	94.1626	82.4405	79.6767	91.8967	92.9977	96.7647
	Kappa hat	0.9266	0.7058	0.7805	0.9072	0.9149	0.9672
2019	PA [%]	99.1003	84.5925	91.7558	98.6311	97.4437	98.881
	UA [%]	99.6485	99.378	94.5953	96.7373	96.0822	99.8172
	Kappa hat	0.995	0.9937	0.9364	0.9534	0.9504	0.9981
2020	PA [%]	96.1536	68.9507	54.1607	92.1101	65.7438	88.5996
	UA [%]	93.224	80.3279	64.7443	83.445	67.6333	95.092
	Kappa hat	0.9029	0.7986	0.5738	0.7707	0.5898	0.9503
2021	PA [%]	96.9235	72.1193	71.7943	93.3405	94.5264	96.6765
	UA [%]	97.5107	72.5924	89.4999	84.4186	95.3593	97.8867
	Kappa hat	0.961	0.7177	0.8812	0.7897	0.9408	0.9785

In the table above, 1 denotes urban, 2 denotes water, 3 denotes vegetation, 4 denotes natural vegetation, 5 denotes bare ground and 6 denotes bare ground-sand.

The table below shows that layer stack (LS) are the ones with best indices and has had indices added to the process.

Table 4.2 Overall classification accuracies of all six classes during the 7-year period.

Year	Urban Fabric	Water	Vegetation	Natural Vegetation	Bare Ground	Bare Ground-Sand
2015	96.1	99.7	100	97.2	97.3	100
2015LS	97.8	99.7	100	98.2	97.5	100
2016	98.5	99.5	88.9	71.4	81.9	96.7
2016LS	99.6	99.9	99.1	94.4	98.8	98.4
2017	87	99	44.5	74.2	44.4	71.8
2017LS	99.2	99.8	90.1	96.3	96.8	99.4
2018	77.8	99.1	79.5	80.6	76.9	99.6
2018LS	99.1	99.9	99.3	95.6	99.3	100
2019	99.1	84.6	91.8	98.6	97.4	98.9

2019LS	99.4	99.7	93.8	99.4	97.3	98.3
2020	96.12	69	54.2	92.1	65.7	88.6
2020LS	99.7	99.8	96.7	99.4	97.9	94.4
2021	96.9	72.1	71.8	93.3	94.5	96.6
2021LS	98.8	100	98.7	99.6	100	99.4

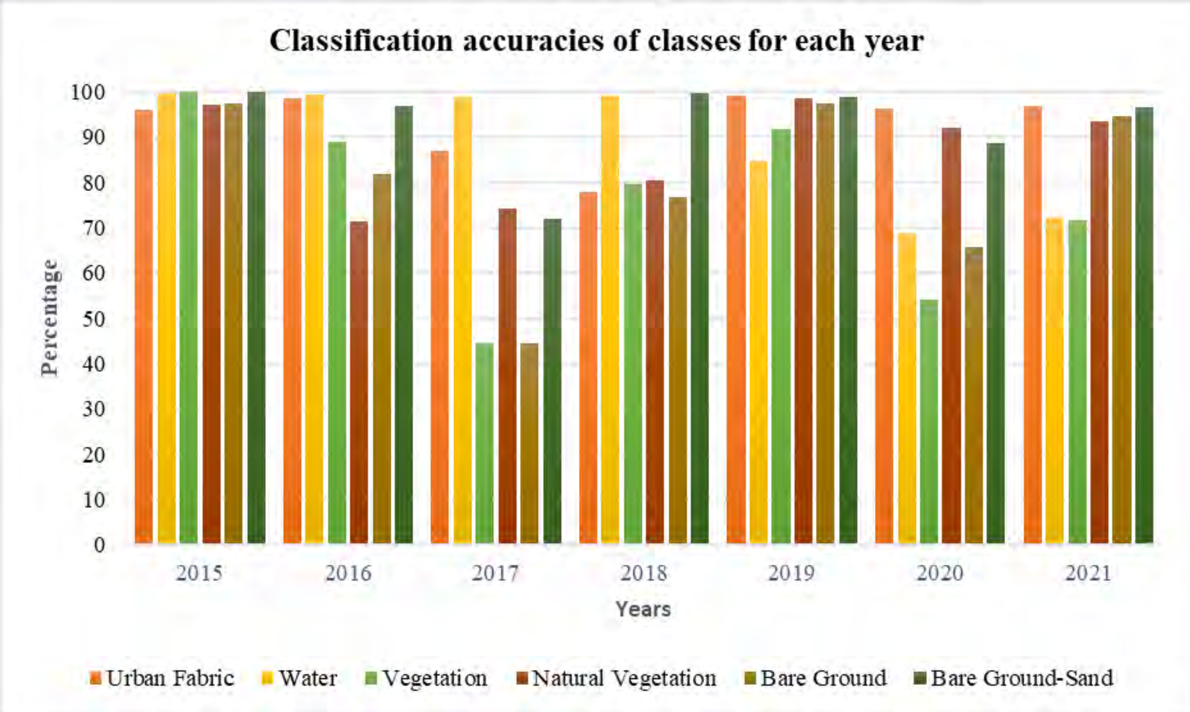


Figure 4.17 Chart indicating the various class accuracies over the 7 year period.

4.4.2 Acreage

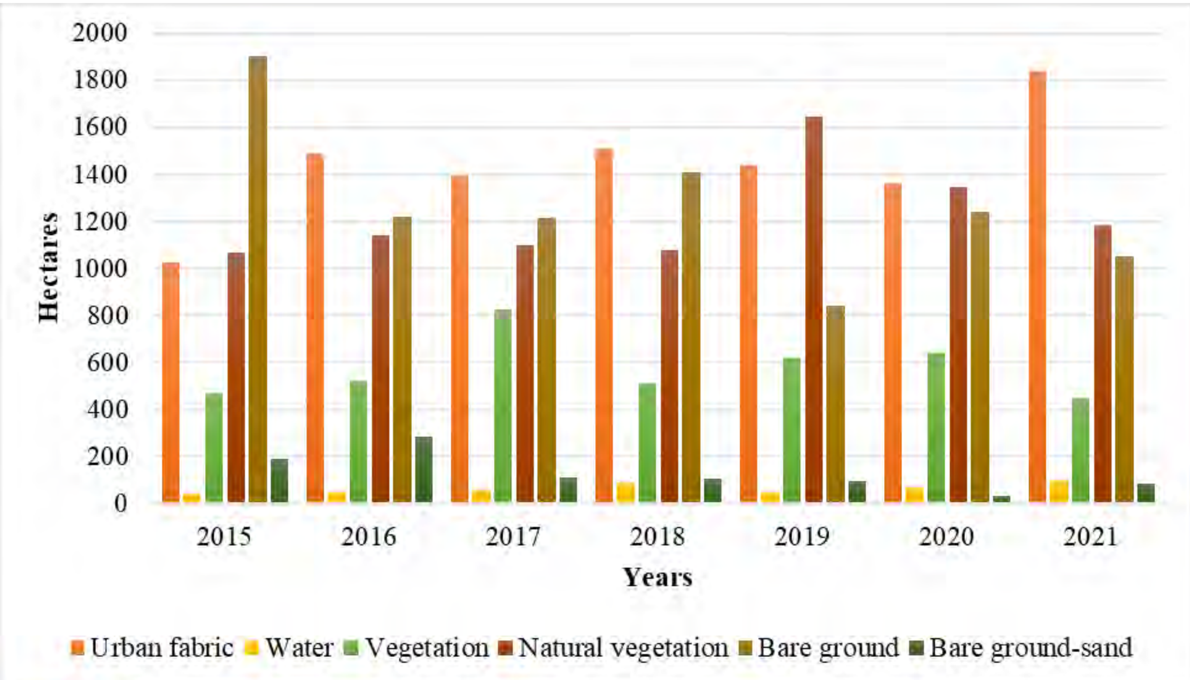


Figure 4.18 Chart indicating the various class acreages over the 7 year period.

4.4.3 Impact of indices

Table 4.3 Overall Accuracy assessment

Year	Classification without indices	Classification with indices
2015	98.4%	98.8%
2016	92.2%	98.9%
2017	70.4%	97.9%
2018	87.0%	97.89%
2019	97.23%	98.8%
2020	80.6%	99.3%
2021	91.8%	99.6%

The table above shows the overall accuracy statistics for classification of layer stacks with indices and layer stacks of only the original bands.

4.5 Change Detection

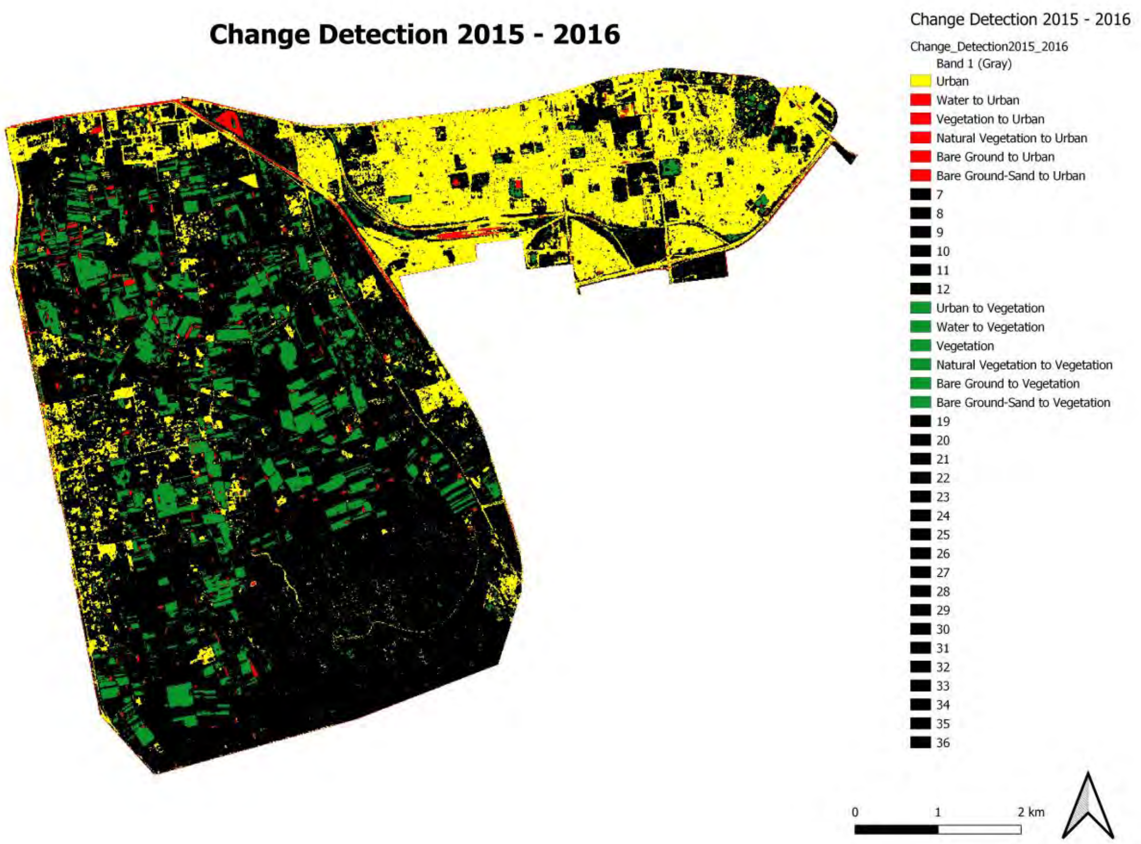


Figure 4.19 Change detection map for years 2015 to 2016.

Emphases were given to urban fabric and vegetation. The colour composite for the maps are as follows: Yellow indicates urban fabric, red is the colour that has changed from the existing feature and green is for vegetation. The black was to map the other features.

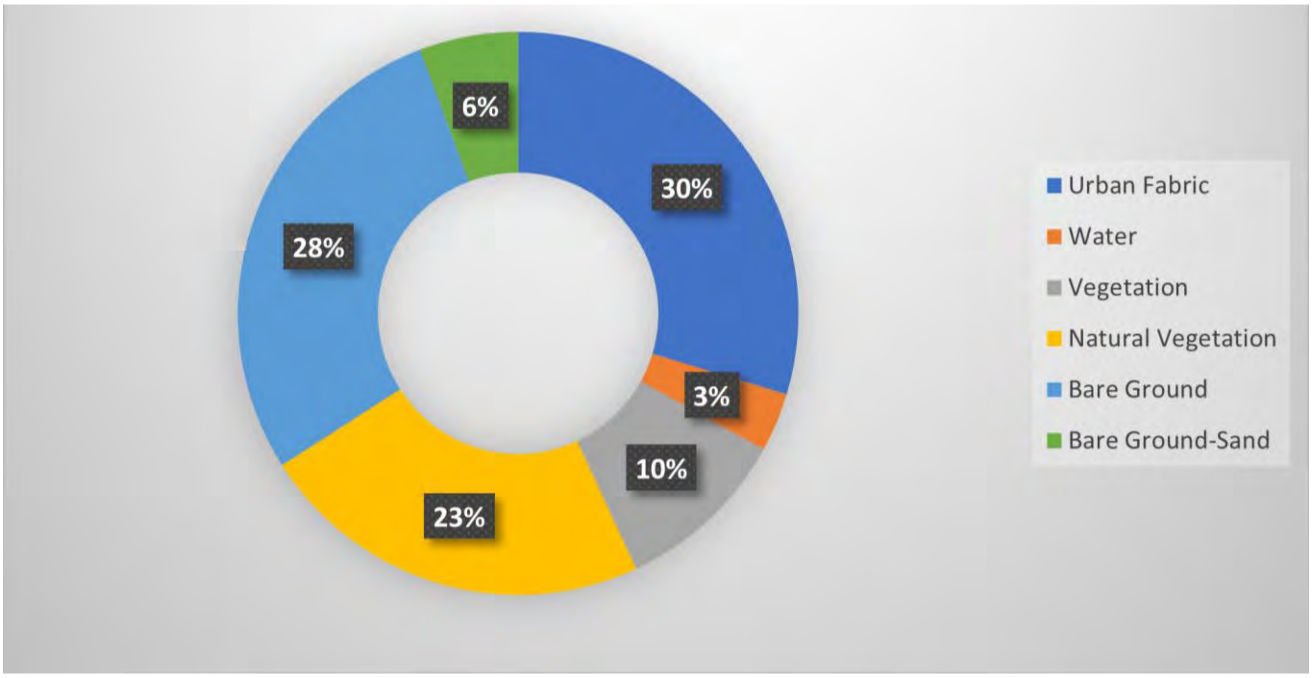


Figure 4.20 Land cover between 2015 and 2016.

The percentages being displayed in the pie chart indicates the amount of changes that occurred.

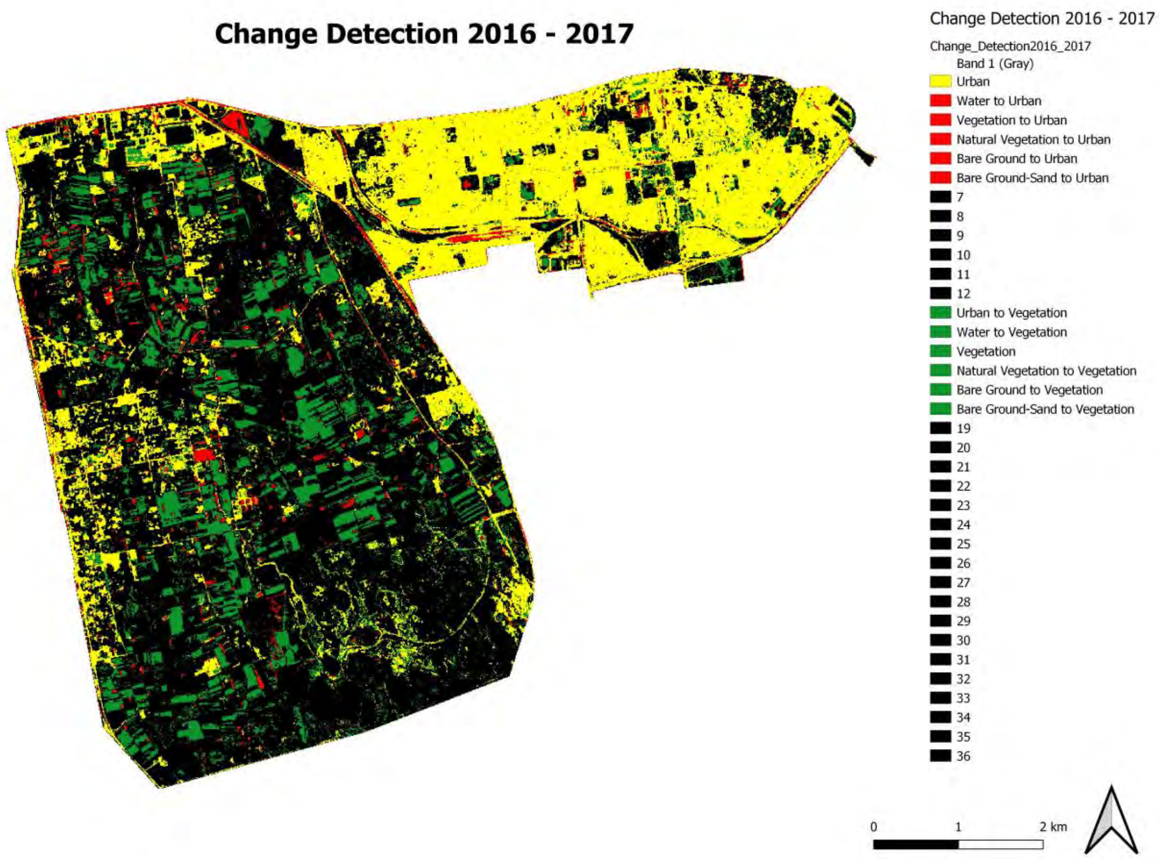


Figure 4.21 Change detection map for years 2016 to 2017. Emphases were given to urban fabric and vegetation.

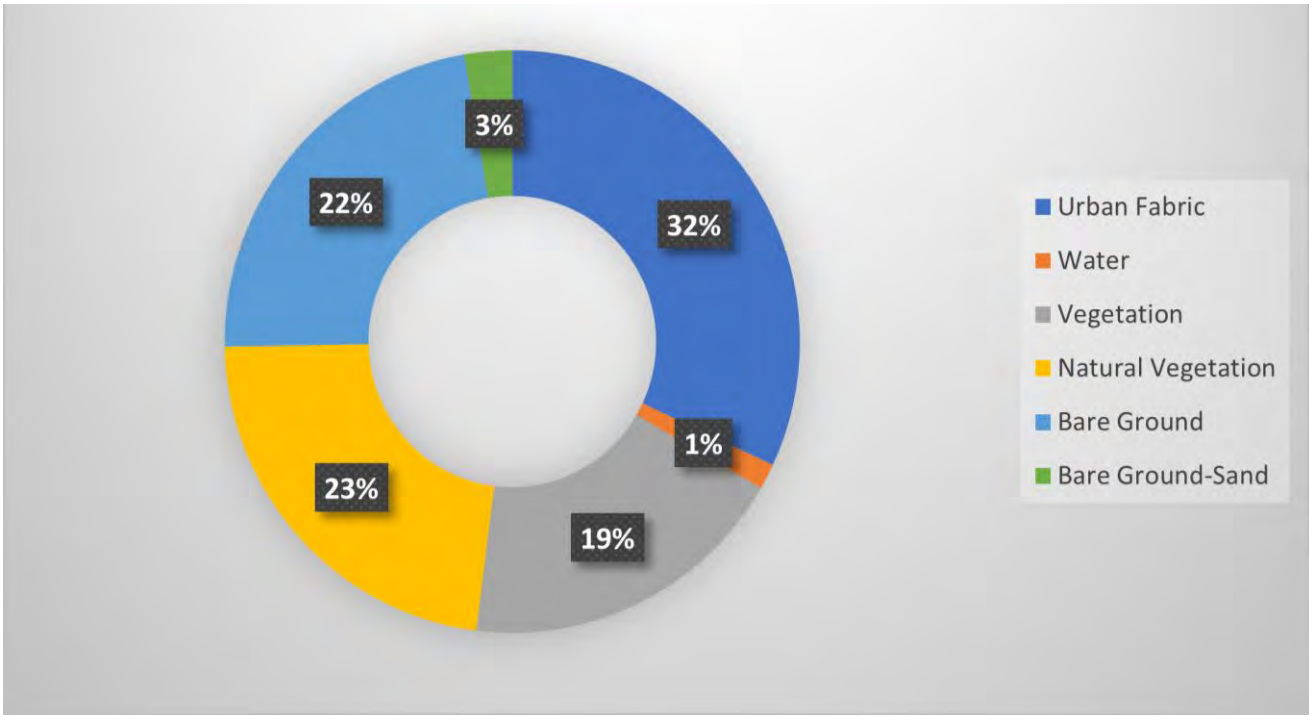


Figure 4.22 Land cover between 2016 and 2017.

The percentages being displayed in the pie chart indicates the amount of changes that occurred.

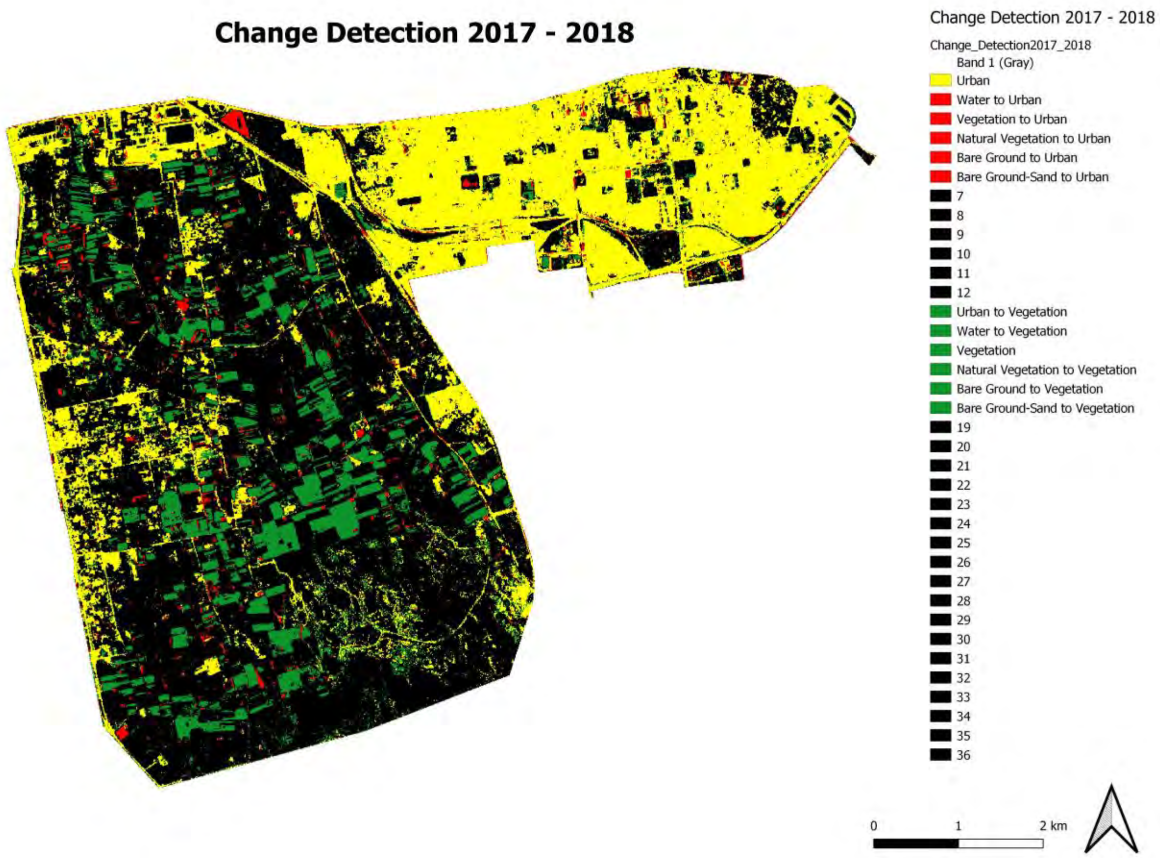


Figure 4.23 Change detection map for years 2017 – 2018. Emphases were given to urban fabric and vegetation.

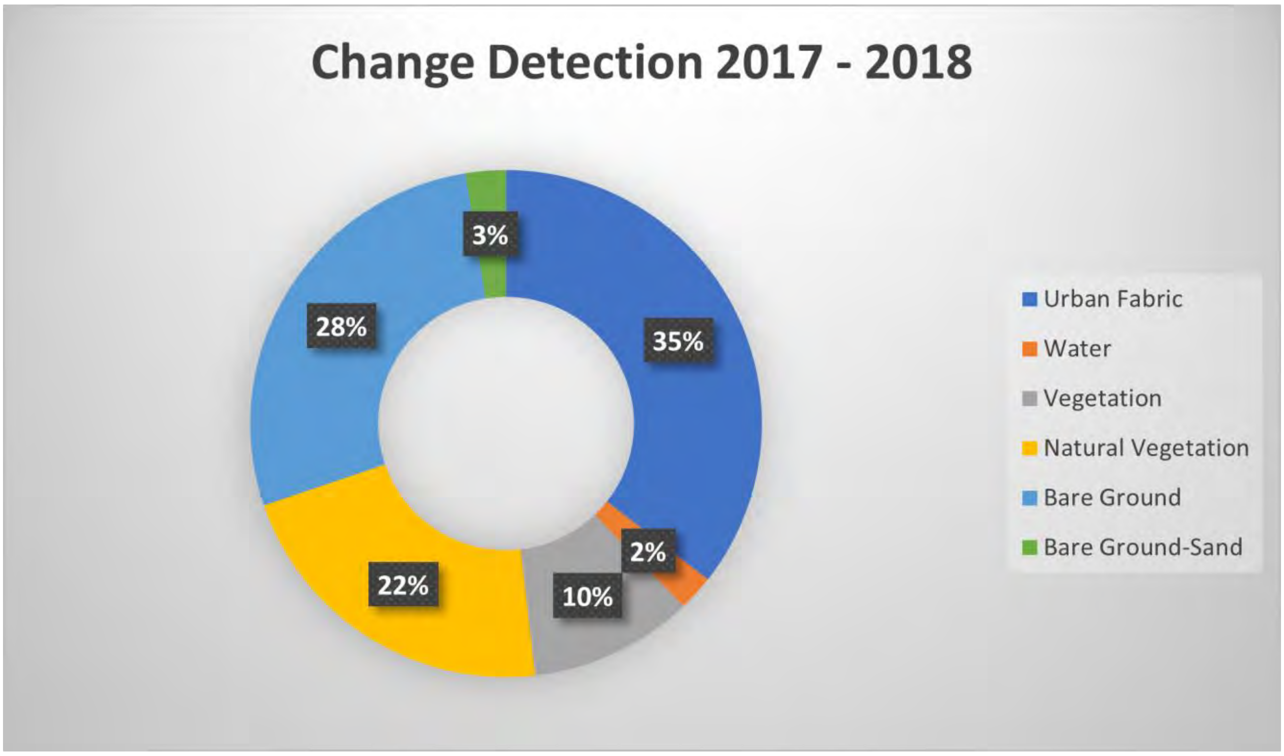


Figure 4.24 Land cover between 2017 and 2018.

The percentages being displayed in the pie chart indicates the amount of changes that occurred.

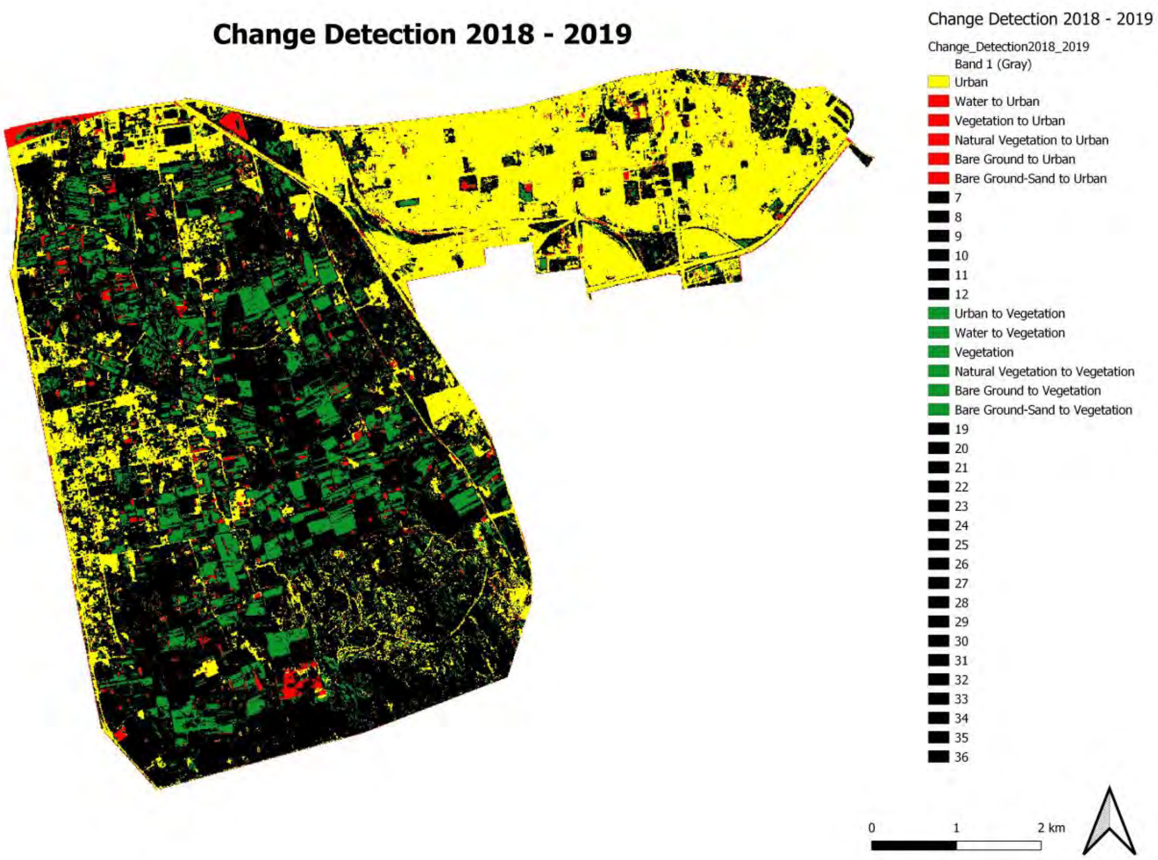


Figure 4.25 Change detection map for years 2018 – 2019. Emphases were given to urban fabric and vegetation.

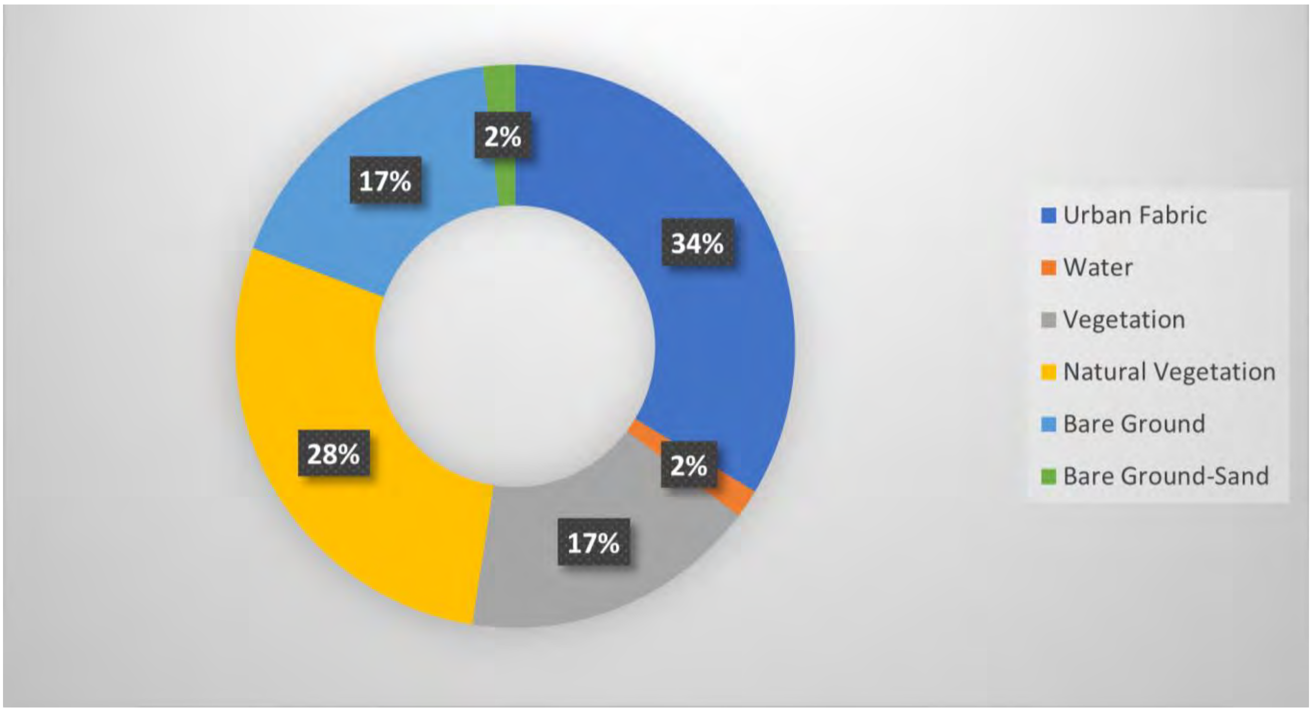


Figure 4.26 Land cover between 2018 and 2019.

The percentages being displayed in the pie chart indicates the amount of changes that occurred.

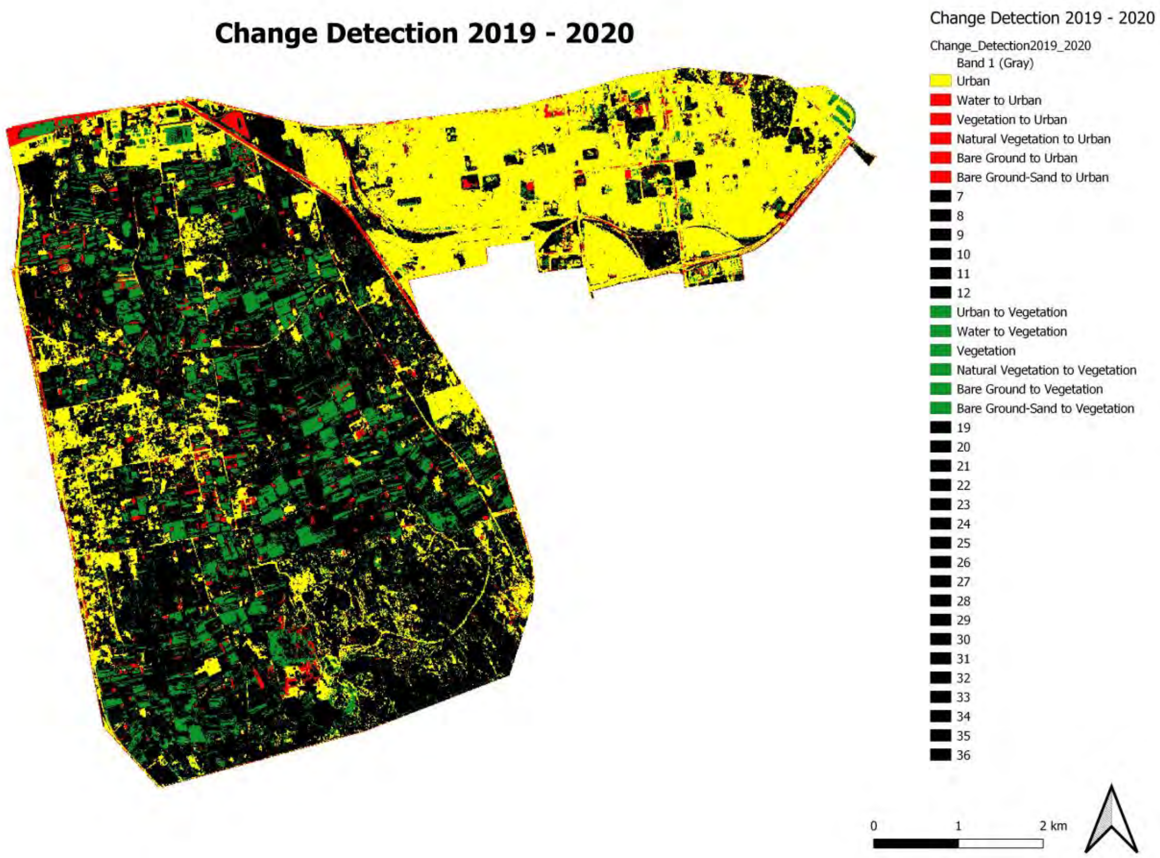


Figure 4.27 Change detection map for years 2019 – 2020. Emphases were given to urban fabric and vegetation.

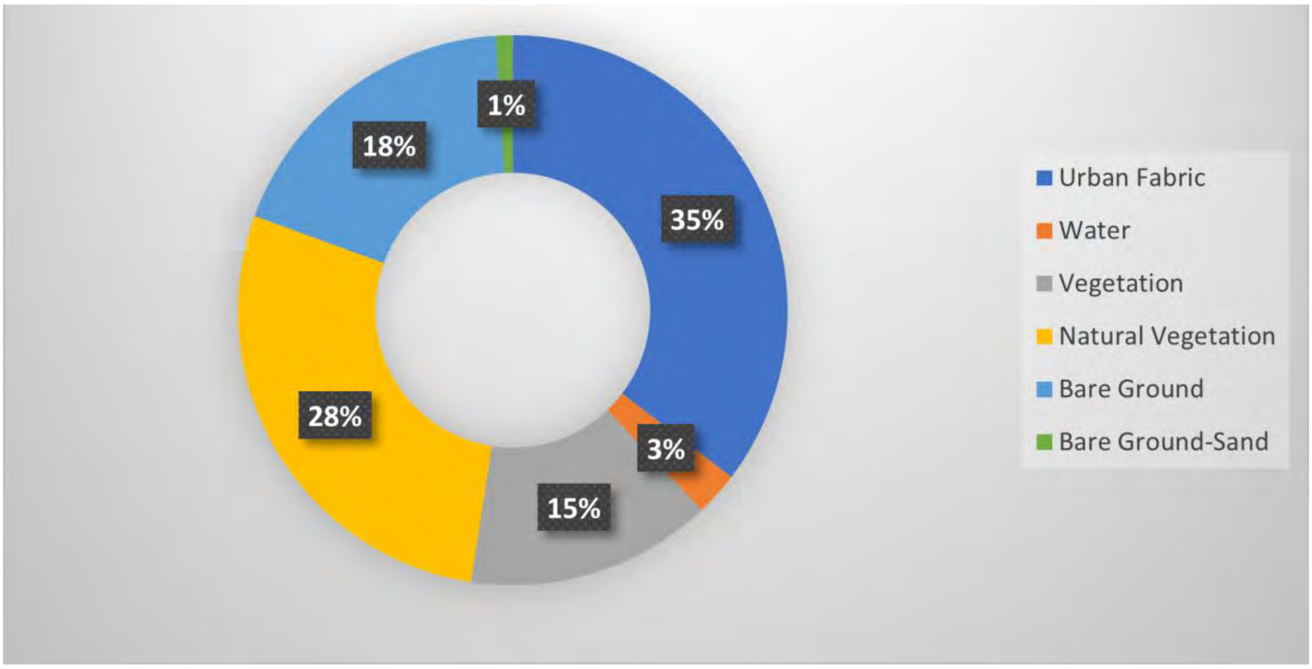


Figure 4.28 Land cover between 2019 and 2020.

The percentages being displayed in the pie chart indicates the amount of changes that occurred.

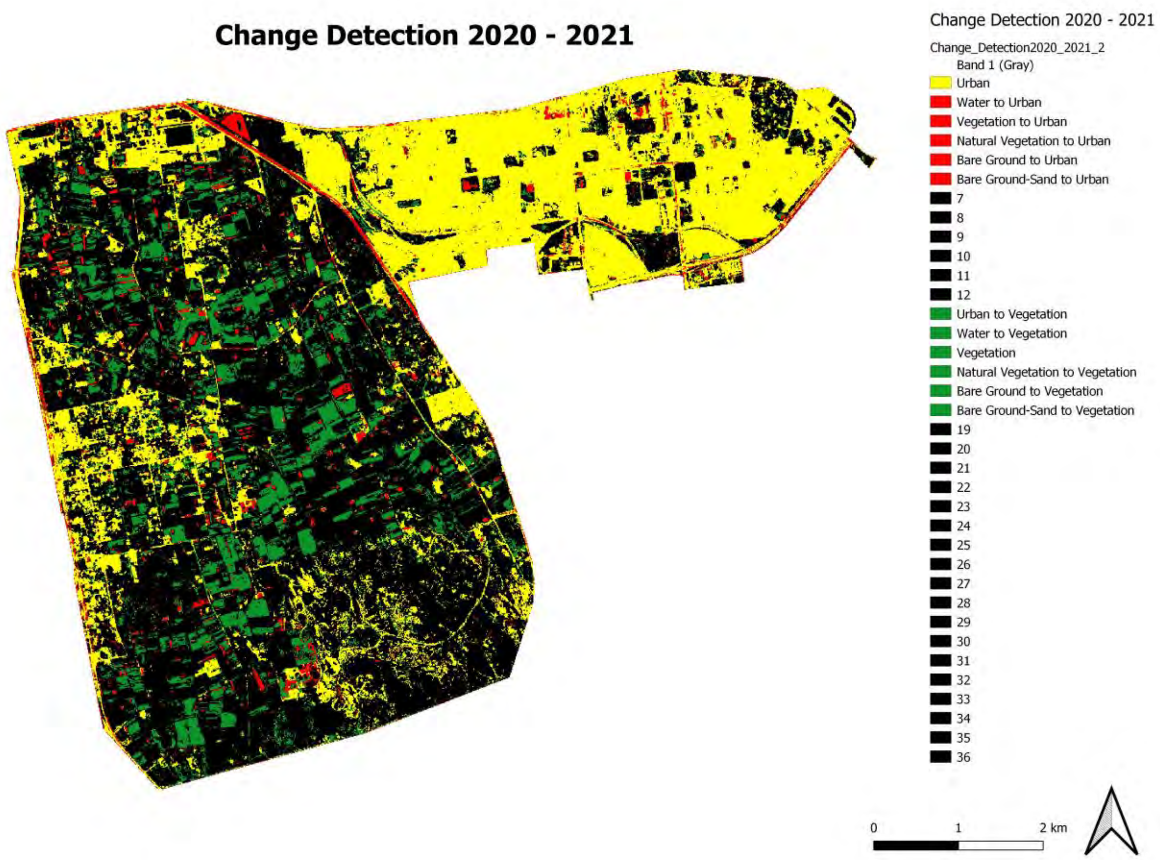


Figure 4.29 Change detection map for years 2020 – 2021. Emphases were given to urban fabric and vegetation.

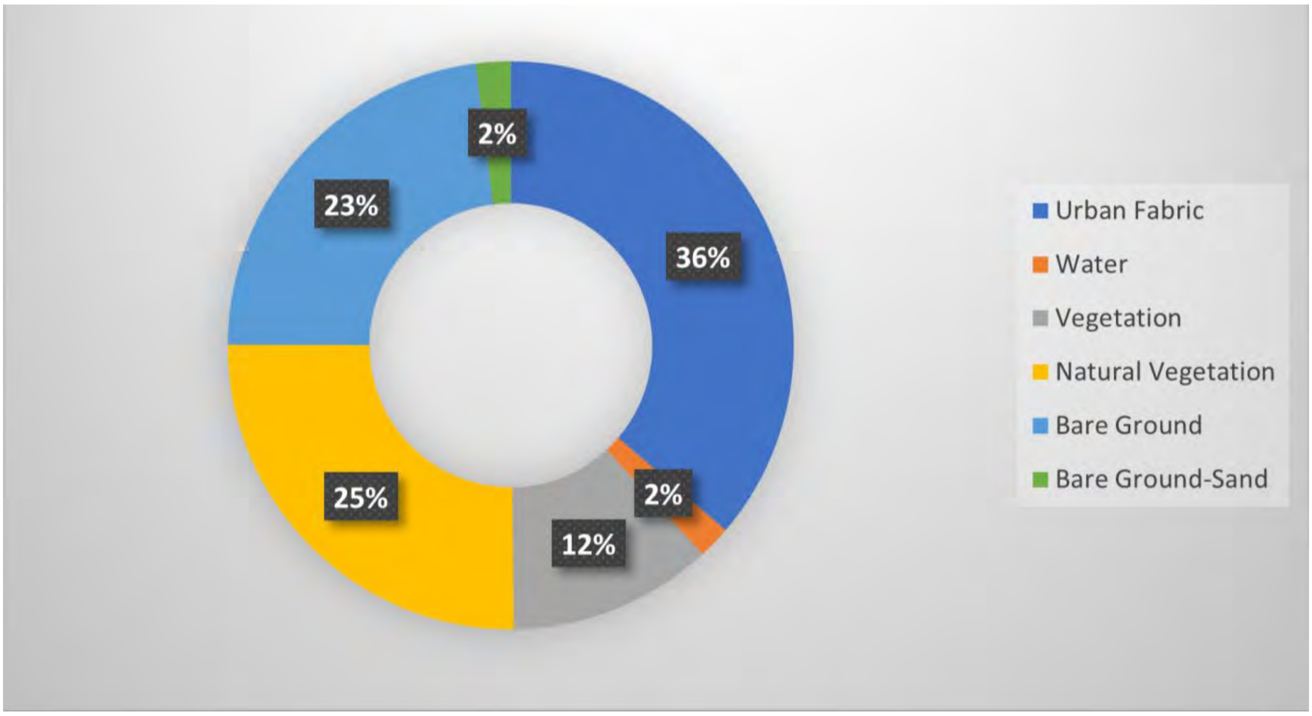


Figure 4.30 Land cover between 2020 and 2021.

The percentages being displayed in the pie chart indicates the amount of changes that occurred.

Chapter 5 Discussion

5.1 Vegetation Indices

It was found that the overall classification accuracy improved with the inclusion of indices for each year. In 2015 the accuracy was found to increase by 0.3%, in 2016 the accuracy improved by 6.6%, in 2017 the accuracy improved by 27.5%, in 2018 the accuracy increased by 10.8%, in 2019 the accuracy increased by 1.6%, in 2020 the accuracy improved by 18.6% and in 2021 the accuracy increased by 7.8%.. Even with variable changes the indices have proven to be an effective inclusion in order increase the accuracy of the initial classification process. The four most relevant indices for the classification of these classes was MSAVI, RDVI, RG and GNDVI. MSAVI for instance addresses the limitation NDVI is applied to high degree of bare soil (references?). RDVI does not operate well in dry areas and suppresses the effects of sun and soil. RG provides pigmentation information on leaves, which is a basic indicator of crop health and is useful to determine the stage of crop growth. GNDVI is used to assess the variability of crop development both in conditions of dense vegetation cover and in conditions of sparse vegetation (Bannari et al., 1995).

5.2 Comparison of classifiers

Of the three classifiers used, two classifiers, that is, Random Forest (RF) and Support Machine Vector (SVM) had similar accuracy statistics. Although the two classifiers showed high accuracies, Random Forest (RF) generally outperformed Support Vector Machines. The results suggest that both the SVM and RF classifiers have the potential to produce high accuracy classifications for mapping the agricultural and urban edge. Evidently, the study area affects the performance of the classifiers. For instance, Lefulebe et al (2022) found that KNN outperformed RF, SVM and NB in their urban forests study. Also, Mazarire et al (2020) found SVM outperformed RF in mapping heterogenous agricultural landscapes. However, studies such as Colette el al (2020) showed similar accuracies for SVM and RF.

5.3 Classification

5.3.1 Classification maps

With the results of the classifications done over the 7-year period an increase in land use change can be seen in the northern, eastern and western parts of the PHA. Parts of the vegetation areas is becoming bare ground as well, presenting a need to monitor and protect the areas left for agricultural purposes. The acreages figure 4.18 shows a steady growth in urban fabric over the 7 year period and reaches it's highest point in 2021. Vegetation has shown an unstable trend over the 7 years ending with it's lowest growth point in 2021.

5.3.2 Accuracies per class

As per the accuracies of vegetation and urban classes shows that accuracy levels for urban fabric maintained high values between 80% and 90%, while vegetation showed poor accuracies in 2017 and 2020 (Refer to table 4.2). Nonetheless, the high accuracies of the urban fabric classification along with the underlying land cover pie charts show the increasing urbanisation of the PHA. The accuracies assessments for the classifications with and without indices has indicated a steady and stable percentage for the accuracies with indices proving it to be a valuable addition to the process.

The pie charts in section 4.5 also show a steady increase in urban land cover within the PHA. There were 14 maps in total created with 7 having indices included with the classified images. These maps were the steppingstone for the inclusion of change detection maps which ultimately produced the final visual results.

5.4 Change detection

In figure 4.19 and 4.21 changes can be seen southern part of the PHA showing an increase in urban fabric as well as vegetation. In figures 4.23 and 4.25 visual changes can be seen occurring to the western part of the PHA indicating an increasing in urban fabric. In figures 4.27 and 4.29 urban fabric is shown more predominantly over the western and southern parts of the PHA. The changes can be seen visually in figures 4.18 to 4.29. The urban fabric has slowly filtered onto the vegetation area and has disturbed the agricultural sector. These maps can determine visually how the area has changed and what impact the urbanization has had on the agricultural sector. Figure 4.19 shows that the changes generally occurred in close proximity to main roads. Changes has happened over the most parts of the PHA, some being slight changes while others showing a bit more. Overall, the changes can be visually seen on the maps. The 7 years has shown an increase within the urban fabric sector.

5.5 Sensors

This study found that PlanetScope data is a viable option for mapping urban areas. The resolution of 3.7 meters makes image interpretation possible and the spectral resolution also led to high accuracies of classification and it supports 20 vegetation indices. PlanetScope has proven to display good results for urban mapping as well as vegetation mapping. The resolution and availability of aerial images makes work for researchers easier. The sensors that can create different bands are Red, Green, Blue and Near

Infrared. These bands can be improve agriculture and can produce a clearer view of what's happening on the ground. The bands can also assist with tracking urban growth in an area. PlanetScope can assist in tracking the pattern of growth for urban structures over a period of time. Spectral resolution helps characterize samples based on fine wavelengths of the electromagnetic spectrum. Spatial resolution focuses on measuring image quality while temporal resolution looks at the amount time required to visit and obtain data for a specified location.

Chapter 6 Conclusions and recommendations

6.1 Conclusions

6.1.1 Vegetation indices

With the assistance and input of vegetation indices within the machine learning realm, the accuracy and true reflection of a land use for this paper or similar outcomes have proven to be vital. The indices have enhanced the efficiency of the classifiers and made it a more unique and distinct version of providing accurate outcomes. Indices improve accuracy (Overall and Kappa) as well as producer vs user accuracies. All the indices has shown an improvement in the accuracy levels to a degree, but the four top indices were MSAVI, RDVI, RG and GNDVI which showed the most significant changes to the accuracies.

6.1.2 Classifiers

The study found that Random Forest outperformed Support Vector Machines and K Nearest Neighbour in the classifications. Nonetheless, SVM and KNN also performed well. Random Forest edged out Support Vector Machine between 1% – 18% over the 7 year period. Compared to findings from (Musungu and Mkhize 2019) the results has shown similar results, although over the last 6 years urban growth has increased slightly and vegetation has shown a decrease over this period. Other studies have used maximum likelihood, minimum distance and spectral angle mapping (Musungu and Mkhize 2019). Although this study showed similar results for vegetation, the later and updated imagery from PlanetScope, with the addition of vegetation indices, showed an improvement to the accuracies and also indicated an increase in urban fabric for this study.

6.1.3 Land cover

The study found that there is increasing urbanisation in the PHA. The urban cover between the year of 2015 – 2021 had increased from 30% of the area land cover to 36% of the land cover. There was a decrease in vegetation in the year 2021 from the year 2015. With the assistance of the vegetation indices changes could be indicated more precisely. The land cover change for Musungu and Mkhize 2019 has shown a decrease for urban fabric and for vegetation, however with the later results from this study has shown an increase in urban fabric and a decrease in vegetation cover.

6.1.4 Sensors

The study found that the constellation of PlanetScope satellites provide sufficient data for urban land cover mapping. The spatial resolution of 3.7 meters coupled with the four available spectral bands Red, Green Blue and NIR facilitated the creation of several urban and vegetation indices. Studies from Musungu and Mkhize 2019 had used Landsat imagery which produced good results, however later imagery with updated sensors have provided an increase in accuracies.

6.1.5 Change Detection

The change detection maps clearly indicate a pattern whereby urban fabric has increased over the time period and the vegetation areas has shown an unstable period over the years. This mapping exercise has been a vital step to display visual changes within the area in question. The change detection method indicated that there was a visual change in urban structures over the 7-year period within the PHA and that vegetation has decreased from the start of 2015. Other studies have shown a decrease in both urban fabric and vegetation whereas this study has shown an increase in urban fabric. Other features such as bare ground has also been seen to increase over the years due to vegetation decreasing. This study and method have definitely proven to be effective when change detection is required for an area.

6.2 Recommendations

6.2.1 Sensors

The study found that the constellation of PlanetScope satellites provide sufficient data for urban land cover mapping. Thus, it is recommended that other studies test the capabilities of PlantScope in urban mapping. The latest constellation has more bands and has almost real-time data available. With easier access, finer scales and better coverage, mapping urban features and agriculture would present improved results for future studies. Sentinel imagery also produces high resolution for vegetation and water. This satellite would not have produced the accurate results for urban fabric as this study has shown with PlanetScope.

6.2.2 Indices

Secondly, it was found that adding vegetation indices improved the performance of all the classifiers. Thus, it is recommended that optimised vegetation indices are included in the classification of urban scenes. Since the PlanetScope satellites have been updated, it provides more bands, scholars can assess an even wider range of indices.

6.2.3 Frequency

This study has proven that satellite imagery combined with remote sensing techniques can prove valuable tool to an organization or public entity. This process and outcomes can serve as a valid decision-making tool for projects, service delivery etc in an area. It can also look at using annual average of indices instead of singel dates in order to mitigate against any errors. Using annual composites and indices gives a better understanding of the land cover.

References

Abderrazak., B, Morin., D, Bonn., F and Huete., A. (1996). *A review of vegetation indices*. Remote Sensing Reviews. 13. 95-120. 10.1080/02757259509532298.

Ahmed, O.S., Shemrock, A., Chabot, D., Dillon, C., Williams, G., Wasson, R. and Franklin, S.E., 2017. Hierarchical land cover and vegetation classification using multispectral data acquired from an unmanned aerial vehicle. *International journal of remote sensing*, *38*(8-10), pp.2037-2052.

Aldoski, J & Mansor, S, Shafri, H & Shafri, M. 2013. *Image Classification in Remote Sensing*. 3.

Agjee, N.H, Ismail, R & Mutuanga, O. 2016. *Identifying relevant hyperspectral bands using Boruta*: a temporal analysis of water hyacinth biocontrol.

Andrade., J, Kenneth., G. Cassman., JI., Rattalino., E, Fahmuddin., A, Abdullahi., B, Nanyan., D, Patricio., G. 2022. *Impact of urbanization trends on production of key staple crops.*

Asokan., A and Anitha., J - Earth Science Informatics, 2019. *Change detection techniques for remote sensing applications*: A survey.

Ardila, J.P., Bijker, W., Tolpekin, V.A. and Stein, A., 2012. *Multitemporal change detection of urban trees using localized region-based active contours in VHR images*. Remote Sensing of Environment, 124: 413-426.

Arjasakusuma, S.; Swahyu Kusuma, S.; Phinn, S. 2020. Evaluating Variable Selection and Machine Learning Algorithms for Estimating Forest Heights by Combining Lidar and Hyperspectral Data. ISPRS International Journal of Geo-Information 9 (9): 507. https://doi.org/10.3390/ijgi9090507

Bannari, A., Morin, D., Bonn, F. and Huete, A., 1995. *A review of vegetation indices*. Remote Sensing Reviews, 13(1-2): 95-120.

Bannari, A., Asalhi, H. and Teillet, P.M., 2002, June. *Transformed difference vegetation index (TDVI) for vegetation cover mapping*. In IEEE International geoscience and remote sensing symposium (Vol. 5, pp. 3053-3055). IEEE.

Basak., R and Haque., I. 2017. *Land cover change detection using GIS and remote sensing techniques*: A spatio-temporal study on Tanguar Haor, Sunamganj, Bangladesh

Berhane, T.M., Lane, C.R., Wu, Q., Autrey, B.C., Anenkhonov, O.A., Chepinoga, V.V. and Liu, H., 2018. Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. Remote Sensing, 10(4): 580

Breiman, L. Random forests. *Mach. Learn.* 2001, 45, 5–32.

Bhandari, Abhishta & Ibrahim, Muhammad & Sharma, Chinmay & Liong, Rebecca & Gustafson, Sonja & Prior, Marita. (2021). *CT-based radiomics for differentiating renal tumours: a systematic review.* Abdominal Radiology. 46. 1-12. 10.1007/s00261-020-02832-9.

Broge, N.H. and Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote sensing of environment*, *76*(2), pp.156-172.

Bugnot, A.B., Hose, G.C., Walsh, C.J., Floerl, O., French, K., Dafforn, K.A., Hahs, A.K., 2019. Urban impacts across realms: making the case for inter-realm monitoring and management. Sci. Total Environ. 648, 711–719. https://doi.org/10.1016/j. scitotenv.2018.08.134.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. and Lopez, A., 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, pp.189-215.

Chailan, R. (2015). Application of Scientific Computing and Statistical Analysis to address Coastal Hazards.

Chang, Z., Du, Z., Zhang, F., Huang, F., Chen, J., Li, W. & Guo, Z. 2020. Landslide Susceptibility Prediction Based on Remote Sensing Images and GIS: Comparisons of Supervised and Unsupervised Machine Learning Models. Remote Sensing, 12(3): 502. https://www.mdpi.com/2072-4292/12/3/502.

Chen, J.M., 1996. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, *22*(3), pp.229-242.

Chen, J., M. Lu, X. Chen, J. Chen and L. Chen, 2013. A spectral gradient difference based approach for land cover change detection. ISPRS J. Photogramm., 85: 1-12.

Clevers, J.G.P.W., 1989. Application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote Sensing of Environment*, *29*(1), pp.25-37.

Coffey., R. Michigan State University Extension - January 18, 2013. *The difference between "land use" and "land cover"*.

Colwell, J. E. (1974). Grass canopy bidirectional reflectance. *Proceedings of 9th International Symposium on Remote Sensing of the Environment*, Ann Arbor, USA, 1061-1065.

Coops, NC, White, JC, Johnson, M, et al. (2006). Assessment of Quick Bird high spatial resolution imagery detect red attack damage due to mountain pine beetle infestation. Remote Sensing Environment 103: 67-80.

Cui, X., Fang, C., Liu, H. & Liu, X., 2019. Assessing sustainability of urbanization by a coordinated development index for an urbanization-resources-environment complex system: a case study of Jing-Jin-Ji region, China.

Da Penha Pacheco, A, da Silva Junior J.A, Ruiz-Armenteros A.M & 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): 1345.

Das, S. and Angadi, D.P., 2022. Land use land cover change detection and monitoring of urban growth using remote sensing and GIS techniques: A micro-level study. *GeoJournal*, *87*(3), pp.2101-2123.

de Castro Paes, É., Veloso, G.V., da Fonseca, A.A., Fernandes-Filho, E.I., Fontes, M.P.F. and Soares, E.M.B., 2022. Predictive modeling of contents of potentially toxic elements using morphometric data, proximal sensing, and chemical and physical properties of soils under mining influence. *Science of The Total Environment*, *817*, p.152972.

Demir, B., F. Bovolo and L. Bruzzone, 2012. Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification. IEEE T. Geoscience. Remote, 50(5): 1930-1941.

Deng, X., Huang, J., Rozelle, S., Zhang, J. and Li, Z., 2015. Impact of urbanization on cultivated land changes in China. *Land use policy*, *45*, pp.1-7.

De Winter, J.C., Gosling, S.D. and Potter, J., 2016. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods*, *21*(3), p.273.

Dr. Alexandru-Ionuţ Petrişor, PhD, Habil, Dr. Raffaele Pelorosso, PhD Dr. Bogdan Zagajewski, PhD, Habil 2020), Special Issue: '*CORINE Land Cover System: Limits and Challenges for Territorial Studies and Planning*'.

Du, J., Fu, Q., Fang, S., Wu, J., He, P. and Quan, Z., 2019. Effects of rapid urbanization on vegetation cover in the metropolises of China over the last four decades. Ecological Indicators, 107, p.105458.

Dudani, S.A., 1976. The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*, (4), pp.325-327.

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X. and Xiang, Y., 2018. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy conversion and management*, *164*, pp.102-111.

Feng, Q. Liu, J. Gong, J. 2015. UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis. Remote Sensing. 7, 1074-1094.

Fernandes, K., 2002. Urban development and new towns in the Third World, lessons from the new Bombay experience. Habitat International. 26: 134–135.

Gao, B.C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, *58*(3), pp.257-266.

García-Nieto, A.P., Geijzendorffer, I.R., Baró, F., Roche, P.K., Bondeau, A. and Cramer, W., 2018. Impacts of urbanization around Mediterranean cities: Changes in ecosystem service supply. *Ecological Indicators*, *91*, pp.589-606.

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E. and Kalogirou, S., 2021. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, *36*(2), pp.121-136.

Gitelson, A.A., 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of plant physiology*, *161*(2), pp.165-173.

Goldblatt, R., Stuhlmacher, M.F., Tellman, B., Clinton, N., Hanson, G., Georgescu, M., Wang, C., Serrano-Candela, F., Khandelwal, A.K., Cheng, W.-H. & Balling, R.C. 2018. *Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover*. Remote Sensing of Environment, 205: 253–275. https://linkinghub.elsevier.com/retrieve/pii/S0034425717305758.

Guha, S., Govil, H. & Diwan, P. 2019. Analytical study of seasonal variability in land surface temperature with normalized difference vegetation index, normalized difference water index, normalized difference built-up index, and normalized multiband drought index. Journal of Applied Remote Sensing, 13(2): 024518.

Hacker, K.P., Seto, K.C., Costa, F., Corburn, J., Reis, M.G., Ko, A.I. and Diuk-Wasser, M.A., 2013. *Urban slum structure: integrating socioeconomic and land cover data to model slum evolution in Salvador, Brazil*. International Journal of Health Geographics, 12(1): 1-12.

Haque, M.I. and Basak, R., 2017. Land cover change detection using GIS and remote sensing techniques: A spatio-temporal study on Tanguar Haor, Sunamganj, Bangladesh. *The Egyptian Journal of Remote Sensing and Space Science*, *20*(2), pp.251-263.

Huang, Y. and Zhao, L., 2018. Review on landslide susceptibility mapping using support vector machines. *Catena*, *165*, pp.520-529.

Huete, A.R., Liu, H., De Lira, G.R., Batchily, K. and Escadafal, R., 1994, August. A soil color index to adjust for soil and litter noise in vegetation index imagery of arid regions. In *Proceedings of IGARSS'94-1994 IEEE International Geoscience and Remote Sensing Symposium* (Vol. 2, pp. 1042-1043). IEEE.

Indego. 2018. Building the City of Cape Town's Resilience and Adding to Regional Competitiveness.

Jie, N.F., Zhu, M.H., Ma, X.Y., Osuch, E.A., Wammes, M., Théberge, J., Li, H.D., Zhang, Y., Jiang, T.Z., Sui, J. and Calhoun, V.D., 2015. Discriminating bipolar disorder from major depression based on SVM-FoBa: efficient feature selection with multimodal brain imaging data. *IEEE transactions on autonomous mental development*, *7*(4), pp.320-331.

Kaplan, G. and Avdan, U., 2018. Monthly Analysis of Wetlands Dynamics Using Remote Sensing Data. ISPRS International Journal of Geo-Information, [online] 7(10), p.411. https://doi.org/10.3390/ijgi7100411.

Karasiak, N. 2017. Dzetsaka: *Classification tool*. Available: https://plugins.qgis.org/plugins/dzetsaka/.

Kim, D.Y., Thomas, V., Olson, J., Williams, M. and Clements, N., 2013. Statistical trend and change-point analysis of land-cover-change patterns in East Africa. International Journal of Remote Sensing, 34(19): 6636-6650

Kohli , D., Sliuzas, R., Kerle, N. & Stein, A. 2012. *An ontology of slums for image-based classification. Computers. Environment and Urban Systems*, 362: 154e163. http://dx.doi.org/10.1016/j.compenvurbsys.2011.11.001.

Kriegler, F., Malila, W., Nalepka, R., & Richardson, W. (1969). *Preprocessing transformations and their effect on multispectral recognition. Proceedings of the 6th International Symposium on Remote Sensing of Environment. Ann Arbor, MI*: University of Michigan, 97-131.

Kuffer, M. & Barrosb, J. 2011. Urban morphology of unplanned settlements: The use of spatial metrics in VHR remotely sensed images. Procedia Environmental Sciences, 7:152-157

Kursa, M. B. and Rudnicki, W. R. 2010. *Feature Selection with the Boruta Package, Journal of Statistical Software*, 36(11), pp. 1–13. doi: 10.18637/jss.v036.i11.

Laliberte, A.S, Browning, D.M and Rango A. 2012. A comparison of three feature selection methods for object-based classification of sub-decimeter resolution UltraCam-L imagery, Int. J. Appl. Earth Observation. Geoinformation., vol. 15, pp. 70-78.

Lee, L.H., Wan, C.H., Rajkumar, R. and Isa, D., 2012. An enhanced support vector machine classification framework by using Euclidean distance function for text document categorization. *Applied Intelligence*, *37*, pp.80-99.

Lefulebe, B.E., Van der Walt, A. & Xulu, S. 2022. *Fine-scale classification of urban land use and land cover with PlanetScope imagery and machine learning strategies in the City of Cape Town, South Africa*. Sustainability, 14(15): 9139. https://www.mdpi.com/2071-1050/14/15/9139.

Leroux, L., Congedo, L., Bellón, B., Gaetano, R. & Bégué, A. 2018. Land cover mapping using Sentinel-2 Images and the Semi-Automatic Classification Plugin: A Northern Burkina Faso case study. In J. Smith (Ed) QGIS and Applications in Agriculture and Forest. Hoboken, NJ, USA: John Wiley & Sons, Inc. 119–151. DOI: 10.1002/9781119457107.ch4: 27-35.

Li, K., Wan, G., Cheng, G., Meng, L. and Han, J., 2020. Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 159: 296-307.

Liang, L., Wang, Z. and Li, J., 2019. The effect of urbanization on environmental pollution in rapidly developing urban agglomerations. *Journal of cleaner production*, *237*, p.117649.

Ling, Z.K., Liu, W., Niu, C.Y., Li, R.S., Hu, Q. and Zang, Y.Q., 2021, December. SAR Image Change Detection Based on Geometric Mean Operator and Extreme Learning Machine. In *2021 CIE International Conference on Radar (Radar)* (pp. 962-965). IEEE.

Louargant, M., Jones, G., Faroux, R., Paoli, J.-N., Maillot, T., Gée, C. & Villette, S. 2018. Unsupervised classification algorithm for early weed detection in row-crops by combining spatial and spectral information. Remote Sensing, 10(5): 761. http://www.mdpi.com/2072-4292/10/5/761.

Lyon, J.G., Yuan, D., Lunetta, R.S. and Elvidge, C.D., 1998. A change detection experiment using vegetation indices. *Photogrammetric engineering and remote sensing*, *64*(2), pp.143-150.

Magidi, J. and Ahmed, F., 2019. Assessing urban sprawl using remote sensing and landscape metrics: A case study of City of Tshwane, South Africa (1984–2015). *The Egyptian Journal of Remote Sensing and Space Science*, *22*(3), pp.335-346.

Mishra, S., Shrivastava, P. and Dhurvey, P., 2017. Change detection techniques in remote sensing: A review. *International Journal of Wireless and Mobile communication for Industrial systems*, *4*(1), pp.1-8.

Mou, L., Lu, X., Li, X. and Zhu, X.X., 2020. Nonlocal graph convolutional networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(12), pp.8246-8257.

Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J. and Wilhelm, V.L., 2017. Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery. *European Journal of Remote Sensing*, *50*(1), pp.564-576.

Perry Jr, C.R. and Lautenschlager, L.F., 1984. Functional equivalence of spectral vegetation indices. *Remote sensing of environment*, *14*(1-3), pp.169-182.

Philippi Environment Development Initiative. June 2017. *PEDI Newsletter. Cape Town* (Accessed: 25/03/2021).

Pham, H. M. & Yamaguchi, Y. 2011. Urban growth and change analysis using remote sensing and spatial metrics from 1975 to 2003 for Hanoi, Vietnam. International Journal of Remote Sensing, 32(7): 1901e1915.

Pham, V.C., Pham, T.T.H., Tong, T.H.A., Nguyen, T.T.H. and Pham, N.H., 2015. The conversion of agricultural land in the peri-urban areas of Hanoi (Vietnam): patterns in space and time. *Journal of Land Use Science*, *10*(2), pp.224-242.

Pham Thi, N., Kappas, M. and Faust, H., 2021. Impacts of agricultural land acquisition for urbanization on agricultural activities of affected households: a case study in Huong Thuy Town, Thua Thien Hue Province, Vietnam. *Sustainability*, *13*(15), p.8559.

Philippi Horticultural Area Summit. 2017, September 12 - 14. Cape Town. https://www.researchgate.net/publication/262685456_Philippi_Horticultural_Area_A_City_asset_or_potential_development_node_A_report_commissioned_by_Rooftops_Canada_Foundation_Inc_-_Foundation_Abri_International_in_partnership_with_the_African_Food_Secu#pf19 (Accessed: 08/08/2020).

PlanetScope, https://earth.esa.int/eogateway/missions/planetscope.

Puliti, S., Talbot, B. and Astrup, R., 2018. Tree-stump detection, segmentation, classification, and measurement using unmanned aerial vehicle (UAV) imagery. *Forests*, *9*(3), p.102.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS journal of photogrammetry and remote sensing*, *67*, pp.93-104.

Samaniego, L. and Schulz, K., 2009. Supervised classification of agricultural land cover using a modified k-NN technique (MNN) and Landsat remote sensing imagery. *Remote Sensing*, *1*(4), pp.875-895.

Sanesi, G., Giannico, V., Elia, M. & Lafortezza, R. 2019. *Remote sensing of urban forests. Remote Sensing*, 11(20):9–13. DOI: 10.3390/rs11202383.

Sejati, A.W., Buchori, I. & Rudiarto, I. 2019. The spatio-temporal trends of urban growth and surface urban heat islands over two decades in the Semarang Metropolitan Region. Sustainable Cities and Society, 46(July 2018):101432. DOI: 10.1016/j.scs.2019.101432.

Sejati, A.W., Buchori, I., Kurniawati, S., Brana, Y.C. & Fariha, T.I. 2020. Quantifying the impact of industrialization on blue carbon storage in the coastal area of Metropolitan Semarang, Indonesia. Applied Geography, 124:102319. DOI: 10.1016/j.apgeog.2020.102319.

Seto, K.C., Güneralp, B. and Hutyra, L.R., 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proceedings of the National Academy of Sciences, 109(40): 16083-16088.

Seto, K.C., Reenberg, A., Boone, C.G., Fragkias, M., Haase, D., Langanke, T., Marcotullio, P., Munroe, D.K., Olah, B. and Simon, D. 2012. *Urban land teleconnections and sustainability. Proceedings of the National Academy of Sciences of the United States of America*, 109(20), 7687e7692.

Setplan. 2017. Socio-Economic Agricultural Plan for the PHA: Spatial Planning and Land Development Management Input. City of Cape Town: Unpublished.

Shahabi, H., Shirzadi, A., Ghaderi, K., Omidvar, E., Al-Ansari, N., Clague, J.J., Geertsema, M., Khosravi, K., Amini, A., Bahrami, S. and Rahmati, O., 2020. *Flood detection and susceptibility mapping using sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on k-nearest neighbor classifier. Remote Sensing*, 12(2): 266.

Sharifi, A., 2020. Remotely sensed vegetation indices for crop nutrition mapping. Journal of the Science of Food and Agriculture, 100(14), pp.5191-5196.

Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W. and Sun, Y., 2020. Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv preprint arXiv:2009.03509*.

Shih, Yang-Hsin & Qin, Gong & Tang, Peifu & Shen, & Liu, Tai-Yi & Gao, Shuai. (2019). *Delineation of Urban Growth Boundaries Using a Patch-Based Cellular Automata Model under Multiple Spatial and Socio-Economic Scenarios*. Sustainability. 11. 6159. 10.3390/su11216159.

Singh, Sukhdeep & Bhardwaj, Anil & Verma, V.K.. (2020). *Remote sensing and GIS based analysis of temporal land use/land cover and water quality changes in Harike wetland ecosystem, Punjab, India*. Journal of Environmental Management. 262. 110355. 10.1016/j.jenvman.2020.110355.

Spain D. Republic of South Africa: *Unravelling the population puzzle*. Country profile. Int Demogr. 1984 Jun;3(6):4-11. PMID: 12313172.

Sripada., R, Heiniger., R, White., J and Meijer., A. 2006. *Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn.*

Stubbings, P., Peskett, J., Rowe, F. & Arribas-Bel, D. 2019. *A hierarchical urban forest index using street-level imagery and deep learning. Remote Sensing*, 11(12):1–22. DOI: 10.3390/rs11121395.

Sudmeier-Rieux, K., Paleo, U.F., Garschagen, M., Estrella, M., Renaud, F.G. and Jaboyedoff, M., 2015. Opportunities, incentives and challenges to risk sensitive land use planning: Lessons from Nepal, Spain and Vietnam. *International Journal of Disaster Risk Reduction*, *14*, pp.205-224.

Sun, S. and Huang, R., 2010, August. An adaptive k-nearest neighbor algorithm. In *2010 seventh international conference on fuzzy systems and knowledge discovery* (Vol. 1, pp. 91-94). IEEE.

Szul, T, Sylwester T, and Krzysztof, P. 2021. Application of the BORUTA Algorithm to Input Data Selection for a Model Based on Rough Set Theory (RST) to Prediction Energy Consumption for Building Heating. Energies 14(10): 2779. https://doi.org/10.3390/en141027

Tysa, S.K., Ren, G., Qin, Y., Zhang, P., Ren, Y., Jia, W. and Wen, K., 2019. Urbanization effect in regional temperature series based on a remote sensing classification scheme of stations. *Journal of Geophysical Research: Atmospheres*, *124*(20), pp.10646-10661.

UN-Habitat. 2016. *World cities report 2016: Urbanization and development –Emerging futures*. New York: United Nations: https://unhabitat.org/world-cities-report-2016 (Accessed: 15/05/2022).

United Nations. 2018. *World urbanization prospects*: The 2018 revision, UN DESA: http://esa.un.org/unpd/wup/ (Accessed: 21/01/2021).

UMVOTO Africa (Pty) Ltd. 2017. Socio-Economic Agricultural Plan for the Existing PHA: Survey of Natural Resources. Cape Town: Unpublished.

Viana, C.M., Girão, I. and Rocha, J., 2019. Long-term satellite image time-series for land use/land cover change detection using refined open source data in a rural region. *Remote Sensing*, *11*(9), p.1104.

Van Coillie, F.M., Verbeke, L.P. and De Wulf, R.R., 2007. Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. *Remote Sensing of Environment*, *110*(4), pp.476-487.

Villa, P., 2012. Mapping urban growth using Soil and Vegetation Index and Landsat data: The Milan (Italy) city area case study. *Landscape and urban planning*, *107*(3), pp.245-254.

Vlahov, D. & Galea, S. 2002. Urbanization, urbanicity, and health. Journal of Urban Health: Bulletin of the New York Academy of Medicine, 79(4): 1S – 12. http://link.springer.com/10.1093/jurban/79.suppl_1.S1.

Vogelmann, J.E., Rock, B.N. and Moss, D.M., 1993. Red edge spectral measurements from sugar maple leaves. *TitleREMOTE SENSING*, *14*(8), pp.1563-1575.

Volpi, M., Tuia, D., Bovolo, F., Kanevski, M. and Bruzzone, L., 2013. Supervised change detection in VHR images using contextual information and support vector machines. International Journal of Applied Earth Observation and Geoinformation, 20: 77-85.

Wang, J., Zhou, W., Pickett, S.T., Yu, W. and Li, W., 2019. A multiscale analysis of urbanization effects on ecosystem services supply in an urban megaregion. *Science of the total environment*, *662*, pp.824-833.

Waring, R.H., Coops, N.C., Fan, W. and Nightingale, J.M., 2006. MODIS enhanced vegetation index predicts tree species richness across forested ecoregions in the contiguous USA. *Remote Sensing of Environment*, *103*(2), pp.218-226.

Weber, C. and Puissant, A., 2003. Urbanization pressure and modeling of urban growth: example of the Tunis Metropolitan Area. *Remote sensing of environment*, *86*(3), pp.341-352.

Wenjing Lv, Wang X, "*Overview of Hyperspectral Image Classification*", Journal of Sensors, vol. 2020, Article ID 4817234, 13 pages, 2020.

Wieland, M. & Pittore, M. 2014. Urban pattern recognition from multi-spectral satellite images are done by performance evaluation of machine learning algorithms. Remote Sensing. 6(4):2912–2939. DOI: 10.3390/rs6042912

Wu, H., Lin, A., Xing, X., Song, D. and Li, Y., 2021. Identifying core driving factors of urban land use change from global land cover products and POI data using the random forest method. International Journal of Applied Earth Observation and Geoinformation, 103: 102475.

Zendehboudi, A., Baseer, M.A. and Saidur, R., 2018. Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of cleaner production*, *199*, pp.272-285.

Zhai, H., Lv, C., Liu, W., Yang, C., Fan, D., Wang, Z. and Guan, Q., 2021. Understanding spatio-temporal patterns of land use/land cover change under urbanization in Wuhan, China, 2000–2019. *Remote Sensing*, *13*(16), p.3331.

Zhu, Z. (2017). Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. ISPRS Journal of Photogrammetry and Remote Sensing. 130. 10.1016/j.isprsjprs.2017.06.013.

Appendices

Appendix A. **Accuracy results**

The figure below shows the classifiers used in this study and the results over a 7 year period.

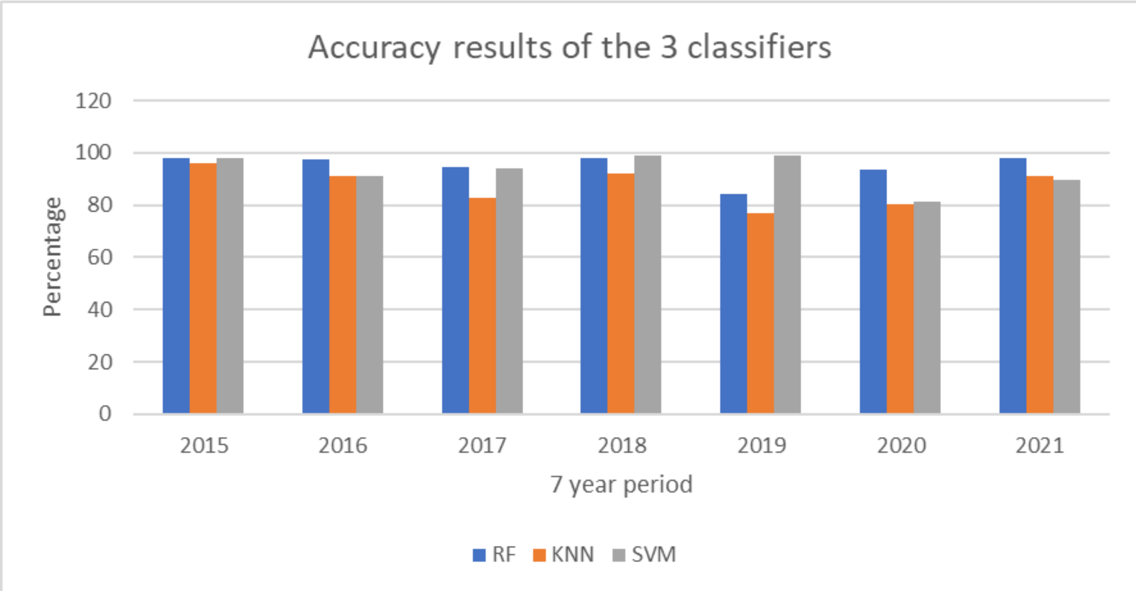
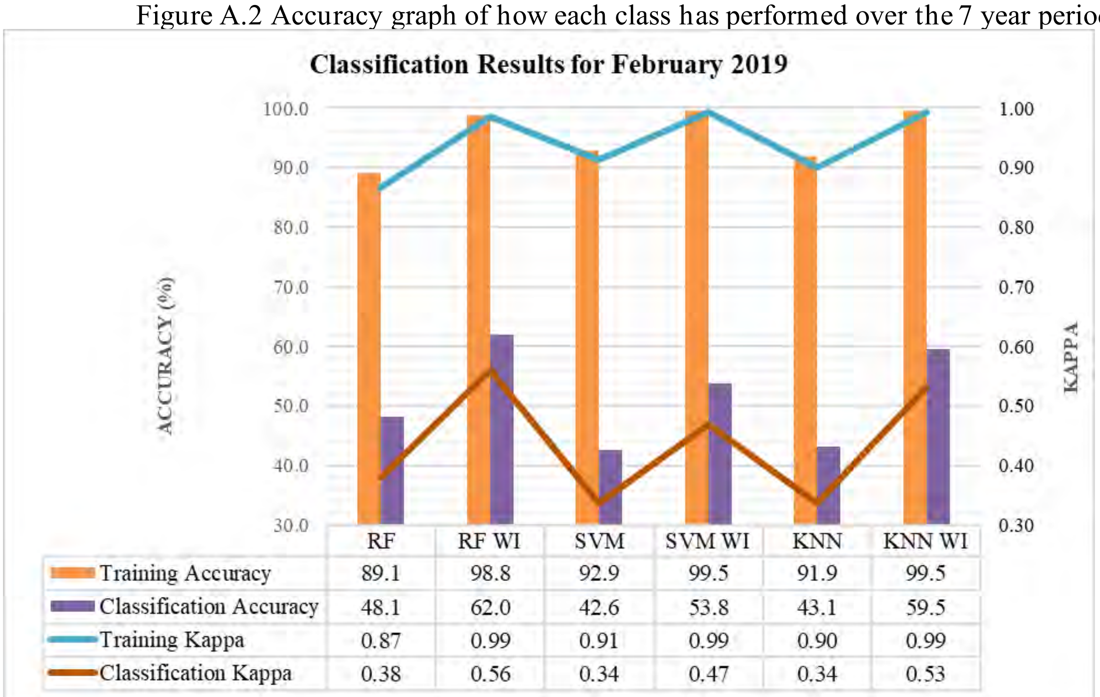


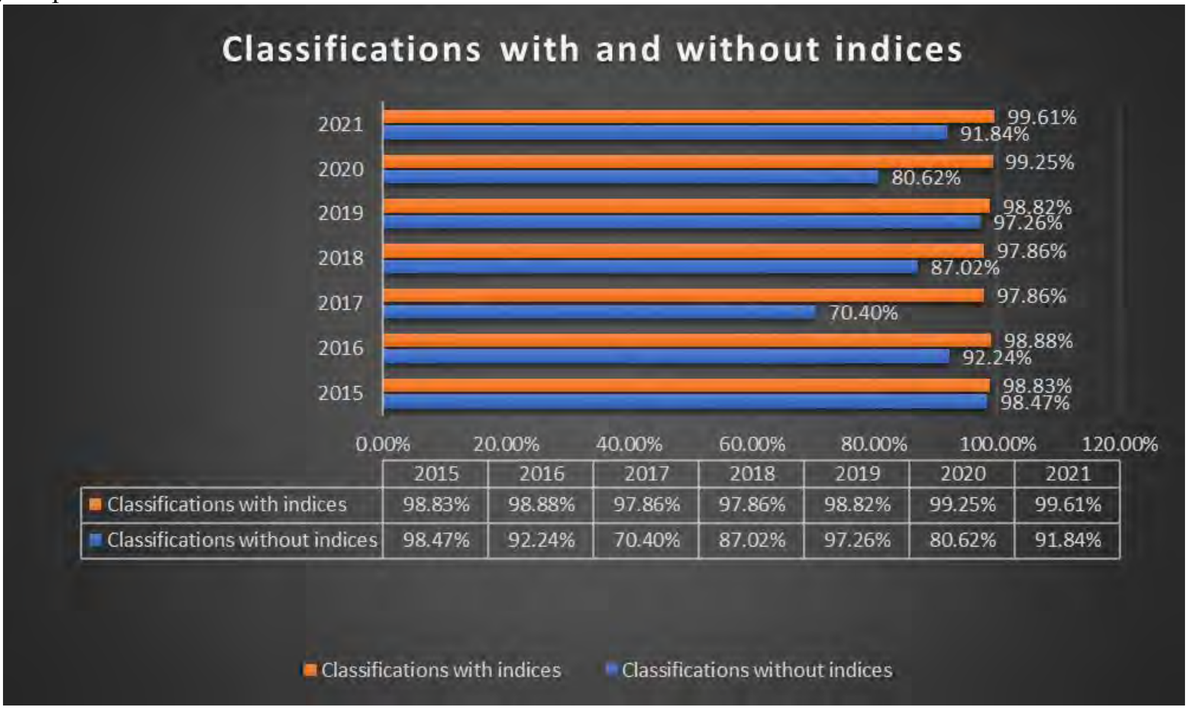
Figure A.2 Accuracy graph of how each class has performed over the 7 year period.



	RF	RF WI	SVM	SVM WI	KNN	KNN WI
Training Accuracy	89.1	98.8	92.9	99.5	91.9	99.5
Classification Accuracy	48.1	62.0	42.6	53.8	43.1	59.5
Training Kappa	0.87	0.99	0.91	0.99	0.90	0.99
Classification Kappa	0.38	0.56	0.34	0.47	0.34	0.53

Appendix B. **Effect of Indices**

Figure B.1 below shows the difference in accuracies for classifications with and without indices for the 7-year period.



Appendix C. **Results of Random Forest error matrix calculations**

The figures below show results from Random Forest Layer Stack classification between 2015 and 2021.

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area        Wi
1       0.2133  0.0000  0.0000  0.0000  0.0049  0.0000  10269300.0000   0.2181
2       0.0000  0.0091  0.0000  0.0000  0.0000  0.0000  426400.0000     0.0091
3       0.0000  0.0028  0.0969  0.0002  0.0000  0.0000  4699100.0000    0.0998
4       0.0000  0.0000  0.0000  0.2252  0.0021  0.0000  10702100.0000   0.2273
5       0.0042  0.0000  0.0000  0.0056  0.3948  0.0000  19046900.0000   0.4046
6       0.0000  0.0000  0.0000  0.0000  0.0000  0.0411  1934500.0000    0.0411
Total   0.2174  0.0118  0.0969  0.2309  0.4018  0.0411  47078300.0000
Area    10236726        557158  4560170 10872531        18917215        1934500 47078300
SE      0.0030  0.0007  0.0007  0.0031  0.0043  0.0000
SE area 141801  32259   33223   148184  204070  0
95% CI area     277929  63228   65116   290440  399977  0
PA  [%] 98.0818 76.5313 100.0000        97.5168 98.2636 100.0000
UA  [%] 97.7707 100.0000        97.0435 99.0698 97.5945 100.0000
Kappa hat       0.9715  1.0000  0.9673  0.9879  0.9598  1.0000

Overall accuracy [%] = 98.0339
Kappa hat classification = 0.9729

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

Figure C.1 Error Matrix for the April 2015 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area        Wi
1       0.3140  0.0000  0.0000  0.0010  0.0000  0.0015  14901100.0000   0.3165
2       0.0000  0.0097  0.0000  0.0003  0.0000  0.0000  470900.0000     0.0100
3       0.0000  0.0000  0.1080  0.0027  0.0000  0.0000  5211400.0000    0.1107
4       0.0003  0.0026  0.0058  0.2289  0.0054  0.0000  11440000.0000   0.2430
5       0.0017  0.0000  0.0000  0.0045  0.2532  0.0000  12212300.0000   0.2594
6       0.0008  0.0000  0.0000  0.0000  0.0000  0.0596  2842600.0000    0.0604
Total   0.3168  0.0123  0.1138  0.2373  0.2587  0.0611  47078300.0000
Area    14916034        578420  5356625 11173241        12177426        2876555 47078300
SE      0.0014  0.0009  0.0014  0.0025  0.0019  0.0009
SE area 66517   42697   66971   119866  91444   42555
95% CI area     130373  83686   131263  234937  179230  83408
PA  [%] 99.1133 79.1810 94.9418 96.4598 97.8986 97.5527
UA  [%] 99.2126 97.2603 97.5875 94.2105 97.6190 98.7179
Kappa hat       0.9885  0.9723  0.9728  0.9241  0.9679  0.9863

Overall accuracy [%] = 97.3544
Kappa hat classification = 0.9651

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

Figure C.2 Error Matrix for the April 2016 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area            Wi
1       0.2912  0.0001  0.0001  0.0001  0.0041  0.0005  13938381.0000   0.2961
2       0.0001  0.0118  0.0002  0.0003  0.0000  0.0000  585828.0000     0.0124
3       0.0003  0.0001  0.1702  0.0019  0.0032  0.0000  8271441.0000    0.1757
4       0.0008  0.0016  0.0177  0.2101  0.0030  0.0000  10982313.0000   0.2333
5       0.0057  0.0009  0.0075  0.0043  0.2396  0.0002  12159819.0000   0.2583
6       0.0001  0.0000  0.0000  0.0000  0.0007  0.0235  1141875.0000    0.0243
Total   0.2982  0.0145  0.1957  0.2168  0.2507  0.0242  47079657.0000
Area    14039019        682740  9212918 10206922        11800518        1137541 47079657
SE      0.0004  0.0002  0.0006  0.0006  0.0006  0.0001
SE area 18528   9292    28069   27601   28505   5479
95% CI area     36315   18213   55015   54097   55869   10738
PA  [%] 97.6537 81.2467 86.9704 96.9263 95.6057 97.1494
UA  [%] 98.3588 94.6871 96.8696 90.0829 92.7807 96.7807
Kappa hat       0.9766  0.9461  0.9611  0.8734  0.9037  0.9670

Overall accuracy [%] = 94.6420
Kappa hat classification = 0.9296

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

Figure C.3 Error Matrix for the April 2017 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area            Wi
1       0.3178  0.0001  0.0000  0.0010  0.0014  0.0000  15080200.0000   0.3203
2       0.0001  0.0190  0.0000  0.0001  0.0000  0.0000  901300.0000     0.0191
3       0.0000  0.0000  0.1079  0.0002  0.0004  0.0000  5107000.0000    0.1085
4       0.0002  0.0027  0.0006  0.2241  0.0014  0.0000  10784900.0000   0.2291
5       0.0057  0.0003  0.0012  0.0022  0.2898  0.0007  14115200.0000   0.2998
6       0.0000  0.0000  0.0000  0.0000  0.0000  0.0231  1089700.0000    0.0231
Total   0.3238  0.0221  0.1097  0.2276  0.2930  0.0238  47078300.0000
Area    15242406        1038639 5164140 10715946        13795578        1121592 47078300
SE      0.0010  0.0008  0.0006  0.0012  0.0013  0.0003
SE area 48332   36429   27429   55877   62795   14249
95% CI area     94731   71400   53760   109519  123078  27929
PA  [%] 98.1592 86.1985 98.3260 98.4752 98.8955 97.1566
UA  [%] 99.2151 99.3333 99.4261 97.8456 96.6561 100.0000
Kappa hat       0.9884  0.9932  0.9936  0.9721  0.9527  1.0000

Overall accuracy [%] = 98.1774
Kappa hat classification = 0.9755

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

Figure C.4 Error Matrix for the April 2018 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area           Wi
1       0.2969  0.0002  0.0006  0.0007  0.0083  0.0000  14391918.0000  0.3067
2       0.0000  0.0107  0.0000  0.0003  0.0000  0.0000  521667.0000    0.0111
3       0.0001  0.0005  0.0921  0.0034  0.0361  0.0000  6205788.0000   0.1323
4       0.0045  0.0065  0.0556  0.2774  0.0061  0.0004  16443081.0000  0.3504
5       0.0021  0.0001  0.0282  0.0008  0.1481  0.0002  8416764.0000   0.1794
6       0.0000  0.0000  0.0000  0.0000  0.0003  0.0197  943938.0000    0.0201
Total   0.3036  0.0180  0.1766  0.2826  0.1990  0.0203  46923156.0000
Area    14244434        845318  8285703 13258622        9335683 953395 46923156
SE      0.0004  0.0004  0.0013  0.0011  0.0010  0.0001
SE area 20941   16849   59911   49830   45930   5603
95% CI area     41045   33023   117425  97667   90024   10981
PA  [%] 97.7973 59.5175 52.1719 98.1812 74.4435 97.1479
UA  [%] 96.7952 96.4431 69.6577 79.1669 82.5710 98.1212
Kappa hat       0.9540  0.9638  0.6315  0.7096  0.7824  0.9808

Overall accuracy [%] = 84.5000
Kappa hat classification = 0.7928

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

Figure C.5 Error Matrix for the April 2019 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area           Wi
1       0.2880  0.0002  0.0003  0.0002  0.0009  0.0000  13629000.0000  0.2895
2       0.0000  0.0137  0.0002  0.0014  0.0000  0.0000  724300.0000    0.0154
3       0.0018  0.0001  0.1263  0.0012  0.0076  0.0000  6454600.0000   0.1371
4       0.0004  0.0039  0.0150  0.2603  0.0064  0.0002  13467100.0000  0.2861
5       0.0042  0.0004  0.0175  0.0040  0.2378  0.0004  12439000.0000  0.2642
6       0.0000  0.0000  0.0000  0.0000  0.0000  0.0077  364300.0000    0.0077
Total   0.2943  0.0183  0.1593  0.2670  0.2526  0.0083  47078300.0000
Area    13856488        863146  7500941 12572127        11892631       392967  47078300
SE      0.0012  0.0010  0.0028  0.0024  0.0027  0.0004
SE area 54418   45591   130380  111395  127334  16557
95% CI area     106660  89357   255544  218333  249574  32452
PA  [%] 97.8441 74.8816 79.2867 97.4593 94.1265 92.7050
UA  [%] 99.4773 89.2361 92.1397 90.9825 89.9921 100.0000
Kappa hat       0.9926  0.8904  0.9065  0.8770  0.8661  1.0000

Overall accuracy [%] = 93.3816
Kappa hat classification = 0.9117

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

Figure C.6 Error Matrix for the April 2020 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area            Wi
1       0.3879  0.0003  0.0000  0.0008  0.0014  0.0000  18382707.0000   0.3905
2       0.0000  0.0188  0.0001  0.0028  0.0000  0.0000  1023354.0000    0.0217
3       0.0000  0.0003  0.0940  0.0008  0.0000  0.0000  4479255.0000    0.0951
4       0.0002  0.0035  0.0077  0.2399  0.0000  0.0000  11832264.0000   0.2513
5       0.0026  0.0003  0.0000  0.0002  0.2201  0.0000  10508940.0000   0.2232
6       0.0001  0.0000  0.0000  0.0000  0.0000  0.0181  853137.0000     0.0181
Total   0.3908  0.0232  0.1018  0.2445  0.2215  0.0181  47079657.0000
Area    18398958        1093202 4792888 11512477        10430282        851850  47079657
SE      0.0003  0.0003  0.0003  0.0004  0.0003  0.0000
SE area 12992   13161   16296   20774   12354   1670
95% CI area     25465   25796   31941   40716   24215   3274
PA  [%] 99.2660 81.0826 92.3708 98.1164 99.3589 99.8189
UA  [%] 99.3538 86.6168 98.8385 95.4646 98.6152 99.6683
Kappa hat       0.9894  0.8630  0.9871  0.9400  0.9822  0.9966

Overall accuracy [%] = 97.8913
Kappa hat classification = 0.9710

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

Figure C.7 Error Matrix for the April 2021 dataset

Appendix D. **Results of K-Nearest Neighbor error matrix calculations**

K-Nearest Neighbors 2015 – 2021

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area          Wi
1       0.2071  0.0000  0.0000  0.0000  0.0007  0.0000  9782800.0000   0.2078
2       0.0000  0.0080  0.0000  0.0000  0.0000  0.0000  374500.0000    0.0080
3       0.0000  0.0026  0.1027  0.0004  0.0000  0.0000  4973300.0000   0.1056
4       0.0000  0.0020  0.0000  0.2093  0.0128  0.0000  10552200.0000  0.2241
5       0.0142  0.0000  0.0000  0.0057  0.3963  0.0000  19593600.0000  0.4162
6       0.0000  0.0000  0.0000  0.0000  0.0000  0.0383  1801900.0000   0.0383
Total   0.2213  0.0125  0.1027  0.2154  0.4098  0.0383  47078300.0000
Area    10419023        588771  4834671 10139736        19294199       1801900 47078300
SE      0.0045  0.0016  0.0007  0.0047  0.0063  0.0000
SE area 210706  73006   34201   219680  295763  0
95% CI area     412984  143091  67033   430573  579696  0
PA  [%] 93.5817 63.6071 100.0000        97.1911 96.6995 100.0000
UA  [%] 99.6678 100.0000        97.2125 93.3921 95.2218 100.0000
Kappa hat       0.9957  1.0000  0.9689  0.9158  0.9190  1.0000

Overall accuracy [%] = 96.1668
Kappa hat classification = 0.9470

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

Figure D.1 Error Matrix for the April 2015 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area          Wi
1       0.2892  0.0000  0.0000  0.0005  0.0033  0.0009  13838900.0000  0.2940
2       0.0000  0.0190  0.0000  0.0014  0.0000  0.0000  960200.0000    0.0204
3       0.0000  0.0000  0.1165  0.0036  0.0007  0.0000  5686000.0000   0.1208
4       0.0009  0.0045  0.0194  0.2003  0.0325  0.0000  12125900.0000  0.2576
5       0.0056  0.0000  0.0000  0.0142  0.2083  0.0000  10735600.0000  0.2280
6       0.0018  0.0000  0.0000  0.0000  0.0000  0.0775  3731700.0000   0.0793
Total   0.2975  0.0235  0.1359  0.2199  0.2448  0.0785  47078300.0000
Area    14004327        1105547 6398262 10354016        11522602       3693546 47078300
SE      0.0021  0.0012  0.0024  0.0043  0.0040  0.0008
SE area 99878   56318   112833  201678  188649  38398
95% CI area     195761  110384  221153  395290  369752  75259
PA  [%] 97.2249 80.9579 85.7422 91.0879 85.0916 98.7914
UA  [%] 98.3871 93.2127 96.4828 77.7778 91.3295 97.7813
Kappa hat       0.9770  0.9305  0.9593  0.7151  0.8852  0.9759

Overall accuracy [%] = 91.0858
Kappa hat classification = 0.8853

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

Figure D.2 Error Matrix for the April 2016 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area    Wi
1       0.2500  0.0001  0.0011  0.0002  0.0145  0.0028  12649959.0000   0.2687
2       0.0006  0.0215  0.0050  0.0010  0.0088  0.0002  1740510.0000    0.0370
3       0.0006  0.0004  0.1503  0.0057  0.0072  0.0000  7727382.0000    0.1641
4       0.0009  0.0039  0.0265  0.1656  0.0040  0.0000  9456858.0000    0.2009
5       0.0263  0.0010  0.0270  0.0265  0.2134  0.0000  13850829.0000   0.2942
6       0.0012  0.0000  0.0024  0.0000  0.0041  0.0274  1654119.0000    0.0351
Total   0.2795  0.0270  0.2121  0.1989  0.2520  0.0304  47079657.0000
Area    13160736        1272672 9986940 9366056 11862984        1430269 47079657
SE      0.0008  0.0004  0.0009  0.0009  0.0011  0.0003
SE area 36528   19401   42475   40954   52979   13251
95% CI area     71596   38027   83251   80271   103840  25971
PA  [%] 89.4370 79.6249 70.8362 83.2495 84.6968 90.3136
UA  [%] 93.0483 58.2222 91.5494 82.4502 72.5413 78.0916
Kappa hat       0.9035  0.5706  0.8927  0.7809  0.6329  0.7741

Overall accuracy [%] = 82.8273
Kappa hat classification = 0.7781

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

Figure D.3 Error Matrix for the April 2017 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area    Wi
1       0.2813  0.0005  0.0000  0.0001  0.0027  0.0000  13399000.0000   0.2846
2       0.0003  0.0373  0.0046  0.0044  0.0007  0.0000  2227900.0000    0.0473
3       0.0001  0.0001  0.1100  0.0028  0.0199  0.0000  6261400.0000    0.1330
4       0.0021  0.0074  0.0008  0.1941  0.0110  0.0000  10133900.0000   0.2153
5       0.0160  0.0006  0.0006  0.0039  0.2730  0.0003  13861900.0000   0.2944
6       0.0005  0.0000  0.0000  0.0000  0.0002  0.0247  1194200.0000    0.0254
Total   0.3002  0.0458  0.1161  0.2053  0.3076  0.0250  47078300.0000
Area    14134696        2157656 5463530 9664218 14481791        1176409 47078300
SE      0.0017  0.0016  0.0015  0.0022  0.0026  0.0003
SE area 82046   75889   72402   105054  121704  14481
95% CI area     160809  148742  141908  205905  238539  28383
PA  [%] 93.6989 81.4367 94.8275 94.5489 88.7572 98.8013
UA  [%] 98.8436 78.8690 82.7439 90.1668 92.7263 97.3294
Kappa hat       0.9835  0.7785  0.8048  0.8763  0.8949  0.9726

Overall accuracy [%] = 92.0497
Kappa hat classification = 0.8956

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

Figure D.4 Error Matrix for the April 2018 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area    Wi
1       0.3402  0.0004  0.0027  0.0028  0.0264  0.0004  17520867.0000   0.3729
2       0.0000  0.0098  0.0005  0.0004  0.0000  0.0000  497547.0000     0.0106
3       0.0015  0.0003  0.0887  0.0109  0.0744  0.0001  8264898.0000    0.1759
4       0.0034  0.0083  0.0457  0.2209  0.0054  0.0001  13334112.0000   0.2838
5       0.0025  0.0006  0.0376  0.0064  0.0944  0.0013  6707169.0000    0.1427
6       0.0000  0.0000  0.0000  0.0000  0.0013  0.0128  662103.0000     0.0141
Total   0.3477  0.0193  0.1752  0.2414  0.2017  0.0146  46986696.0000
Area    16338274        908537  8230309 11343065        9478316 688195  46986696
SE      0.0007  0.0004  0.0013  0.0010  0.0012  0.0002
SE area 30589   18072   62494   48304   58297   9161
95% CI area     59955   35420   122488  94677   114262  17956
PA  [%] 97.8510 50.5137 50.6565 91.5175 46.7766 87.4728
UA  [%] 91.2464 92.2397 50.4445 77.8521 66.1029 90.9199
Kappa hat       0.8658  0.9209  0.3992  0.7080  0.5754  0.9078

Overall accuracy [%] = 76.6851
Kappa hat classification = 0.6857

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

Figure D.5 Error Matrix for the April 2019 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area    Wi
1       0.2911  0.0002  0.0065  0.0005  0.0136  0.0000  14679100.0000   0.3118
2       0.0000  0.0177  0.0004  0.0000  0.0000  0.0000  851400.0000     0.0181
3       0.0011  0.0006  0.0852  0.0078  0.0428  0.0005  6494700.0000    0.1380
4       0.0035  0.0062  0.0295  0.2500  0.0144  0.0006  14315300.0000   0.3041
5       0.0058  0.0002  0.0486  0.0109  0.1456  0.0004  9957000.0000    0.2115
6       0.0008  0.0000  0.0001  0.0000  0.0000  0.0157  780900.0000     0.0166
Total   0.3023  0.0249  0.1701  0.2691  0.2165  0.0171  47078400.0000
Area    14230398        1172447 8010139 12668741        10190732        805943  47078400
SE      0.0020  0.0012  0.0043  0.0034  0.0042  0.0006
SE area 92811   54623   201020  161660  196500  27829
95% CI area     181909  107061  394000  316853  385141  54545
PA  [%] 96.2890 71.1533 50.0465 92.8880 67.2630 91.5755
UA  [%] 93.3457 97.9839 61.7241 82.2039 68.8419 94.5122
Kappa hat       0.9046  0.9793  0.5388  0.7565  0.6023  0.9442

Overall accuracy [%] = 80.5162
Kappa hat classification = 0.7416

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

Figure D.6 Error Matrix for the April 2020 dataset

```
      > AREA BASED ERROR MATRIX
      > Reference
V_Classified    1       2       3       4       5       6       Area           Wi
1       0.3268  0.0002  0.0003  0.0004  0.0064  0.0002  15756588.0000   0.3342
2       0.0009  0.0218  0.0002  0.0068  0.0000  0.0000  1398645.0000    0.0297
3       0.0001  0.0003  0.0962  0.0161  0.0005  0.0000  5332221.0000    0.1131
4       0.0019  0.0073  0.0310  0.2219  0.0019  0.0000  12442995.0000   0.2639
5       0.0122  0.0005  0.0000  0.0016  0.2260  0.0003  11343654.0000   0.2406
6       0.0000  0.0000  0.0000  0.0000  0.0000  0.0185  871758.0000     0.0185
Total   0.3420  0.0300  0.1276  0.2468  0.2347  0.0190  47145861.0000
Area    16123497        1413368 6014342 11635428        11064728        894498  47145861
SE      0.0006  0.0004  0.0008  0.0009  0.0006  0.0001
SE area 26918   19369   35958   41348   28007   4018
95% CI area     52760   37963   70478   81041   54894   7875
PA  [%] 95.5599 72.6567 75.3742 89.9144 96.2940 97.4578
UA  [%] 97.7851 73.4216 85.0164 84.0789 93.9262 100.0000
Kappa hat       0.9663  0.7260  0.8283  0.7886  0.9206  1.0000

Overall accuracy [%] = 91.1132
Kappa hat classification = 0.8813

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

Figure D.7 Error Matrix for the April 2021 dataset

Appendix E. **Results of Support Vector Machine error matrix calculations**

Support Vector Machine 2015 – 2021

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area        Wi
1       0.2359  0.0000  0.0000  0.0000  0.0008  0.0016  11213900.0000   0.2382
2       0.0000  0.0069  0.0000  0.0000  0.0000  0.0000  322600.0000     0.0069
3       0.0000  0.0030  0.1038  0.0004  0.0000  0.0000  5042300.0000    0.1071
4       0.0000  0.0000  0.0000  0.2315  0.0000  0.0000  10900700.0000   0.2315
5       0.0074  0.0000  0.0000  0.0062  0.3600  0.0000  17586000.0000   0.3735
6       0.0000  0.0000  0.0000  0.0000  0.0000  0.0428  2012800.0000    0.0428
Total   0.2433  0.0098  0.1038  0.2381  0.3608  0.0443  47078300.0000
Area    11452556        462664  4884728 11208406        16984092        2085855 47078300
SE      0.0033  0.0007  0.0008  0.0028  0.0041  0.0011
SE area 154446  34556   36587   129510  192774  51573
95% CI area     302714  67730   71710   253840  377837  101083
PA  [%] 96.9593 69.7266 100.0000        97.2547 99.7849 96.4976
UA  [%] 99.0228 100.0000        96.8750 100.0000        96.3696 100.0000
Kappa hat       0.9871  1.0000  0.9651  1.0000  0.9432  1.0000

Overall accuracy [%] = 98.0764
Kappa hat classification = 0.9740

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

Figure E.1 Error Matrix for the April 2015 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area    Wi
1       0.3097  0.0000  0.0000  0.0035  0.0035  0.0005  14938600.0000   0.3173
2       0.0000  0.0206  0.0000  0.0012  0.0000  0.0000  1026300.0000    0.0218
3       0.0000  0.0001  0.0950  0.0024  0.0000  0.0000  4588300.0000    0.0975
4       0.0008  0.0053  0.0124  0.2613  0.0488  0.0000  15467700.0000   0.3286
5       0.0024  0.0000  0.0000  0.0052  0.1456  0.0000  7207900.0000    0.1531
6       0.0023  0.0000  0.0000  0.0000  0.0000  0.0795  3849500.0000    0.0818
Total   0.3152  0.0259  0.1074  0.2736  0.1979  0.0800  47078300.0000
Area    14837547        1220900 5056134 12882062        9317356 3764302 47078300
SE      0.0022  0.0014  0.0022  0.0048  0.0044  0.0007
SE area 103686  67641   101693  227418  205634  34725
95% CI area     203225  132576  199318  445740  403043  68061
PA  [%] 98.2801 79.4338 88.4625 95.5084 73.5514 99.3691
UA  [%] 97.6153 94.4954 97.4823 79.5429 95.0769 97.1698
Kappa hat       0.9652  0.9435  0.9718  0.7184  0.9386  0.9692

Overall accuracy [%] = 91.1715
Kappa hat classification = 0.8842

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

Figure E.2 Error Matrix for the April 2016 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area    Wi
1       0.1663  0.0000  0.0000  0.0019  0.0019  0.0003  14938600.0000   0.1704
2       0.0000  0.0111  0.0000  0.0006  0.0000  0.0000  1026400.0000    0.0117
3       0.0000  0.0000  0.0510  0.0013  0.0000  0.0000  4588300.0000    0.0523
4       0.0004  0.0028  0.0067  0.1403  0.0262  0.0000  15467700.0000   0.1764
5       0.0013  0.0000  0.0000  0.0028  0.0782  0.0000  7207900.0000    0.0822
6       0.0143  0.0000  0.0000  0.0000  0.0000  0.4926  44441900.0000   0.5069
Total   0.1823  0.0139  0.0577  0.1469  0.1063  0.4928  87670800.0000
Area    15986389        1220994 5056134 12882067        9317356 43207860        87670800
SE      0.0035  0.0008  0.0012  0.0026  0.0023  0.0033
SE area 309268  67641   101693  227418  205634  293431
95% CI area     606166  132577  199318  445740  403043  575125
PA  [%] 91.2173 79.4353 88.4625 95.5084 73.5514 99.9450
UA  [%] 97.6153 94.4954 97.4823 79.5429 95.0769 97.1698
Kappa hat       0.9708  0.9442  0.9733  0.7602  0.9449  0.9442

Overall accuracy [%] = 93.9488
Kappa hat classification = 0.9112

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

Figure E.3 Error Matrix for the April 2017 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area    Wi
1       0.1687  0.0000  0.0000  0.0001  0.0004  0.0000  14829800.0000   0.1692
2       0.0030  0.4674  0.0000  0.0060  0.0000  0.0000  41769500.0000   0.4764
3       0.0000  0.0000  0.0624  0.0001  0.0000  0.0000  5478500.0000    0.0625
4       0.0002  0.0004  0.0000  0.1224  0.0003  0.0000  10820500.0000   0.1234
5       0.0012  0.0001  0.0001  0.0001  0.1536  0.0000  13601700.0000   0.1551
6       0.0001  0.0000  0.0000  0.0000  0.0000  0.0133  1170800.0000    0.0134
Total   0.1732  0.4679  0.0625  0.1287  0.1544  0.0133  87670800.0000
Area    15185834        41018754        5483170 11282774        13536522        1163747 87670800
SE      0.0022  0.0037  0.0001  0.0030  0.0004  0.0001
SE area 190051  322847  10434   265678  35570   4980
95% CI area     372500  632779  20451   520729  69718   9760
PA  [%] 97.3800 99.8906 99.7719 95.1329 99.5113 100.0000
UA  [%] 99.7179 98.0952 99.8569 99.1971 99.0345 99.3976
Kappa hat       0.9966  0.9642  0.9985  0.9908  0.9886  0.9939

Overall accuracy [%] = 98.7789
Kappa hat classification = 0.9827

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

Figure E.4 Error Matrix for the April 2018 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area    Wi
1       0.1687  0.0000  0.0000  0.0001  0.0004  0.0000  14829800.0000   0.1692
2       0.0030  0.4674  0.0000  0.0060  0.0000  0.0000  41769500.0000   0.4764
3       0.0000  0.0000  0.0624  0.0001  0.0000  0.0000  5478500.0000    0.0625
4       0.0002  0.0004  0.0000  0.1224  0.0003  0.0000  10820500.0000   0.1234
5       0.0012  0.0001  0.0001  0.0001  0.1536  0.0000  13601700.0000   0.1551
6       0.0001  0.0000  0.0000  0.0000  0.0000  0.0133  1170800.0000    0.0134
Total   0.1732  0.4679  0.0625  0.1287  0.1544  0.0133  87670800.0000
Area    15185834        41018754        5483170 11282774        13536522        1163747 87670800
SE      0.0022  0.0037  0.0001  0.0030  0.0004  0.0001
SE area 190051  322847  10434   265678  35570   4980
95% CI area     372500  632779  20451   520729  69718   9760
PA  [%] 97.3800 99.8906 99.7719 95.1329 99.5113 100.0000
UA  [%] 99.7179 98.0952 99.8569 99.1971 99.0345 99.3976
Kappa hat       0.9966  0.9642  0.9985  0.9908  0.9886  0.9939

Overall accuracy [%] = 98.7789
Kappa hat classification = 0.9827

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

Figure E.5 Error Matrix for the April 2019 dataset

```
        > AREA BASED ERROR MATRIX
        > Reference
V_Classified    1       2       3       4       5       6       Area         Wi
1       0.3002  0.0001  0.0052  0.0006  0.0145  0.0003  15105000.0000   0.3208
2       0.0000  0.0167  0.0001  0.0010  0.0000  0.0000  840700.0000     0.0179
3       0.0019  0.0007  0.0981  0.0093  0.0452  0.0005  7329400.0000    0.1557
4       0.0024  0.0059  0.0218  0.2523  0.0138  0.0006  13970600.0000   0.2968
5       0.0046  0.0004  0.0496  0.0087  0.1347  0.0004  9337600.0000    0.1983
6       0.0002  0.0000  0.0000  0.0000  0.0000  0.0103  495100.0000     0.0105
Total   0.3093  0.0237  0.1748  0.2720  0.2081  0.0121  47078400.0000
Area    14561426        1117598 8231105 12804438        9795453 568381  47078400
SE      0.0019  0.0012  0.0042  0.0032  0.0041  0.0006
SE area 88282   55745   196240  152889  195157  26256
95% CI area     173032  109261  384631  299662  382507  51462
PA  [%] 97.0633 70.3429 56.1093 92.7775 64.7190 85.4212
UA  [%] 93.5703 93.5115 63.0122 85.0331 67.8924 98.0645
Kappa hat       0.9069  0.9335  0.5518  0.7944  0.5946  0.9804

Overall accuracy [%] = 81.2327
Kappa hat classification = 0.7501

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

Figure E.6 Error Matrix for the April 2020 dataset

```
      > AREA BASED ERROR MATRIX
      > Reference
V_Classified    1       2       3       4       5       6       Area          Wi
1       0.3617  0.0005  0.0003  0.0007  0.0188  0.0008  18052479.0000   0.3829
2       0.0021  0.0291  0.0006  0.0076  0.0000  0.0000  1861191.0000    0.0395
3       0.0002  0.0002  0.0901  0.0131  0.0003  0.0000  4895217.0000    0.1038
4       0.0012  0.0062  0.0344  0.2257  0.0064  0.0000  12909861.0000   0.2738
5       0.0067  0.0002  0.0000  0.0012  0.1767  0.0003  8722872.0000    0.1850
6       0.0000  0.0000  0.0000  0.0000  0.0005  0.0145  704241.0000     0.0149
Total   0.3719  0.0362  0.1254  0.2483  0.2026  0.0155  47145861.0000
Area    17532402        1708065 5914056 11705815        9553252 732271  47145861
SE      0.0006  0.0004  0.0008  0.0009  0.0006  0.0001
SE area 28163   20205   36321   42619   30626   6370
95% CI area     55200   39601   71190   83534   60027   12486
PA  [%] 97.2743 80.4441 71.8505 90.9010 87.2115 93.2401
UA  [%] 94.4719 73.8257 86.8047 82.4231 95.5136 96.9512
Kappa hat       0.9120  0.7284  0.8491  0.7662  0.9437  0.9690

Overall accuracy [%] = 89.7912
Kappa hat classification = 0.8616

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

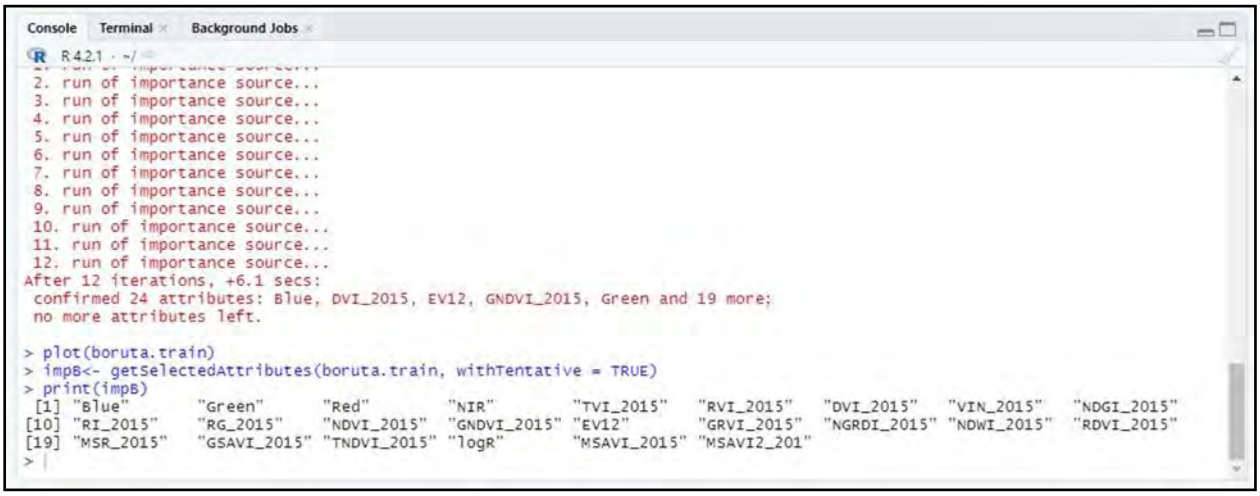
Figure E.7 Error Matrix for the April 2021 dataset

Appendix F. **Computing classifiers in R studio**

```
> #assessing the results
> classifiersResults<-resamples(models)
> summary(classifiersResults)

Call:
summary.resamples(object = classifiersResults)

Models: rf, knn, svmRadial
Number of resamples: 10

Accuracy
               Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
rf        0.9635158 0.9712666 0.9743708 0.9731658 0.9760993 0.9788790    0
knn       0.9519071 0.9562999 0.9588431 0.9589884 0.9622490 0.9675061    0
svmRadial 0.8391552 0.8432464 0.8539634 0.8535258 0.8625871 0.8701504    0

Kappa
               Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
rf        0.9550683 0.9646059 0.9683760 0.9669624 0.9706935 0.9741049    0
knn       0.9407495 0.9461354 0.9492022 0.9495149 0.9537233 0.9601414    0
svmRadial 0.8021500 0.8072652 0.8193211 0.8196598 0.8311258 0.8392183    0
```

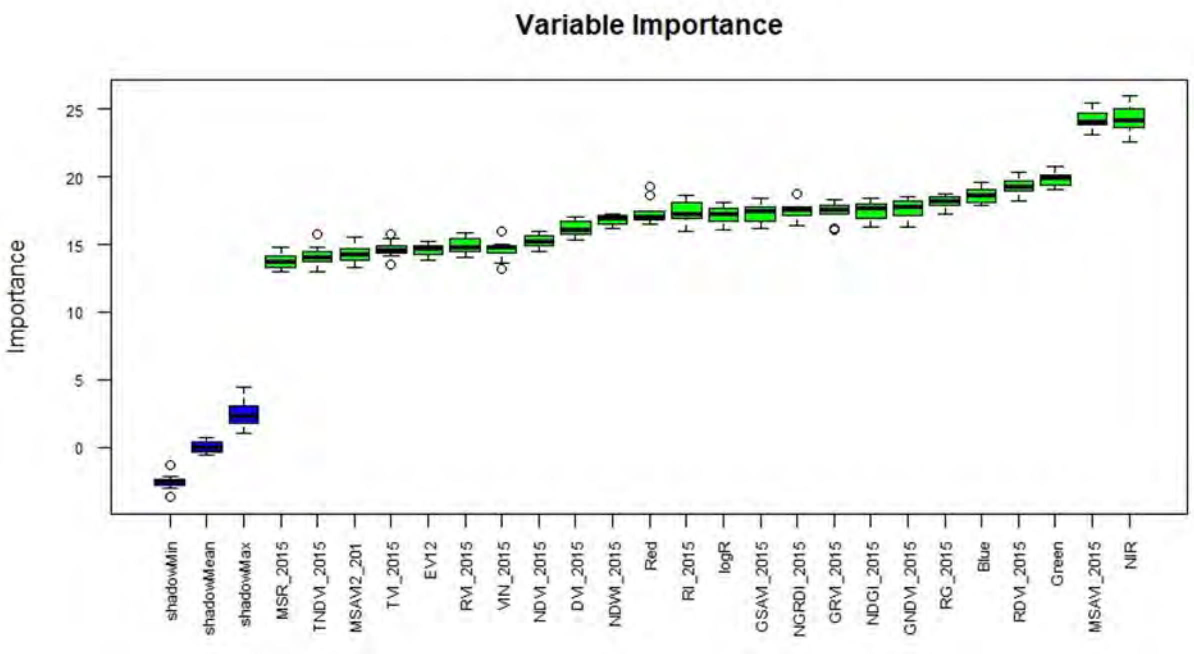
Figure F.1 period Error Matrix for the April 2021 dataset.

Appendix G. **Computing classifiers with indices in R studio**



```
Console   Terminal   Background Jobs
R 4.2.1 · ~/
 2. run of importance source...
 3. run of importance source...
 4. run of importance source...
 5. run of importance source...
 6. run of importance source...
 7. run of importance source...
 8. run of importance source...
 9. run of importance source...
 10. run of importance source...
 11. run of importance source...
 12. run of importance source...
After 12 iterations, +6.1 secs:
 confirmed 24 attributes: Blue, DVI_2015, EV12, GNDVI_2015, Green and 19 more;
 no more attributes left.

> plot(boruta.train)
> impB<- getSelectedAttributes(boruta.train, withTentative = TRUE)
> print(impB)
 [1] "Blue"       "Green"      "Red"        "NIR"        "TVI_2015"   "RVI_2015"   "DVI_2015"   "VIN_2015"   "NDGI_2015"
[10] "RI_2015"    "RG_2015"    "NDVI_2015"  "GNDVI_2015" "EV12"       "GRVI_2015"  "NGRDI_2015" "NDWI_2015"  "RDVI_2015"
[19] "MSR_2015"   "GSAVI_2015" "TNDVI_2015" "logR"       "MSAVI_2015" "MSAVI2_201"
>
```

Appendix H. **Producing graphs from indices in R studio**



Appendix I. **Error matrix for years 2015 with and without indices**

Accuracy assessments with and without indices over the 7 years (2015 – 2021)

Accuracy without indices 2015> ERROR MATRIX (pixel count)

> Reference

V_Classified	1	2	3	4	5	6	Total
1	304	0	0	0	7	0	311
2	0	55	0	0	0	0	55
3	0	16	558	2	0	0	576
4	0	0	0	212	6	0	218
5	6	0	0	4	280	0	290
6	0	0	0	0	0	112	112
Total	310	71	558	218	293	112	1562

> AREA BASED ERROR MATRIX

> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.1138	0.0000	0.0000	0.0000	0.0026	0.0000	10210700.0000	0.1165
2	0.0000	0.4668	0.0000	0.0000	0.0000	0.0000	40926800.0000	0.4668
3	0.0000	0.0016	0.0541	0.0002	0.0000	0.0000	4899700.0000	0.0559
4	0.0000	0.0000	0.0000	0.1145	0.0032	0.0000	10320400.0000	0.1177
5	0.0046	0.0000	0.0000	0.0031	0.2147	0.0000	19491200.0000	0.2223
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0208	1822000.0000	0.0208
Total	0.1184	0.4684	0.0541	0.1177	0.2205	0.0208	87670800.0000	
Area	10384143	41062903	4746584	10322209	19332960	1822000	87670800	
SE	0.0021	0.0004	0.0004	0.0020	0.0029	0.0000		
SE area	184484	33579	35552	176532	253581	0		
95% CI area	361589	65815	69682	346003	497018	0		
PA [%]	96.1165	99.6686	100.0000	97.2307	97.3420	100.0000		
UA [%]	97.7492	100.0000	96.8750	97.2477	96.5517	100.0000		
Kappa hat	0.9745	1.0000	0.9670	0.9688	0.9558	1.0000		

Overall accuracy [%] = 98.4726
Kappa hat classification = 0.9782

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

Accuracy with indices

> ERROR MATRIX (pixel count)
> Reference

V_Classified	1	2	3	4	5	6	Total
1	306	0	0	0	1	0	307
2	0	55	0	0	0	0	55
3	0	16	558	2	0	0	576
4	0	0	0	213	8	0	221
5	4	0	0	3	284	0	291
6	0	0	0	0	0	112	112
Total	310	71	558	218	293	112	1562

> AREA BASED ERROR MATRIX
> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.1244	0.0000	0.0000	0.0000	0.0004	0.0000	10943900.0000	0.1248
2	0.0000	0.4679	0.0000	0.0000	0.0000	0.0000	41020700.0000	0.4679
3	0.0000	0.0015	0.0510	0.0002	0.0000	0.0000	4618800.0000	0.0527
4	0.0000	0.0000	0.0000	0.1248	0.0047	0.0000	11348500.0000	0.1294
5	0.0028	0.0000	0.0000	0.0021	0.2005	0.0000	18008700.0000	0.2054
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0197	1730200.0000	0.0197
Total	0.1272	0.4694	0.0510	0.1271	0.2056	0.0197	87670800.0000	
Area	11155794	41149000	4474462	11139389	18021954	1730200	87670800	
SE	0.0015	0.0004	0.0004	0.0020	0.0025	0.0000		
SE area	128186	31654	33514	178781	218972	0		
95% CI area	251244	62042	65687	350410	429185	0		
PA [%]	97.7810	99.6882	100.0000	98.1894	97.5227	100.0000		
UA [%]	99.6743	100.0000	96.8750	96.3801	97.5945	100.0000		
Kappa hat	0.9963	1.0000	0.9671	0.9585	0.9697	1.0000		

Overall accuracy [%] = 98.8320
Kappa hat classification = 0.9834

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

Appendix J. **Error matrices for 2016 with and without indices**

Accuracy without indices 2016

> ERROR MATRIX (pixel count)
> Reference

V_Classified		1	2	3	4	5	6	Total
1	621	0	0	3	44	5	673	
2	0	204	0	13	0	0	217	
3	0	0	1214	33	27	0	1274	
4	1	17	58	675	76	0	827	
5	8	0	0	44	608	0	660	
6	14	0	0	0	0	614	628	
Total	644	221	1272	768	755	619	4279	

> AREA BASED ERROR MATRIX
> Reference

V_Classified		1	2	3	4	5	6	Area	Wi	
1	0.1620	0.0000	0.0000	0.0008	0.0115	0.0013	15389100.0000		0.1755	
2	0.0000	0.4489	0.0000	0.0286	0.0000	0.0000	41868000.0000		0.4776	
3	0.0000	0.0000	0.0673	0.0018	0.0015	0.0000	6195000.0000		0.0707	
4	0.0001	0.0025	0.0084	0.0975	0.0110	0.0000	10469200.0000		0.1194	
5	0.0014	0.0000	0.0000	0.0078	0.1081	0.0000	10291400.0000		0.1174	
6	0.0009	0.0000	0.0000	0.0000	0.0000	0.0386	3458100.0000		0.0394	
Total	0.1644	0.4514	0.0757	0.1365	0.1321	0.0399	87670800.0000			
Area	14414541		39574986		6637478	11968375		11580079		3495341
	87670800									
SE	0.0019	0.0077	0.0011	0.0080	0.0024	0.0006				
SE area	166212	678029	100032	699487	211892	54905				
95% CI area		325775	1328936	196063	1370994	415308	107613			
PA [%]	98.5120	99.4562	88.9380	71.3964	81.8696	96.7290				
UA [%]	92.2734	94.0092	95.2904	81.6203	92.1212	97.7707				
Kappa hat		0.9075	0.8908	0.9490	0.7871	0.9092	0.9768			

Overall accuracy [%] = 92.2424
Kappa hat classification = 0.8918

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

Accuracy with indices 2016

ErrMatrixCode	Reference	Classified	PixelSum
1	1	1	639
10	1	4	1
15	1	5	2
21	1	6	2
5	2	2	217
14	2	4	4
13	3	3	1269
19	3	4	3
12	4	2	2
18	4	3	13
24	4	4	742
29	4	5	11
11	5	1	2
28	5	4	8
32	5	5	745
16	6	1	2
36	6	6	617

> ERROR MATRIX (pixel count)

> Reference

V_Classified	1	2	3	4	5	6	Total
1	639	0	0	0	2	2	643
2	0	217	0	2	0	0	219
3	0	0	1269	13	0	0	1282
4	1	4	3	742	8	0	758
5	2	0	0	11	745	0	758
6	2	0	0	0	0	617	619
Total	644	221	1272	768	755	619	4279

> AREA BASED ERROR MATRIX

> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.1587	0.0000	0.0000	0.0000	0.0005	0.0005	14003900.0000	0.1597
2	0.0000	0.4752	0.0000	0.0044	0.0000	0.0000	42049000.0000	0.4796
3	0.0000	0.0000	0.0537	0.0005	0.0000	0.0000	4753000.0000	0.0542
4	0.0002	0.0006	0.0005	0.1204	0.0013	0.0000	10785100.0000	0.1230
5	0.0004	0.0000	0.0000	0.0022	0.1501	0.0000	13391600.0000	0.1527
6	0.0001	0.0000	0.0000	0.0000	0.0000	0.0306	2688200.0000	0.0307
Total	0.1594	0.4759	0.0542	0.1276	0.1519	0.0311	87670800.0000	
Area	13975032	41721904	4747488	11183990	13319314	2723072	87670800	
SE	0.0006	0.0031	0.0003	0.0032	0.0009	0.0004		

SE area	52459	272397	27978	283078	80901	31382	
95% CI area		102820	533897	54836	554833	158567	61509
PA [%]	99.5832	99.8636	99.1009	94.3979	98.8184	98.4004	
UA [%]	99.3779	99.0868	98.9860	97.8892	98.2850	99.6769	
Kappa hat		0.9926	0.9826	0.9893	0.9758	0.9798	0.9967

Overall accuracy [%] = 98.8761
Kappa hat classification = 0.9840

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

Appendix K. **Error matrix for 2017 with and without indices**

Accuracy without indices 2017

ErrMatrixCode	Reference	Classified	PixelSum
1	1	1	21763
3	1	2	62
6	1	3	76
10	1	4	90
15	1	5	1385
21	1	6	108
2	2	1	25
5	2	2	1858
9	2	3	61
14	2	4	305
20	2	5	65
4	3	1	133
8	3	2	495
13	3	3	20097
19	3	4	2222
25	3	5	1782
30	3	6	275
7	4	1	27
12	4	2	117
18	4	3	765
24	4	4	14706
29	4	5	1524
11	5	1	1339
17	5	2	796
23	5	3	1024
28	5	4	522
32	5	5	13502
35	5	6	494
16	6	1	395
22	6	2	19
36	6	6	3488

> ERROR MATRIX (pixel count)

> Reference

V_Classified	1	2	3	4	5	6	Total
1	21763	25	133	27	1339	395	23682
2	62	1858	495	117	796	19	3347
3	76	61	20097	765	1024	0	22023
4	90	305	2222	14706	522	0	17845

5	1385	65	1782	1524	13502	0	18258
6	108	0	275	0	494	3488	4365
Total	23484	2314	25004	17139	17677	3902	89520

> AREA BASED ERROR MATRIX

> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.1415	0.0002	0.0009	0.0002	0.0087	0.0026	13480758.0000	0.1540
2	0.0090	0.2684	0.0715	0.0169	0.1150	0.0027	42309900.0000	0.4834
3	0.0003	0.0002	0.0815	0.0031	0.0042	0.0000	7814043.0000	0.0893
4	0.0006	0.0019	0.0140	0.0927	0.0033	0.0000	9840321.0000	0.1124
5	0.0109	0.0005	0.0140	0.0120	0.1064	0.0000	12592143.0000	0.1439
6	0.0004	0.0000	0.0011	0.0000	0.0019	0.0136	1484631.0000	0.0170
Total	0.1627	0.2712	0.1829	0.1248	0.2394	0.0189	87521796.0000	
Area	14240668	23736131	16011571	10926257	20955792	1651376	87521796	
SE	0.0012	0.0042	0.0030	0.0016	0.0036	0.0006		
SE area	104672	363680	262816	140004	315258	56805		
95% CI area	205156	712813	515120	274407	617905	111337		
PA [%]	86.9930	98.9514	44.5345	74.2191	44.4366	71.8397		
UA [%]	91.8968	55.5124	91.2546	82.4096	73.9511	79.9084		
Kappa hat	0.9032	0.3896	0.8930	0.7990	0.6575	0.7952		

Overall accuracy [%] = 70.3985

Kappa hat classification = 0.6199

Area unit = metre^2

SE = standard error

CI = confidence interval

PA = producer's accuracy

UA = user's accuracy

Accuracy with indices 2017

ErrMatrixCode	Reference	Classified	PixelSum
1	1	1	2071
6	1	3	3
10	1	4	2
15	1	5	13
21	1	6	1
2	2	1	1
5	2	2	185
9	2	3	3
14	2	4	6

4	3	1	1
8	3	2	3
13	3	3	2194
19	3	4	21
25	3	5	23
7	4	1	2
12	4	2	1
18	4	3	18
24	4	4	1501
29	4	5	15
11	5	1	4
17	5	2	1
23	5	3	11
28	5	4	7
32	5	5	1573
27	6	3	2
36	6	6	350

> ERROR MATRIX (pixel count)

> Reference

V_Classified	1	2	3	4	5	6	Total
1	2071	1	1	2	4	0	2079
2	0	185	3	1	1	0	190
3	3	3	2194	18	11	2	2231
4	2	6	21	1501	7	0	1537
5	13	0	23	15	1573	0	1624
6	1	0	0	0	0	350	351
Total	2090	195	2242	1537	1596	352	8012

> AREA BASED ERROR MATRIX

> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.1715	0.0001	0.0001	0.0002	0.0003	0.0000	15095900.0000	0.1722
2	0.0000	0.4566	0.0074	0.0025	0.0025	0.0000	41115800.0000	0.4690
3	0.0001	0.0001	0.0986	0.0008	0.0005	0.0001	8788000.0000	0.1002
4	0.0002	0.0005	0.0017	0.1195	0.0006	0.0000	10724100.0000	0.1223
5	0.0010	0.0000	0.0017	0.0011	0.1180	0.0000	10684400.0000	0.1219
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0144	1262600.0000	0.0144
Total	0.1728	0.4573	0.1095	0.1240	0.1219	0.0145	87670800.0000	
Area	15152708	40094747	9596555	10873427	10686482	1266881	87670800	
SE	0.0004	0.0055	0.0043	0.0025	0.0025	0.0001		
SE area	33699	479142	376308	222637	222908	6630		
95% CI area	66051	939119	737563	436369	436900	12995		
PA [%]	99.2417	99.8480	90.0558	96.3166	96.8407	99.3782		
UA [%]	99.6152	97.3684	98.3416	97.6578	96.8596	99.7151		
Kappa hat	0.9953	0.9515	0.9814	0.9733	0.9642	0.9971		

Overall accuracy [%] = 97.8600
Kappa hat classification = 0.9701

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

Appendix L. **Error matrices for 2018 with and without indices**

Accuracy without indices 2018

> ERROR MATRIX (pixel count)
> Reference

V_Classified	1	2	3	4	5	6	Total
1	3081	2	0	10	178	1	3272
2	26	277	10	14	9	0	336
3	1	1	1380	36	314	0	1732
4	4	30	4	1032	53	0	1123
5	85	4	4	28	1607	0	1728
6	7	0	0	0	4	329	340
Total	3204	314	1398	1120	2165	330	8531

> AREA BASED ERROR MATRIX
> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.1593	0.0001	0.0000	0.0005	0.0092	0.0001	14830400.0000	0.1692
2	0.0375	0.3997	0.0144	0.0202	0.0130	0.0000	42508300.0000	0.4849
3	0.0000	0.0000	0.0589	0.0015	0.0134	0.0000	6483300.0000	0.0740
4	0.0004	0.0030	0.0004	0.1021	0.0052	0.0000	9744600.0000	0.1111
5	0.0072	0.0003	0.0003	0.0024	0.1363	0.0000	12849600.0000	0.1466
6	0.0003	0.0000	0.0000	0.0000	0.0002	0.0138	1254600.0000	0.0143
Total	0.2047	0.4032	0.0741	0.1268	0.1773	0.0139	87670800.0000	
Area	17950374	35346916	6495259	11114439	15545269	1218543	87670800	
SE	0.0072	0.0101	0.0046	0.0054	0.0045	0.0001		
SE area	627425	885045	400260	473188	397188	12880		
95% CI area	1229754	1734689	784510	927448	778488	25245		
PA [%]	77.7961	99.1431	79.5300	80.5706	76.8712	99.6280		
UA [%]	94.1626	82.4405	79.6767	91.8967	92.9977	96.7647		
Kappa hat	0.9266	0.7058	0.7805	0.9072	0.9149	0.9672		

Overall accuracy [%] = 87.0224
Kappa hat classification = 0.8208

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

Accuracy with indices 2018

ErrMatrixCode	Reference	Classified	PixelSum
1	1	1	3179
10	1	4	1
15	1	5	24
5	2	2	310
14	2	4	4
13	3	3	1393
19	3	4	2
25	3	5	3
7	4	1	4
12	4	2	3
18	4	3	2
24	4	4	1104
29	4	5	7
11	5	1	13
28	5	4	2
32	5	5	2150
36	6	6	330

> ERROR MATRIX (pixel count)
> Reference

V_Classified	1	2	3	4	5	6	Total
1	3179	0	0	4	13	0	3196
2	0	310	0	3	0	0	313
3	0	0	1393	2	0	0	1395
4	1	4	2	1104	2	0	1113
5	24	0	3	7	2150	0	2184
6	0	0	0	0	0	330	330
Total	3204	314	1398	1120	2165	330	8531

> AREA BASED ERROR MATRIX
> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.1905	0.0000	0.0000	0.0002	0.0008	0.0000	16786900.0000	0.1915
2	0.0000	0.4691	0.0000	0.0045	0.0000	0.0000	41522600.0000	0.4736
3	0.0000	0.0000	0.0558	0.0001	0.0000	0.0000	4901600.0000	0.0559
4	0.0001	0.0004	0.0002	0.1153	0.0002	0.0000	10191700.0000	0.1162
5	0.0016	0.0000	0.0002	0.0005	0.1467	0.0000	13065500.0000	0.1490
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0137	1202500.0000	0.0137
Total	0.1922	0.4695	0.0562	0.1206	0.1477	0.0137	87670800.0000	
Area	16850342	41161248	4930834	10577181	12948695	1202500	87670800	

SE	0.0004	0.0026	0.0002	0.0026	0.0005	0.0000	
SE area	37422	229765	17306	231499	41512	0	
95% CI area		73347	450340	33919	453737	81364	0
PA [%]	99.0936	99.9110	99.2646	95.5764	99.3312	100.0000	
UA [%]	99.4681	99.0415	99.8566	99.1914	98.4432	100.0000	
Kappa hat		0.9934	0.9819	0.9985	0.9908	0.9817	1.0000

Overall accuracy [%] = 99.1102
Kappa hat classification = 0.9873

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

Appendix M. **Error matrices for 2019 with and without indices**

Accuracy without indices 2019

> ERROR MATRIX (pixel count)

> Reference

V_Classified	1	2	3	4	5	6	Total
1	33454	9	5	39	60	5	33572
2	0	2876	0	18	0	0	2894
3	19	25	10624	245	318	0	11231
4	61	111	301	15210	37	3	15723
5	80	0	415	30	13096	9	13630
6	2	0	0	1	0	1638	1641
Total	33616	3021	11345	15543	13511	1655	78691

> AREA BASED ERROR MATRIX

> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.2986	0.0001	0.0000	0.0003	0.0005	0.0000	14078898.0000	0.2996
2	0.0000	0.0141	0.0000	0.0001	0.0000	0.0000	664632.0000	0.0141
3	0.0002	0.0003	0.1379	0.0032	0.0041	0.0000	6850008.0000	0.1458
4	0.0012	0.0022	0.0058	0.2954	0.0007	0.0001	14346135.0000	0.3053
5	0.0013	0.0000	0.0065	0.0005	0.2052	0.0001	10032624.0000	0.2135
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0215	1014399.0000	0.0216
Total	0.3013	0.0166	0.1503	0.2995	0.2105	0.0218	46986696.0000	
Area	14156782	780800	7061994	14070678	9892439	1024003	46986696	
SE	0.0002	0.0002	0.0006	0.0005	0.0004	0.0001		
SE area	11054	10177	26045	22951	20838	3065		
95% CI area	21665	19947	51048	44983	40843	6008		
PA [%]	99.1003	84.5925	91.7558	98.6311	97.4437	98.8810		
UA [%]	99.6485	99.3780	94.5953	96.7373	96.0822	99.8172		
Kappa hat	0.9950	0.9937	0.9364	0.9534	0.9504	0.9981		

Overall accuracy [%] = 97.2613

Kappa hat classification = 0.9635

Area unit = metre^2

SE = standard error

CI = confidence interval

PA = producer's accuracy

UA = user's accuracy

Accuracy with indices 2019

ErrMatrixCode Reference Classified PixelSum

> ERROR MATRIX (pixel count)
> Reference

V_Classified	1	2	3	4	5	6	Total
1	2991	3	0	2	3	1	3000
2	0	258	1	0	0	0	259
3	2	4	979	8	20	0	1013
4	3	6	19	1390	5	1	1424
5	6	0	27	1	1201	0	1235
6	1	0	0	0	0	155	156
Total	3003	271	1026	1401	1229	157	7087

> AREA BASED ERROR MATRIX
> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.1808	0.0002	0.0000	0.0001	0.0002	0.0001	15900300.0000	0.1814
2	0.0000	0.4684	0.0018	0.0000	0.0000	0.0000	41225500.0000	0.4702
3	0.0002	0.0004	0.0893	0.0007	0.0018	0.0000	8102900.0000	0.0924
4	0.0003	0.0006	0.0020	0.1483	0.0005	0.0001	13316300.0000	0.1519
5	0.0005	0.0000	0.0021	0.0001	0.0916	0.0000	8256400.0000	0.0942
6	0.0001	0.0000	0.0000	0.0000	0.0000	0.0099	869400.0000	0.0099
Total	0.1818	0.4696	0.0952	0.1492	0.0941	0.0100	87670800.0000	
Area	15942336	41170332	8348288	13079631	8251734	878478	87670800	
SE	0.0003	0.0018	0.0020	0.0007	0.0007	0.0001		
SE area	30661	161858	173960	59274	57051	12108		
95% CI area	60095	317241	340961	116178	111820	23731		
PA [%]	99.4371	99.7474	93.8029	99.3786	97.3020	98.3322		
UA [%]	99.7000	99.6139	96.6436	97.6124	97.2470	99.3590		
Kappa hat	0.9963	0.9927	0.9629	0.9719	0.9696	0.9935		

Overall accuracy [%] = 98.8255
Kappa hat classification = 0.9834
Area unit = metre^2
SE = standard error
CI = confidence interval
PA = producer's accuracy
UA = user's accuracy

Appendix N. **Error matrices for 2020 with and without indices**

Accuracy without indices 2020

> ERROR MATRIX (pixel count)
> Reference

V_Classified		1	2	3	4	5	6	Total
1	3013	2	28	3	186	0	3232	
2	0	245	1	59	0	0	305	
3	9	2	595	47	263	3	919	
4	15	31	135	1245	63	3	1492	
5	34	2	254	54	723	2	1069	
6	8	0	0	0	0	155	163	
Total	3079	282	1013	1408	1235	163	7180	

> AREA BASED ERROR MATRIX
> Reference

V_Classified		1	2	3	4	5	6	Area	Wi
1	0.2903	0.0002	0.0027	0.0003	0.0179	0.0000	14660900.0000		0.3114
2	0.0000	0.0161	0.0001	0.0039	0.0000	0.0000	945600.0000		0.0201
3	0.0014	0.0003	0.0936	0.0074	0.0414	0.0005	6808300.0000		0.1446
4	0.0031	0.0064	0.0278	0.2560	0.0130	0.0006	14443400.0000		0.3068
5	0.0065	0.0004	0.0487	0.0104	0.1387	0.0004	9653600.0000		0.2051
6	0.0006	0.0000	0.0000	0.0000	0.0000	0.0114	566600.0000		0.0120
Total	0.3019	0.0234	0.1729	0.2779	0.2109	0.0129	47078400.0000		
Area	14214208		1101628	8138720	13084669		9931055	608119	47078400
SE	0.0020	0.0013	0.0042	0.0035	0.0042	0.0006			
SE area	94149	60210	198564	162740	196680	26467			
95% CI area		184531	118011	389185	318970	385492	51875		
PA [%]	96.1536	68.9507	54.1607	92.1101	65.7438	88.5996			
UA [%]	93.2240	80.3279	64.7443	83.4450	67.6333	95.0920			
Kappa hat		0.9029	0.7986	0.5738	0.7707	0.5898	0.9503		

Overall accuracy [%] = 80.6212
Kappa hat classification = 0.7422

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

Accuracy with indices 2020

> ERROR MATRIX (pixel count)

> Reference

V_Classified	1	2	3	4	5	6	Total
1	3072	0	0	0	6	0	3078
2	0	271	0	0	0	0	271
3	0	1	985	10	8	0	1004
4	4	9	9	1397	10	3	1432
5	3	1	19	1	1211	0	1235
6	0	0	0	0	0	160	160
Total	3079	282	1013	1408	1235	163	7180

> AREA BASED ERROR MATRIX

> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.1902	0.0000	0.0000	0.0000	0.0004	0.0000	16708600.0000	0.1906
2	0.0000	0.4752	0.0000	0.0000	0.0000	0.0000	41662500.0000	0.4752
3	0.0000	0.0001	0.0766	0.0008	0.0006	0.0000	6846500.0000	0.0781
4	0.0004	0.0010	0.0010	0.1478	0.0011	0.0003	13279500.0000	0.1515
5	0.0002	0.0001	0.0015	0.0001	0.0974	0.0000	8706700.0000	0.0993
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0053	467000.0000	0.0053
Total	0.1909	0.4763	0.0791	0.1486	0.0994	0.0056	87670800.0000	
Area	16734273	41759830	6934345	13030173	8717359	494820	87670800	
SE	0.0003	0.0003	0.0006	0.0007	0.0006	0.0002		
SE area	25858	29425	50674	58727	50706	16051		
95% CI area	50682	57673	99321	115105	99384	31459		
PA [%]	99.6520	99.7669	96.8647	99.4226	97.9368	94.3777		
UA [%]	99.8051	100.0000	98.1076	97.5559	98.0567	100.0000		
Kappa hat	0.9976	1.0000	0.9795	0.9713	0.9784	1.0000		

Overall accuracy [%] = 99.2519

Kappa hat classification = 0.9893

Area unit = metre^2

SE = standard error

CI = confidence interval

PA = producer's accuracy

UA = user's accuracy

Appendix O. **Error matrices for 2021 with and without indices**

Accuracy without indices 2021

> ERROR MATRIX (pixel count)
> Reference

V_Classified	1	2	3	4	5	6	Total
1	34981	23	29	48	761	32	35874
2	96	2691	20	900	0	0	3707
3	7	11	8609	946	46	0	9619
4	97	466	2005	15051	210	0	17829
5	502	17	0	35	11692	15	12261
6	0	0	0	0	25	1158	1183
Total	35683	3208	10663	16980	12734	1205	80473

> AREA BASED ERROR MATRIX
> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.3511	0.0002	0.0003	0.0005	0.0076	0.0003	16974441.0000	0.3600
2	0.0007	0.0210	0.0002	0.0070	0.0000	0.0000	1364994.0000	0.0290
3	0.0001	0.0001	0.0832	0.0091	0.0004	0.0000	4381623.0000	0.0929
4	0.0016	0.0075	0.0322	0.2420	0.0034	0.0000	13512366.0000	0.2866
5	0.0088	0.0003	0.0000	0.0006	0.2042	0.0003	10095462.0000	0.2141
6	0.0000	0.0000	0.0000	0.0000	0.0004	0.0170	816975.0000	0.0173
Total	0.3622	0.0291	0.1159	0.2592	0.2160	0.0175	47145861.0000	
Area	17077291	1373948	5462200	12220805	10184415	827202	47145861	
SE	0.0005	0.0004	0.0007	0.0009	0.0006	0.0001		
SE area	24309	19485	34914	40632	25982	5385		
95% CI area	47646	38190	68431	79638	50924	10554		
PA [%]	96.9235	72.1193	71.7943	93.3405	94.5264	96.6765		
UA [%]	97.5107	72.5924	89.4999	84.4186	95.3593	97.8867		
Kappa hat	0.9610	0.7177	0.8812	0.7897	0.9408	0.9785		

Overall accuracy [%] = 91.8383
Kappa hat classification = 0.8893

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

Accuracy with indices 2021

> ERROR MATRIX (pixel count)
> Reference

V_Classified	1	2	3	4	5	6	Total
1	3183	0	0	2	1	1	3187
2	1	289	0	0	0	0	290
3	0	0	953	6	0	0	959
4	0	1	9	1504	0	0	1514
5	7	0	0	0	1147	0	1154
6	0	0	0	0	0	115	115
Total	3191	290	962	1512	1148	116	7219

> AREA BASED ERROR MATRIX
> Reference

V_Classified	1	2	3	4	5	6	Area	Wi
1	0.1954	0.0000	0.0000	0.0001	0.0001	0.0001	17149700.0000	0.1956
2	0.0016	0.4697	0.0000	0.0000	0.0000	0.0000	41321300.0000	0.4713
3	0.0000	0.0000	0.0625	0.0004	0.0000	0.0000	5514100.0000	0.0629
4	0.0000	0.0001	0.0008	0.1348	0.0000	0.0000	11900400.0000	0.1357
5	0.0008	0.0000	0.0000	0.0000	0.1230	0.0000	10847100.0000	0.1237
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0107	938200.0000	0.0107
Total	0.1977	0.4698	0.0633	0.1354	0.1230	0.0108	87670800.0000	
Area	17336460	41186673	5550343	11867059	10786684	943581	87670800	
SE	0.0017	0.0016	0.0003	0.0003	0.0003	0.0001		
SE area	145030	142704	27394	29485	25381	5381		
95% CI area	284258	279700	53693	57791	49747	10547		
PA [%]	98.7986	99.9809	98.7254	99.6186	99.9501	99.4297		
UA [%]	99.8745	99.6552	99.3743	99.3395	99.3934	100.0000		
Kappa hat	0.9984	0.9935	0.9933	0.9924	0.9931	1.0000		

Overall accuracy [%] = 99.6089
Kappa hat classification = 0.9944

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