MONITORING AND EVALUATING URBAN LAND USE LAND COVER CHANGE USING MACHINE LEARNING CLASSIFICATION TECHNIQUES: A CASE STUDY OF POLOKWANE MUNICIPALITY

By

Tshivhase Funani

2604729

"The research report is submitted to the Faculty of Science, University of Witwatersrand Johannesburg in fulfilment of the Degree

in Master of Science (Geographical Information System and Remote sensing) at the school of Geography, Archaeology & Environmental Studies."

Supervisor: Prof Paida Mhangara (Wits University)

March 2023

DECLARATION

Presenting this report to the University of the Witwatersrand to meet the requirements for the degree. I agree that the head of my department or one of his or her representatives may give permission for a lot of copies of this report to be made for academic research. It is understood that any copying or dissemination of this report for the purpose of making a financial gain will not be permitted without my consent.

"SCHOOL OF GEOGRAPHY, ARCHAEOLOGY AND ENVIRONMENTAL STUDIES
The University of Witwatersrand Braamfontein Campus
1 Jan Smuts Avenue
Braamfontein
Johannesburg, South Africa"

Tshivhase Funani (Miss) ____ September 2023____

Surname & Initials (Title) **Date**

ABSTRACT

Remote sensing is one of the tools which is very important to produce Land use and land cover maps through the process of image classification. Image classification requires quality multispectral imagery and secondary data, a precise classification technique, and user experience skill. Remote sensing and GIS were used to identify and map land-use/land-cover in the study region. Big Data issues arise when classifying a huge number of satellite images and features, which is a very intensive process. This study primarily uses GEE to evaluate the two classifiers, Support Vector Machine, and gradient boosting, using multi-temporal Landsat-8 images, and to assess their performance while accounting for the impact of data dimension, sample size, and quality. Land use/Landcover (LULC) classification, accuracy assessment, and landscape metrics comprise this study. Gradient Tree Boost and SVM algorithms were used in 2008, 2013, 2017, and 2022. Google Earth Engine was used for supervised classification. The results of change detection showed that urbanization has occurred and most of the encroachments were on agricultural land. In this study, XG boost, and support vector machine (SVM)) were used and compared for image classification to oversight spatio-temporal land use changes in Polokwane Municipality. The Google Earth Engine has been utilized to pre-process the Landsat imagery, and then upload it for classification. Each classification method was evaluated using field observations and high-resolution Google Earth imagery. LULC changes were assessed, utilizing Geographic Information System (GIS) techniques, as well as the dynamics of change in LULCC were analysed using landscape matrix analysis over the last 15 years in four different periods: 2008–2013, 2018 and 2022. The results showed that XGBoost performed better than SVM both in overall accuracies and Kappa statistics as well as F-scores and the ratio of Z-score.

The overall accuracy of gradient boosting in 2008 was 0.82, while SVM showed results of 0.82 overall accuracy and kappa statistics of 0.69. The average F-score for SVM in 2008 was from 0.58-1.00, in 2013 an average of 0.86-0.97, and in 2022 it was 0.76. Z values were not statistically significant as all values were below the z score of 1.96. The ratios for the two classifiers were also taken to know which classifier performs the best. The results showed 212:212 which indicates that during 2008 SVM and XG boost performed the same way as they classified the same number of cases. During 2013 the ratio was 345:312 which shows that XGBoost performed better than SVM. The results of 2017 show 374:316 which shows that XGBoost performed better than SVM. Lastly, in 2022 the ratio was 298:277 which shows that XGBoost performed better than SVM. Overall z scores result show that XGBoost performs better than SVM. Overall, this study offers useful insight into LULC changes that might aid shareholders and decision makers in making informed decisions about controlling land use changes and urban growth.

Keywords

Accuracy assessment, Geographic Information Systems (GIS), Land Use Land Cover (LULC), Remote Sensing, Google Earth Engine(GEE), urbanization.

ACKNOWLEDGEMENTS

Firstly, I would like to begin by expressing my most sincere gratitude to the Almighty God for giving me the dream, the courage, the health, and other resources necessary to finish this dissertation research.

For the duration of this dissertation, my supervisor, Professor Paida Mhangara, provided me with academic assistance as well as essential guidance, and I would want to express my gratitude for this kindness.

I would also like to thank Dr. Makananise Oscar for his support in organizing my work and proofreading my study. I would not have done either of those things without him.

For their encouragement, I would want to extend my appreciation to my mother, Tshivhase Mukondeleli, my sister, Khodani Ramuhuyu, and my friend, William Mathosa. I would also like to thank my grandmother, Rachel Tshivhase.

In addition, I would like to express my gratitude to a few of the Wits doctoral students studying geographic information systems and remote sensing who contributed to the success of the image categorization.

Firstly, I would like to give special thanks to God the Almighty for giving me a dream, strength, health, and other resources to successfully complete this dissertation.

I would also like to acknowledge some of the Wits PHD students at GIS and Remote Sensing who assisted in breaking through the image classification.

TABLE OF CONTENTS

List of Figures

Figure 1: Methodology Flow chart 22
Figure 2:XG Boost classification results (2008, 2013, 2017, 2022).. 43
Figure 3:SVM classification results (2008, 2013, 2017, 2022) 44

List of Tables

Table 1: Landsat 8 Spectral properties with important bands for this study 24
Table 2:Landset5 TM   Spectral properties with important bands for this study”........................ 25
Table 3:Landscape Matrix Analysis for SVM Classifier (2008, 2013, 2018 and 2022) 36
Table 4:Landscape Matrix Analysis for XGBoost Classifier (2008, 2013, 2018 and 2022)........ 39
Table 5:Accuracy assessments, Confusion Matrix and F-scores for the validation samples results of XGBoost classification (2008, 2013, 2018 and 2022)... 44
Table 6:Accuracy assessments, Confusion Matrix and F-scores for the validation samples results of SVM classification (2008, 2013, 2018 and 2022)... 46
Table 7:Accuracy assessments, Confusion Matrix and F-scores for the training samples results of XGBoost classification (2008, 2013, 2018 and 2022).. 47
Table 8:Accuracy assessments, Confusion Matrix and F-scores for the training samples results of SVM classification (2008, 2013, 2018 and 2022) 51
4.3.7 Table 9:McNemar's Test statistics between XG Boost and Support Vector Machine 54

DECLARATION ii

ABSTRACT iii

ACKNOWLEDGEMENTS v

ACRONYMS ix

CHAPTER 1: INTRODUCTION / BACKGROUND 1

1.1 INTRODUCTION 1

1.2 PROBLEM STATEMENT 4

1.4 RESEARCH QUESTIONS 6

1.5 THE AIM 6

1.6 MAIN OBJECTIVE 6

The main objective of this study is to track and record any changes in urban Land use Land Cover change of Polokwane Municipality between 2008 to 2022 6

1.6.1 Specific Objectives 6

1.7 SIGNIFICANCE OF THE STUDY 6

1.8 RESEARCH STRUCTURE 7

CHAPTER 2: LITERATURE REVIEW 9

2.1. CONCEPTS OF URBANIZATION 10

2.2. URBANIZATION AND LAND USE LAND COVER CHANGE 10

2.3. LAND USE LAND COVER CLASSIFICATIONS 11

2.4. THE USE OF SVM AND GRADIENT BOOSTING FOR LULC CLASSIFICATION 13

2.5. CLASSIFICATION ON GOOGLE EARTH ENGINE 15

2.6. SATELITE REMOTE SENSING FOR URBAN ANALYSIS 16

2.7. CHANGES IN LANDSCAPE PATTERNS USING LANDSCAPE FRAGMENTATION ANALYSIS 17

CHAPTER 3: RESEARCH METHODOLOGY 18

3.1 GEOGRAPHIC LOCATION/ STUDY AREA 18

3.2 TOPOGRAPHY OF THE STUDY AREA 20

3.3 SOILS AND GEOLOGY OF THE STUDY AREA 20

3.4 METHODOLOGY FLOW CHART 21

3.4.1 Data Acquisition 22

3.5 FIELD DATA COLLECTION, GROUND REFERENCE AND PRE-PROCESSING 22

3.6 DATA SETS 23

3.6.1 Landsat 8 OLI (11 bands) 23

3.6.2 Landsat 5 TM (seven bands) 25

3.7 DATA ANALYSIS 26
3.7.1 Support Vector Machine 26
3.7.2 Extreme gradient boosting classification 26
3.7.3 Landscape Fragmentation Analysis 27
3.7.4. Google Earth Engine 27
3.8 MODEL VALIDITY AND RELIABILITY (ACCURACY ASSESSMENT) 28
3.8.1 McNemar’s Test Statistics 29
3.8.2 Kappa Statistics 30
3.8.3 Assessment of Pixel-Based Accuracy 30
3.8.4. Assessment of Producers and users Accuracy 30
CHAPTER 4: RESULTS INTERPRETATION AND DISCUSSION 32
4.1. LAND COVER/ USE CLASSIFICATION 32
4.2. RESULTS FROM OBSERVATIONS AND STRUCTUAL MEETINGS AND DIALOGUES WITH THE PUBLIC 33
4.2.1. Public meetings: 33
4.2.2. Formal meetings: 34
4.3. RESULTS FOR QUANTIFYING LAND USE LAND COVER CHANGE USING LANDSCAPE METRICS / ACCURACY ASSESSMENTS AND F-SCORES 35
4.3.1. Landscape Matrix Analysis / Accuracy assessments and F-scores for XGradient_ Boost 36
4.4. RESULTS AND DATA INTERPRETATION OF LULC CLASSIFICATION AND ACCURACIES 42
4.4.1 Results for Gradient Boosting classification in 2008, 2013, 2017 and 2022 43
4.4.2 Results for Support Vector Machine classification in 2008, 2013, 2017 and 2022 44
4.4.3 Results for Validation test of XGBoost classifier (2008, 2013, 2018 and 2022) 44
4.4.4. Results for Validation test of SVM classifier (2008, 2013, 2018 and 2022) 46
Table 6:Accuracy assessments, Confusion Matrix and F-scores for the validation samples results of SVM classification (2008, 2013, 2018 and 2022) 46
4.4.5 Accuracy assessment results for XGBoost (2008, 2013, 2018 and 2022) 47
4.5 DISCUSSION 55
CHAPTER 5: CONCLUSION AND RECOMMENDATION 60
5.2 RECOMMENDATION 61
5.3 LIMITATION OF THE STUDY 63

ACRONYMS

RS	Remote Sensing
GIS	Geographical Information System
LULC	Land Use Land Cover Change
NDVI	Normalized Difference Vegetation Index
CTA	Classification Tree Analysis
SVM	Support Vector Machine
USGS	United States Geological Survey
SDF	Spatial Development Framework
UN	United Nations
AVHRR	Advanced Very High-Resolution Radiometer
OCA	Overall core area
SPA	Smallest patch area
GPA	Greatest patch area
MPA	Mean patch area.
NP	Number of patches
WHO	World Health Organization
GEE	Google Earth Engine
PD	Patch Density
LA	Landscape Proportions

CHAPTER 1: INTRODUCTION / BACKGROUND

1.1 INTRODUCTION

Urbanization is a complex socioeconomic process that shifts the distribution of a population from dispersed rural settlements to dense urban settlements (United Nations, 2019). In spatial terms, the urbanization process is manifested in the physical development of urban settlements and the transition of landscapes into urban forms (Foley et al.,2005 and Grimm et al.,2008). According to Thapa et al., (2011) supported by Pauleit et al., (2010), In the Global South, rapid and unplanned urban sprawl leads to problems such as fragmented landscape, reduction in arable land, increase in urban poverty, and environmental degradation, which pose a huge threat to sustainable development in these regions (Thapa et al., 2011; Pauleit et al., 2010; Lu et al., 2000-2016). According to United Nations, (2015), by 2030, Sustainable Development Goal 11 of the United Nations intends to make cities and human settlements more inclusive, safe, resilient, and sustainable. Building policies to promote the sustainable development of cities, especially in developing countries, need accurate and timely monitoring and understanding of the spatial growth of urban settlements (Inostroza et al.,2019).

Geospatial techniques have enabled the analysis and forecasting of urban growth at regional and global scales. These methods are useful for observing and understanding the dynamics of urban landscapes. Models include landscape matrix, support vector machine and XG boosting machine learning algorithms. Polokwane is the largest city in Limpopo and has seen to have massive urban growth in recent decades not only in the city’s centre, but also in the surrounding suburbs (Lu et al., 2019). Furthermore, the urban land expansion rate is higher than the population increase rate, the population density in the urban area will significantly decline, and the phenomena of urban growth will occur. Due to institutional inefficiency and governance failure, rural lands have been converted into residential and industrial areas without considering the Land use planning schemes in Polokwane (Akhtar & Dhanani, 2013). According to Akhtar and Dhanani (2013), the massive conversion of rural lands for urban areas has caused the sprawl phenomenon since 2000, which has led to loss of agricultural lands, an increase in commuting costs, and flooding (Akhtar & Dhanani, 2013). The unplanned urban growth has also resulted in a range of social problems such as a lack of health care, shortage of education facilities and infrastructures, an increase in criminal incidents, and sociocultural imbalance (Schetke et al., 2016; Qureshi et al., 2010). The introduction of new urban forms and structures that adapt to climate change issues can mitigate the environmental problems caused by dispersed urban area growth and create more efficient urban economies (D'Acci, 2019). Therefore, the spatiotemporal modelling of urban sprawl is crucial to better understand the changing urban patterns of Polokwane divisions, thus helping local governments in prioritizing the demands of the local population and formulating strategies and practical solutions to achieve the goal of urban sustainable development.

Due to the significant population expansion that has occurred in metropolitan areas over the past few decades (Su et al., 2014; Dadashpoor and Alidadi, 2017; Dadashpoor et al., 2016), land use changes and urbanization have reshaped the landscape patterns of these places (Salvati et al., 2018). According to Liu et al.'s (2010) definition, landscape pattern is the spatial arrangement of a variety of landscape features that range in size and shape. Natural elements and human activities (Déjeant-Pons, 2006; Gustafson, 1998; Dadashpoor and Salarian, 2018; Dadashpoor and Zarei, 2016) might potentially influence the composition and operation of an ecosystem. This alteration in the landscape pattern is a significant example of this interaction. Landscape or spatial heterogeneity (McGarial and Marks, 1995) refers to the characteristics of landscape patterns such as the configuration and composition of elements (Kwok et al., 2016; Zhang et al., 2006). These characteristics influence the ecological processes that led to biodiversity and have profound effects on the ecological, social, and economic functions of systems (Tischendorf and Walz, 2014; Lausch et al., 2015). Landscape or spatial heterogeneity refers to the characteristics.

Landscape metrics are among the tools widely used for analysing (Wu et al., 2017), monitoring and planning of landscape patterns (Peng et al., 2010). These metrics are often used for the diagnosis and measurement of spatial changes in the composition and configuration of the landscape (Herold et al., 2002). They can also provide analyses such as fragmentation, diversity, changes in landscape patterns, and descriptions of the landscape structure. For this purpose, tools and methods such as Remote Sensing (RS), Geographic Information System (GIS), land-use models and statistical methods have been used to investigate the changes in landscape patterns and their relationships with driving forces (Deng et al., 2009; Feng et al., 2018; Steiniger and Hay, 2009)

The technologies of remote sensing (RS) and geographic information systems (GIS) can give advanced assistance for effective monitoring of the spatiotemporal LULC changes, ecological conditions, and thermal trends at the local, regional, and global scales (Dewan et al, 2021).According to Rasool et al. 2021, remotely sensed Landsat satellite pictures used for earth observation are valuable data sources for identifying the LULC and LST of the earth's surfaces. (Naim et al., 2021; Liu et al., 2012) Landsat data are often employed because of their adequate temporal and spatial resolutions, their capacity for historical preservation, and their ability to provide global coverage. Understanding the LULC transition is critical for both the identification of Spatio-temporal shifts in land features and the successful long-term implementation of community development initiatives. Many researchers all over the world employ the application of remote sensing (RS) in conjunction with geographic information systems (GIS) to determine the shifting patterns of LULC (Selcuk et al., 2003). Thematic Mapper (TM), Operational Land Imager (OLI), and Thermal Infrared Sensor (TIRS) sensors are built into Landsat satellite images. These sensors are used to collect data for LULC and LST calculation. These sensors each include multiple bands, some of which are optimized for detecting thermal infrared radiation for the purpose of calculating LST, while other bands are employed for the purpose of estimating LULC. Researchers (Hu and Jia 2010; Keshtkar and Voigt, 2016) from all over the world have used the data obtained from Landsat's sensors to estimate the LULC and LST movements over the course of several years. Therefore, monitoring images generated from remote sensing systems are fundamental as satellite imagery enables large-scale monitoring of deforestation and understanding of landscape dynamics (Noma et al., 2013; Oliveira, 2017) . One of the concerning aspects of analysing data on a large scale is the data size. The Google Earth Engine (GEE) cloud-based platform is of great help in this regard because it allows analysis of large spatial regions using high-performance computing resources without the need to download data (Fadli et al., 2019).

The classification of images acquired by remote sensing in the context of land use mapping consists of assigning each pixel of the image to a class. The set of classes selected to represent the scene of interest constitutes a taxonomy, or nomenclature, which varies according to the needs of the end user. The attribution of these classes is based on the visual analysis specific to the pixel and can be based on its visual description (Khatami, 2016). Furthermore, existing works all use supervised classification methods, based on machine learning algorithms (Khatami, 2016). The supervised term comes from the training step of the algorithm, which consists of modelling the classes in play from a reference data set. After training the data, one can then infer the unlabelled data classes by applying the model. The reference dataset is a set of pixels, described by attributes (observations), whose classes are known in advance. Each class is thus modelled in relation to these attributes, whether they are generative or discriminative methods. According to Shao and Lunetta (2012) supported by Mountrakis et al., (2011), support vector machines (SVM) have also been shown to have the ability to generalize complex classifications, with previous research being able to achieve a 95.2% accuracy. Similar to our study, they also chose to utilize Sentinel-2 data. However, while there have been newer algorithms developed, such as Light Gradient Boosting Machine (LightGBM), not many researchers have utilized them. For a while now, RF and SVM have been adequate to perform the tasks assigned, however, as the study area increases in size, it becomes increasingly costly in time and effort to perform the functions. With LightGBM, there is the ability to have rapid and accurate results with the ability to reach accuracy rates of 92.07% when categorizing crop types, a task that is not easily accomplished by older methods.

By using city of Polokwane as a case study, this study aims to compare the effectiveness of extreme gradient boosting (Xgboost) and Support Vector Machine in the classification of urban land use and land cover, as well as to quantify the changes in land use and land cover from 2008 to 2022 using satellite data from Landsat 5 TM and Landsat 8 OLI using the Google Earth Engine (GEE) cloud platform. Secondly, we aim to examine the changes in landscape patterns due to fragmentation using landscape fragmentation analysis from 2008–2022.

1.2 PROBLEM STATEMENT

Polokwane, the capital city of Limpopo province in South Africa, is currently experiencing rapid urban growth. The expansion of the city's urban areas brings about significant changes in land use and land cover (LULC). The investigation of these LULC changes is crucial to understanding the environmental impact and sustainable management of the city (Ramoelo, Cho, & Naidoo, 2018). To analyse and monitor these changes effectively, the utilization of machine learning algorithms can be beneficial. Machine learning algorithms can process large volumes of remote sensing data and identify patterns, trends, and anomalies in the LULC changes (Contreras-Moreira, & Arias-

González, 2016). According to Schaumburg et al., (2011), employing these algorithms, researchers can gain valuable insights into the dynamic transformation of Polokwane's urban landscape. The incorporation of machine learning algorithms in LULC change detection provides a data-driven and efficient approach to influence appropriate urban planning and decision-making processes.

The number of people living in cities is growing, which has effects on both how cities are built and how they look. Green spaces in cities are becoming more important in urban planning because they can improve people's health and happiness. The city's population and size have both contributed to a greater emphasis on green space (Tan et al., 2013). The loss of parks and other green areas is often a consequence of rapid urbanization.

The challenge of the current techniques used in Polokwane municipality to monitor land use/land cover (LULC) changes and urban growth lies in their reliance on traditional and manual data collection methods. These methods, such as field surveys and manual interpretation of satellite imagery, are time-consuming and labour-intensive. Additionally, they often lack spatial and temporal resolution, hindering the accurate and timely monitoring of changes in LULC and urban growth. To address these challenges and improve monitoring efforts, the proposal of new techniques involves the integration of remote sensing, geographic information systems (GIS), and machine learning algorithms. These modern technologies can provide higher-resolution remote sensing data, automate data processing and analysis, and allow for more efficient monitoring and prediction of LULC changes and urban growth in Polokwane municipality (Amy, 2016; Mahdavi et al., 2019; Zevenbergen et al., 2019).

According to Wiatkowska et al. (2002), for Polokwane's urban growth management, geographic data needs to be collected and analysed so that urban growth can be tracked in a systematic way and growth dynamics and spatial patterns can be understood. Wiatkowska et al. (2002) outline that to get a better solution to this illegal situation, land use and land cover need to be changed in a technical way. That is why we can monitor shifts in land use and cover with clear images. The findings of several change detection methods are compared visually and quantitatively to determine the best solution. The use of geographic information systems (GIS) and remote sensing (RS) is helpful in detecting environmental changes (Safaa, 2023).

1.4 RESEARCH QUESTIONS

- How effective is extreme gradient boosting (Xgboost) and Support Vector Machines in urban land use and land cover classification using the Google Earth Engine (GEE) cloud platform?
- How much landscape fragmentation occurred due to urbanisation from 2008–2022 ?

1.5 THE AIM

The aim of the study is to compare the effectiveness of extreme gradient boosting (Xgboost) and Support Vector Machine in the classification of urban land use and land cover, as well as to quantify the changes in land use and land cover from 2008 to 2022 using satellite data from Landsat 5 TM and Landsat 8 OLI using the Google Earth Engine (GEE) cloud platform. Secondly, we aim to examine the changes in landscape patterns due to fragmentation using landscape fragmentation analysis from 2008–2022.

1.6 MAIN OBJECTIVE

The main objective of this study is to track and record any changes in urban Land use Land Cover change of Polokwane Municipality between 2008 to 2022.

1.6.1 Specific Objectives

- To evaluate the effectiveness of extreme gradient boosting (Xgboost) and Support Vector Machine in the classification of urban land use and land cover, as well as to quantify the changes in land use and land cover from 2008 to 2022 using satellite data of Landsat 5 TM and Landsat 8 OLI together with the use of Google Earth Engine (GEE) cloud platform.
- To examine the changes in landscape patterns of urban green spaces and vegetation fragmentation using landscape fragmentation analysis from 2008–2022 using Landsat 5 TM and Landsat 8 OLI satellite data
- To compare the vegetation, built-up area, impervious surface, and soil fraction of land cover using landscape metrics.

1.7 SIGNIFICANCE OF THE STUDY

Polokwane municipality is experiencing the increase in urban growth caused by the migrants from other SADC countries bordering Limpopo. The municipality serves as a city of trans passing by both the people from Limpopo and the migrants from Zimbabwe and other countries bordering Limpopo. The municipality still maintain the use of policy solution to urban growth challenges and the change in LULC which is time consuming and costly and does not cover large area in terms of monitoring the change. Therefore, monitoring these kinds of changes is important for coordinated action at all levels, including the national, provincial, and district levels, because they reflect changes in economic and social conditions. Scientists and decision-makers in a region would benefit from having access to data on the spatial extents of land use and land cover in the region (IGBP/IHDP, 1999). For example, built-up areas, forests, farmlands, and grasslands are regularly mapped by remote sensing to give the government up-to-date information on how one land cover is moving into another.

As a result, accurate maps at large scales are necessary for effective resource management. Polokwane is extremely abundant in natural resources, and there is a current need for high-quality, high-spatial-resolution digital maps to facilitate the proper management of these resources. Polokwane municipality solely rely on policy solutions and without leveraging geospatial technology this is leading to limited accuracy, time-consuming data collection, lack of spatial analysis, limited monitoring capacity, difficulties in enforcement, insufficient data for planning and decision-making, and challenges in monitoring large areas (Aplin et al., 1999, Dai and Khorram, 1998). Combining policy solutions with geospatial technology will provide more robust and efficient monitoring of LULC in Polokwane. Here, remote sensing data can be utilized to create regularly updated, spatially explicit land cover maps with clearly delineated classification schemes and land cover categories in this approach, an inventory of available resources may be built, updated, and used for both security and efficiency. Therefore, the research is guided by the following research questions.

1.8 RESEARCH STRUCTURE

This research consists of five chapter. The chapters include chapter 1 for introduction/ background, chapter 2 for literature review, chapter 3 for research methodology, chapter 4 for results interpretation and lastly chapter 5 for conclusion and recommendation. This research is organized as follows:

Chapter 1 - Highlights the general introduction of the research. It introduces urban growth, urbanization as a cause of fragmentation, loss of biodiversity because of urbanization, it is followed by urbanization with land use land cover change, the use of SVM and gradient boosting for LULC classification.

Chapter 2- Highlights the conceptual framework of urbanization and urban growth. It explores the loss of green areas and biodiversity caused by the conversion of urban green land to built-up area due to urbanization. The chapter also explores the methods used in LULC classifications in other areas with use of multispectral images. It also reviews similar change detection methods used elsewhere and their performances in LULC classification. Landscape Metrix analysis which determines how LULC are spread out, how they look and how they are put together was also reviewed in this chapter. The chapter also review the comparison of SVM and Gradient boost used in classification of monitoring urban growth in other areas. The chapter review the effectiveness of GEE in LULC classification of other case studies done by other scholars and lastly the chapter also includes satellite remote sensing for urban analysis.

Chapter 3- This chapter analyse the geographical location of the study area which is followed by the methodology flowchart as well as data acquisition. The chapter also includes the field data collection, ground reference and pre-processing, data analysis for SVM and gradient boosting classification. It also highlights the model validity and reliability of the which includes Mcnemar's test statistics, kappa statistics. Assessment of pixel-based accuracy and assessment of procedures as well as users were done and detailed in this chapter.

Chapter 4- There results interpretation and discussion were outlined in this chapter which includes land cover/ use classification. The results for quantifying land use land cover change using landscape metrics and accuracy assessments together with F-scores were outlined in this chapter. There results for landscape matrix analysis/ accuracy assessments and f-scores for XG boosts were detailed in this section. This follows the interpretation of LULC classification and accuracies. In this interpretation of accuracies, McNamar's test statistics of SVM and XG boost was shown and outlined. Lastly there is a section of the discussion of the overall classification, accuracies and landscape matrix analysis.

Chapter 5- The chapter concludes the research by integrating the objectives with the findings as well as giving recommendations which can assist planners and other entities in managing LULC changes.

CHAPTER 2: LITERATURE REVIEW

Historical and current remote sensing and GIS data can be used to identify land use land cover changes. Some studies that are important to this discussion will be discussed below. The research question and conceptual framework are based on participants' perceptions and assumptions. Satellite remote sensing for urban analysis; a review of change detection approaches; global change detection; the characteristics of satellite sensors for urban mapping; and a comparison of machine learning algorithms will be discussed below.

2.1. CONCEPTS OF URBANIZATION

Expanding human populations have led to more space being taken up by cities. It is an idea for building more safe places in more places. Human settlements adapt to their environments, but they also must accommodate the economic and social activities that take place within (Bull, 1976). For a society to develop, it needs to build settlements, which contribute to "people need safe, comfortable, and efficient locations to live and work" (UN-Habitat I 1976: 8).

Housing and settlement issues developed because of rapid urbanization, especially in developing countries. Solutions to these issues must consider economic, social, and environmental factors. Regional growth, urbanization, and national social and economic advancement were all part of the UN's environmental planning project in 1955. Self-sufficiency and community organization were seen as particularly important developing countries. In 1964, the United Nations (UN) Centre for Housing, Building, and Planning was established (Rivkin, 1967).

Population of Polokwane has been steadily increasing over the years, the demand for housing, infrastructure, and services has risen, leading to the expansion of urban areas. This expansion often results in the conversion of agricultural or natural land to residential, commercial, or industrial areas, causing land use and land cover change. Kgaugelo et al. (2019) examined the urbanization process and its impact on land cover change in Polokwane City. The findings of this study revealed a significant increase in urban areas and a corresponding decrease in agricultural and natural areas over the past decade.

In conclusion, the concepts of urbanization and land use and land cover change are closely linked in Polokwane Local Municipality. The growth and development of urban areas in response to the increasing population have led to the conversion of agricultural and natural land for housing, infrastructure, and commercial purposes. Understanding this linkage is crucial for sustainable urban planning and environmental management in the municipality.

2.2. URBANIZATION AND LAND USE LAND COVER CHANGE

The term "urbanization" is used to describe the process through which rural residents leave their homes and relocate to urban centres and the resulting societal changes. Researchers in urban planning, geography, sociology, architecture, economics, education, statistics, and public health all study urbanization, which is defined as the process by which towns and cities grow as more people live and work in core areas.

The term "urbanization" can be taken in two senses: either as a static state (such as the percentage of the population or land area that is in urban areas) or as a dynamic process (Griess &Grundmann, 2017). Indicators of urbanization include the percentage of the population living in urban areas and the rate of urbanization. Huge social, economic, and environmental shifts brought on by urbanization present an opportunity for sustainability through "the use of resources more efficiently, the creation of more sustainable land use, and the protection of natural ecosystem biodiversity" (Day & Ellis, 2014: 65). Target 11 of the Sustainable Development Agenda is dedicated to making cities more resilient and sustainable in the face of rapid urbanization. 2016 UNFPA.

The spread of cities has led to the loss of natural habitats and species diversity. Many authors, including Cheng et al. (2009), Liang and Wang (2002), and Weng et al. (2009), have documented the loss of many species of plants in the rapidly urbanizing regions of China (2008). Urban sprawl, tall buildings, and the overexploitation of groundwater all contribute to subsidence. principal

2.3. LAND USE LAND COVER CLASSIFICATIONS

The term "Land Use/Land Cover" (LULC) is used to describe the practice of categorizing the human and non-human components of a landscape according to recognized scientific and statistical methods and reliable data sources. It lends itself to a wide range of classifications. LULC characteristics include urban/built-up areas, agriculture, forests, bare soil, waterbodies, etc (Twiss & Buchroithner, 2019).

According to Cihlar et al., (1998:5), “Classification algorithms should be accurate, reproducible by others using the same input data, robust (not sensitive to small changes in the input data), able to fully exploit the information content of the data, applicable uniformly across the entire domain of interest, and objective (not dependent on the analyst's decision)”. Many existing approaches to classifying digital images fail to do so (Cihlar et al., 1998:5).

As part of the FAO AFRI project, AFRI Cover was initiated between 1997 and 1999 before its official launch (Di Gregorio and Jansen, 2000). Its authors, Di Gregorio A., and L.J.M. Jansen, received feedback from all over the world and released a revised edition (LCCS 2) in 2005. (2005). Standardizing the attribute nomenclature, according to LCCS, is more important than standardizing the final categories. The diagnostic features used by LCCS as classifiers to place patients into distinct LC subtypes are numerous and complex. The classifiers can be used by an application ontology to give each LC class in that ontology a more nuanced set of semantics (classification system), Ahlquist (2008)

Anderson (1971) made a list of ways to classify how land is used and how it is covered. Data at LULC levels I and II may be useful if you are interested in the whole country, the states, or even just a single county. Level III and IV land use and land cover data are usually used by organizations that need and make local information at the intrastate, regional, or municipal level. The original plan was for these categories to be made by the user groups themselves, so that the framework could be made to fit the needs of each user group.

Land Use and Land Cover (LULC) change analysis in Polokwane are significant and provide valuable insights into the changing dynamics of the city's landscape. McGarigal et al., (2012) highlighted that evaluating the effectiveness of extreme gradient boosting (Xgboost) and Support Vector Machine (SVM) in the classification of urban land use and land cover is important as it informs us about the accuracy and reliability of these machine learning algorithms in detecting and classifying different types of land cover. This knowledge enables researchers and urban planners to better understand the composition and distribution of the land cover, facilitating effective land use planning and management decisions.

According to Sudhira et al., (2009), quantifying the changes in land use and land cover from 2008 to 2022 using satellite data of Landsat 5 TM and Landsat 8 OLI, along with the use of Google Earth Engine (GEE) cloud platform, allows us to track and monitor the transformations caused by urban growth over time. By identifying the specific changes and trends in land use and land cover, decision-makers can assess the impact of urbanization on the environment, identify areas of concern, and develop appropriate strategies for sustainable urban development.

Additionally, Sudhira et al., (2009) argues that examining the changes in landscape patterns of urban green spaces and vegetation fragmentation provides valuable information regarding the quality and health of urban ecosystems. Landscape fragmentation analysis utilizing Landsat 5 TM and Landsat 8 OLI satellite data allows us to understand how urbanization affects the spatial distribution and connectivity of green spaces. This knowledge is essential for advocating for the preservation and enhancement of green areas, promoting biodiversity, and improving the overall quality of life in the urban environment.

Lastly, comparing various landscape metrics such as vegetation, built-up area, impervious surface, and soil fraction of land cover helps us to better understand the composition and configuration of the urban landscape. This analysis provides insights into the extent of vegetation cover, the degree of imperviousness of surfaces, and the quality of soil, which in turn contributes to assessing the ecological health and resilience of the urban ecosystem. By understanding and monitoring these landscape metrics, urban planners and policymakers can make informed decisions regarding land management and environmental sustainability.

2.4. THE USE OF SVM AND GRADIENT BOOSTING FOR LULC CLASSIFICATION

Several research have shown that high-resolution images and machine learning techniques improve urban LULC classifications (Klein et al., 2009). Klein et al. (2009) used TerraSAR-X data to semi-automatedly classify Cape Town's metropolitan districts using SVMs. They were 82.3% accurate. In 2020, Ha et al. investigated the classification of green infrastructure in two Croatian cities: Varaždin and Osijek. SVM was compared to random forest, artificial neural network, and Naïve classifier using Sentinel-2 satellite imagery. The SVM approach produced more accurate classification results than other methods, with Varaždin's Kappa value of 0.87 and Osijek's at 0.89.

LULC analysis of remotely sensed recordings has been extensively studied. From 1986 through 2001, Hua et al. (2017) used DTs, SVMs, and MLCs to map Pallisa District, Uganda, land cover. Their studies assessed local land cover. They used knowledge mining to determine decision-making classification bands and thresholds. The study tested categorization models and found that land cover elements occurred at an unpredictable speed. Several image classification models can segment and label a multi-dimensional component space into homogenous portions, depending on the classes needed. Parametric model classifiers need a normal dataset and precise statistical parameter collection from training data. The most common parametric classifier is the maximum-likelihood classifier (MLC), which generates decision surfaces based on class mean and covariance.

MLC was first used on IRS LISS-III images between 2001 and 2011, dividing the data into eight groups. The study also used a novel methodological framework for post-classification modifications (Shaharudin, Ahmad, and Nor, 2020). Overall categorization precision increased from 67.84% to 82.75% in 2001 and from 71.93% to 87.43% in 2011. Ohmaid et al. (2020) examined land use changes in Chunati Wildlife Sanctuary (CWS) between 2005 and 2015 using Landsat TM and Landsat 8 OLI/TIRS imagery. The land use change evaluation using ArcGIS with ERDAS envision (Jamali, 2019). The greatest likelihood classification approach was used to classify land use supervised. Over 10 years (2005–2015), the degraded forest increased by 256 ha at 25.56% per year. This was discovered. Non-parametric classifiers cannot isolate classes from multi-dimensional feature spaces because they do not appropriate information. Decision trees, SVMs, and expert systems are the most common non-parametric classifiers (Otukei and Blaschke, 2010).

Pixel classifier-based ML systems have been used to analyse remote sensing photos (Swetanisha, Panda and Behera, 2022). Otukei and Blaschke (2010), Islam et al. (2018), and Aliyu, Mokji, and Sheikh (2020) use very high-resolution satellite images and derived land cover maps to map urban land use at the street block level. The method emphasises home use. A random forest (RF) classifier classifies street blocks. This classifier has 84% accuracy for five land-use classes and 79% for six. The RF classifier used in Dakar and Ouagadougou spanned approximately 1,000 kilometres with a spatial resolution of 0.5 metres. In 2019, Jamali compared eight machine learning methods for image categorization in northern Iran. These approaches were created using R and the Waikato knowledge analysis environment. Machine learning models like RF, SVM, decision tree, K-nearest-neighbors (KNN), and PCA have been successfully deployed in several application fields (Islam et al., 2018; Aliyu, Mokji, and Sheikh, 2020; Swetanisha, Panda, and Behera, 2022). The ensemble model we constructed using SVM and XGBoost is more accurate than other machine learning models (Kulkarni, 2020).

2.5.CLASSIFICATION ON GOOGLE EARTH ENGINE

Further work is being done all the time to improve the accuracy of LULC maps. Due to the availability of free, high-resolution remote sensing images, researchers now have more options than ever before for improving upon existing map products. Several methods exist for achieving this goal, such as careful sample selection during training, the incorporation of a larger number of input features, the use of multi-temporal images with higher resolution, the application of sophisticated classification algorithms, etc. This complex set of interrelated issues is the "Big Data" challenge, which in turn mandates massive computing resources and more data storage for image categorizations (Azzari & Lobell, 2017). According to Giri, Pengra, Long, and Loveland, NASA Earth Exchange (NEX) and GEE provide a forum for addressing such issues (2013). In addition, Giri, Pengra, Long, and Loveland (2013) note that GEE has grown as a new method for investigating spatial data.

The mapping of land cover on a worldwide scale is required for remote sensing. Using the cloud computing resources offered by Google Earth, Gong et al. (2013) produced a worldwide land cover map with a resolution of 30 meters. GEE is capable of carrying out such analysis on a single platform. Midekisa et al. (2017) utilized GEE to generate annual maps of Africa over the course of 15 years. Hansen et al. (2013) obtained multi-temporal satellite image data spanning 12 years using GEE. This allowed them to map worldwide forest loss and growth at a resolution of 30 meters. Another global problem that calls for an investigation on a more expansive scale is urbanization. To classify and analyze huge amounts of LULC data, an environment with a high computational capacity is required. GEE was utilized in the research conducted by Goldblatt et al. (2016), Trianni, Angiuli, Lisini, and Gamba (2014), and Patel et al. (2015). These methods are also capable of producing cost-effective datasets on a national and global scale. In the field of agriculture, GEE has been utilized for crop mapping, smallholder farming, and comparative analysis using various in-built machine learning classifiers on larger regions with multi-temporal datasets (Aguilar, Zurita-Milla, Izquierdo-Verdiguier, & de by, 2017; Dong et al., 2016; Shelestov, Lavreniuk, Kussul, Novikov, & Skakun, 2017). According to Padarian, Minsny, and McBratney (2015), the performance of GEE power in Digital Soil Mapping (DSM) was anywhere from 40 to 100 times faster than the desktop workstation.

2.6. SATELITE REMOTE SENSING FOR URBAN ANALYSIS

The use of remote sensing is becoming an increasingly important part of managing the environment and planning how to use urban land. It accomplishes this by collecting data that is multispectral, multiresolution, and multitemporal, and then transforming those data into information that is useful for understanding and monitoring the processes that occur on urban land, as well as developing databases of urban land cover. This allows it to better understand and manage the processes that take place on urban land (Hu et al., 2016). In recent years, the use of digital multi-spectral satellite images, rather than the more conventional method of aerial photography, has emerged as the principal focus of urban remote sensing. Since the launch of the first generation of Landsat MSS satellites in 1972, technological advancements have led to the creation of the second generation of earth observation satellites, such as Landsat TM, Landsat ETM+, SPOT, and Indian Remote Sensing Linear Imaging Self-Scanner (LISS) sensors, as well as the third generation of satellites with very high geometric resolution. Satellite remotely sensed technology has made possible to monitor land surface changes in a large-scale, high-precision, and long-term sequence.

Remotely sensed images can effectively provide a good data source for effectively extracting, analysing, and simulating updated LULC information and changes (Pradhan et al., 2008; Singh et al., 2018). Many researchers in China and abroad use remotely sensed (RS) technology and geographic information systems (GIS), with the support of land-use models and statistical methods, to carry out land-use landscape pattern research (Bürgi et al., 2005, Liu and Yang, 2015; Feng et al., 2018). Akinyemi et al. used Landsat and SPOT satellite data to study changes of LULC in northern Rwanda, Dakaria Province in Egypt, Hue City in Vietnam, Nagardai City in Iran, and the Atlanta Metropolitan Area in the United States (Deng et al., 2009; Gong et al., 2013; Hegazy and Kaloop, 2015; Parsa and Salehi, 2016; Akinyemi, 2017). Researchers such as Ma have used multi-temporal Landsat satellite imagery to analyze the changes of LULC in Syria from 2010 to 2018 (Disperati and Virdis, 2015; Dadashpoor and Nateghi, 2017).

By analysing the changes of LULC, exploring the evolution of landscape pattern becomes more feasible (da Silva et al., 2015; Dadashpoor et al., 2019; Mohamed et al., 2020), Alam, Deng, Hasan, Salvati and et al. used Landsat satellite imagery to analyse changes of LULC in Kashgar

Valley, China, in highly urbanized areas, and in southern China, which is conducive to understanding the changes in the land-use landscape pattern (Deng et al., 2019; Salvati et al., 2018; Hasan et al., 2019; Alam et al., 2020). With the development of remotely sensed technology, Gong used Landsat satellite data to establish a global land-use database with a spatial resolution of 30 m for the first time, which has enhanced the application value of Landsat data in global land-use change monitoring (Gong et al., 2013). Liu's team released a fine classification product of global 30-m resolution land cover in 2015 and 2020, which reflected the distribution of land cover in global land regions (except Antarctica) in a timely manner (Liu et al., 2021; Zhang et al., 2020). It provides the latest data support for surface-related applications (Zhang et al., 2020). The change in land-use type and the expansion of urbanization have direct impacts on the landscape pattern and the social and economic functions of cities.

2.7.CHANGES IN LANDSCAPE PATTERNS USING LANDSCAPE FRAGMENTATION ANALYSIS

Some scholars have also analysed and optimized the ecological networks of urban planning to satisfy the requirements of good urban ecosystem. De Montis et al. (2016) studied the urban and rural landscape planning in the Nuoro region of Italy. They noted the important role of the network model in protecting and improving the species diversity to prevent landscape fragmentation and establish the connection between the isolated ecological patches. According to Mas, Gao & Pacheco, (2010), various quantitative landscape pattern indices have been used to reflect the spatiotemporal transformation of landscape patterns and ecological processes on different scales, which laid a foundation for the study of the changes process and characteristics of landscape pattern (Mas, Gao & Pacheco, 2010). FRAGSTATS was a commonly used landscape pattern analysis software with powerful calculation and analysis function. It can calculate three levels of landscape pattern indices, namely patch level, patch type level, and land type level (Mcgarigal & Marks, 1995)

Cabral & Costa, (2017) stated that some scholars used Landscape fragmentation indices, such as Patch Density Index (PD) and Mean Patch Size Index (MPS) to describe the relationship between landscape size and landscape fragmentation. Landscape shape indices, such as the area perimeter fractal dimension (PAFRAC) and the Landscape Shape Index (LSI) also was used to reflect the complexity of the patch shape. Landscape dominance indices, such as the Largest Patch

Index (LPI) and Percent of Landscape (PLAND), can determine the dominance type of the landscape. Landscape aggregation indices, such as the Aggregation Index (AI) and Contagion Index (CONTAG) has been used to indicate the degree of aggregation between different patch types. Landscape diversity indices, such as the Shannon Diversity Index (SHDI) and Shannon Evenness Index (SHEI), indicate the richness and evenness of the landscape, respectively.

McGarigal & Marks, (1995) utilized four groups of landscape indices including landscape fragmentation, shape, aggregation and diversity index, to analyse and describe the landscape structure and spatial pattern characteristics. Landscape fragmentation indices, such as largest patch index (LPI) and landscape division index (DIVISION),were also used to describe human activities affecting landscape patches. McGarigal & Marks, (1995) further employed Landscape shape indices, such as landscape shape index (LSI) and patch cohesion index (COHESION), measure patch shape and connectivity to detect the changes in landscape patterns of LULC changes in different years. Landscape aggregation indices, such as contagion (CONTAG) and the interspersion juxtaposition index (IJI) were used to demonstrate mixed dispersion among different patch types. Landscape diversity indices, such as Shannon's diversity index (SHDI) and Simpson's evenness index (SHEI), illustrate landscape richness and evenness, respectively. All indices were calculated using the software FRAGSTATS 4.2

There have been a few studies conducted throughout the years that have utilized landscape fragmentation metrics to determine what LULC change detection entails. The model was used to understand the changes that occurred in 2008, 2013, 2017, and 2022. As a result, it became much simpler to determine the patch density of each class and the number of patches, as well as the interpretation of great patch metrics.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 GEOGRAPHIC LOCATION/ STUDY AREA

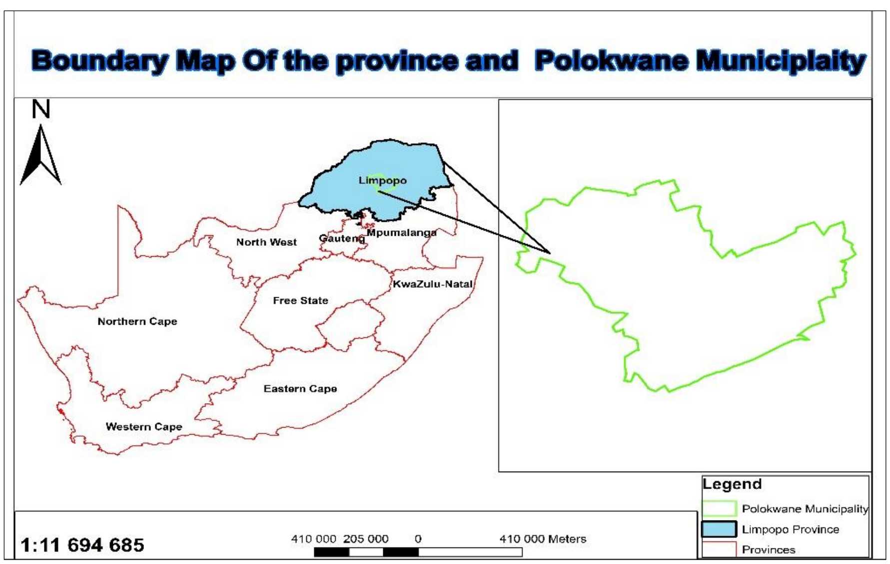
City of Polokwane is in Limpopo Province, and “the municipalities of Molemole, Greater Tzaneen, Lepelle-Nkumpi, Mogalakwena, and Aganang are found in the immediate vicinity (in Capricorn

District Municipality). It is located at 23.8983S and 29.4490E (Polokwane Municipal Spatial Development Framework 2010)".

Polokwane, the capital of Limpopo and the economic centre of the province, is on the Great North Road, which is the main route to Zimbabwe and the urban centre with the greatest population that lives in the surrounding region. According to Polokwane Municipal Spatial Development Framework (2010), Polokwane is a popular tourist destination and port of entry to Limpopo and Zimbabwe due to its proximity to Kruger National Park, Magoebaskloof, Botswana, Zimbabwe, Mozambique, and Swaziland.

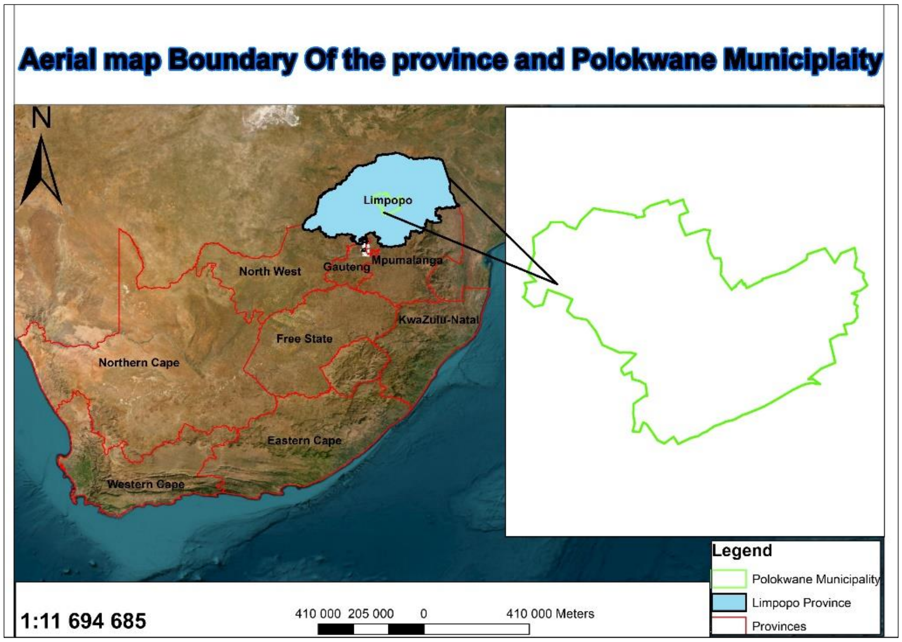
According to Polokwane Municipal Spatial Development Framework 2010, Polokwane's population grew from 458,120 in 1996 to 557,404 in 2008. Polokwane's population grew by 99 284 people between 1996 and 2008, an average annual growth rate of 1.65%, lower than South Africa's 2% growth rate during the same period. The population of Moletji has continuously decreased over the years, reaching 98,454 residents in 2002. Similarly, Pietersburg Part 1 had a population of 86,412, and Seshego Part 1 had 78,001 residents in 2008. However, there has been a notable difference in population growth rates between the two areas. Pietersburg Part 1 has experienced an annual increase of 5.03 percent since 1996, whereas Seshego Part 1 has only grown by 1.81 percent annually.

"Map 1: Boundary Map of the Study area



Source; ArcGIS, September 2023

Map 2: Aerial Map of Polokwane Municipality



Source: ArcGIS 10.4, September 2023

3.2 TOPOGRAPHY OF THE STUDY AREA

According to the Limpopo provincial Environmental Management Framework of 2003, the study area is split into two rough topographical units: "moderately undulating plains" in the east and "strongly undulating plains" in the west. Most of the eastern half of the municipal area is made up of plains with slight hills. The Strydpoort Mountains border the Polokwane Municipal Area to the south, the Waterberg Mountains to the west and north, and the Great Escarpment to the east. This area is known as the "Petersburg Plateau." The highest point of the plateau can be found in the south, close to the Strydpoort Mountains. (Polokwane Municipal Spatial Development Framework, 2010) outlined that these mountains act as a watershed between the Olifants River system and the Sand River system.

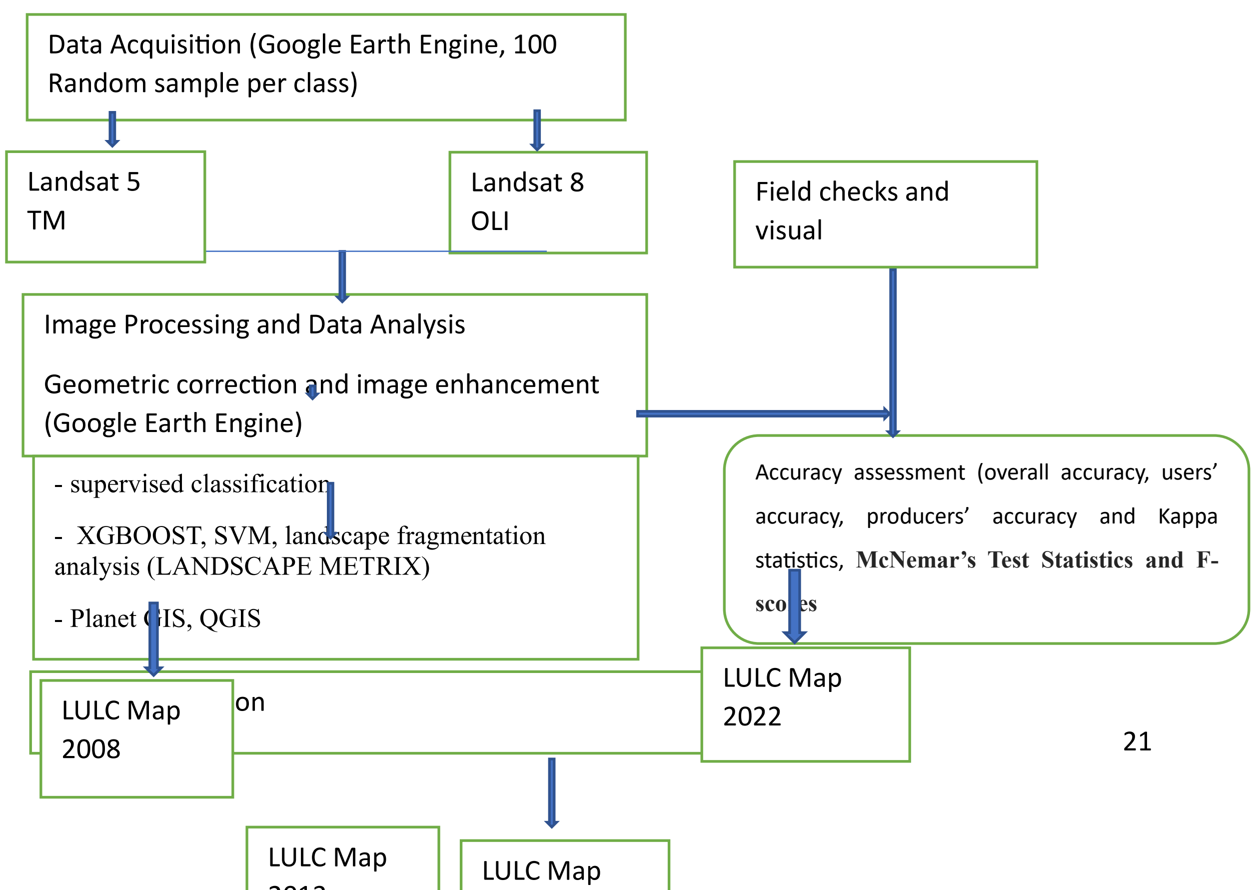
3.3 SOILS AND GEOLOGY OF THE STUDY AREA

Depending on factors including slope, soil type, depth, and underlying geology, the Polokwane Municipal Area can be broken up into several distinct land-type units (Dept. of Agriculture and

Water Supply, 1988). “This data can be used to infer a variety of features, including erosion risk, agricultural viability, and development potential”.

3.4 METHODOLOGY FLOW CHART

The figure below represents the methods and techniques used in this study:



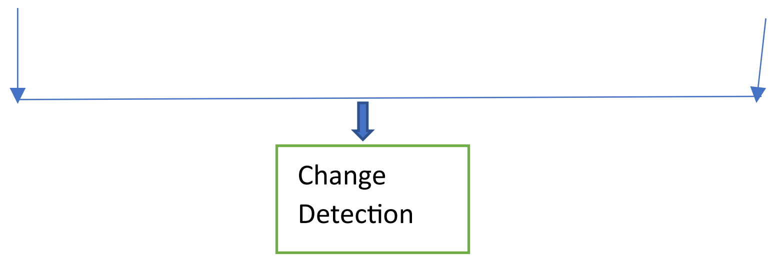


Figure 1: Methodology Flow chart

3.4.1 Data Acquisition

The study used high-resolution satellite sensors such as Landsat 5 TM (seven bands) and Landsat 8 OLI (11 bands) to detect changes in LULC. Non-Parametric Machine Learning classifiers such as Support Vector Machine (SVM), AND gradient boosting have been reported to provide highly accurate LULC classification results using remotely sensed images (Foody & Mathur, 2004; Nery et al., 2016). Images were uploaded through GEE coding for each year. The images were already atmospherically corrected and ready for classification using Boost (extreme gradient boosting) and Support Vector Machine (SVM) algorithms. Multiple samples were created according to the categories of class on GEE. Classes included bare soil, agriculture, business, waterbodies, and built-up areas. The study area was obtained using the Limpopo cadastral data obtained from the Planet GIS software.

3.5 FIELD DATA COLLECTION, GROUND REFERENCE AND PRE-PROCESSING

Primary and secondary data sources were used. A short field survey was conducted to verify land use changes and new developments. The most important information was gathered through direct field observation, taking direct geographical coordinate systems of some of the land uses and land cover, creating spectral signatures on GEE, meetings with key witnesses, field perception, direct dialogue with the public, and organized and unstructured witness meetings with some of the municipal town planners to verify land uses on their GIS data. This study used satellite images from GEE that were already atmospherically corrected and ready for use.

The position data for validation through ground-based field work was conducted and 100 sample size per class was selected. On this data collection exercise, data points collected were for bare soil, built-up areas, waterbodies, agriculture, and business. Systematic random sampling was applied for each class and the data was splitted to 70% training samples and 30% validation testing. Results will be obtained and interpreted on chapter 4 and chapter 5.

The ground truth data collection for land use and land cover (LULC) change in Polokwane City involved systematic random sampling, stratified sampling, field surveys by trained personnel, and thorough validation procedures. These measures ensured that the sample accurately reflects the variability in land use and cover patterns across the city. The data was dispersed evenly across the city, capturing the diverse range of land use and cover patterns. The collected information was cross-checked and verified through multiple visits, revisits, and comparisons with existing data sources, ensuring the reliability and representativeness of the final sample.

3.6 DATA SETS

3.6.1 Landsat 8 OLI (11 bands)

The data collected by the Landsat satellite is repeated every 16 days. Given the sensor's feature, the flexibility with which data can be selected has grown, which is especially helpful in situations when cloud cover is a significant barrier in satellite data selection (Jamal & Ahmad, 2020). The filtered collection by date feature of the GEE archive was utilized to choose data for analysis. The image was placed together after being filtered by year. For each year, the median value of each pixel was used in the final product. The resultant image object is the median of all the images in the filtered set for each filtered collection. Carleer et al, (2004) persists that the use of median filter on Google Earth Engine corresponds to the homogeneous regions of the image. Before transforming the original data into a gradient image, a median filter can be applied on the image to reduce the noise. The presence of noise in an image causes an over detection of edges by the morphological gradient. The median filter locally homogenizes the image and avoids extreme gradients, and thus disturbing contours (Carleer et al,. 2004). Collections are an effective method of removing many of the cloud-contaminated pixels since clouds appear in a variety of places in the images (Tassi A., 2020; Patel et al., 2015). Landsat 8 images with a cloud cover of less than 10% were used for this study. In this study, the data collection from Landsat 8 was utilized to measure various frequency ranges over the electromagnetic spectrum. These frequency ranges are not visible to the human eye. Landsat 8 includes 11 different bands, and each range is referred to as a band. Landsat numbers its red, green, and blue sensors as 4, 3, and 2 (Kern et al., 2022).

Table 1: Landsat 8 Spectral properties with important bands for this study

Band Number	Band Name	Spectrum range **(µm)**	Resolution
1	Coastal	0.433–0.453	30m
2	Blue	0.450-0.515	30m
3	Green	0.525–0.600	30m
4	Red	0.630–0.680	30m
5	NIR	0.845-0.885	30m
6	SWIR 1	1.560-1.660	30m
7	SWIR 2	2.100-2.300	30m
8	PAN	0.500-0.680	15m
9	Cirrus	1.360-1.390	30m
10	TIRS 1	10.6-11.2	100m
11	TIRS 2	11.5-12.5	100m

Source: (Wiweka, Suwarsono, Jalu T Nugroho, & Samsul, 2014, p. 5)

Only the bands that are sensitive to light with extremely short wavelengths (bands 1–4 and 8) can detect visible light; all the remaining bands are in regions of the spectrum that are invisible to the human eye. So, bands 1-4, 5 and 8 were used in the research to get high-resolution images in terms of spectral resolution.

Band 5 was used most of the time, because it can measure NIR which stands for 'near infrared'. 'This part of the electromagnetic spectrum is important to the study of ecology because healthy plants can reflect it'. The water in their leaves is what scatters the wavelengths back into the air. If we merely looked at the amount of visible greenery on the plant, we wouldn't obtain as accurate a reading as we would using an index such as the normalized difference vegetation index (NDVI) (Wiweka, Suwarsono, Jalu T Nugroho, & Samsul, 2014, p. 5).

The changes in land use that occurred over time are reflected in bands 2, 3, and 4, which may be seen in the colours blue, green, and red, respectively.

According to Wiweka et al,. (2014) **Bands 6 and 7** cover different parts of the shortwave infrared, or SWIR. Because of this, the two bands were put together to make a reflection of the bare soil. In geology, they are especially useful for telling the difference between wet and dry ground and for telling the difference between rocks and soils that look the same in other bands.

Suwarsono et al,. (2014) outlines that band 8 was more important than the other bands since it contains the panchromatic, or pan, information. It operates in the same manner as traditional black-and-white film in that, rather than gathering visible colours in their individual channels, it combines them into a single channel. As a result of this sensor's ability to take in more light all at once, its resolution is the highest of any of the bands, coming in at 15meters (50ft).

3.6.2 Landsat 5 TM (seven bands)

This research made use of both Landsat 5TM and Landsat 8 to obtain results that were complementary to one another. Landsat 4 and Landsat 5 were both equipped with the Landsat Thematic Mapper TM sensor, which can produce images with a resolution of 30m and consists of six different spectral bands. The length of the scene is approximately 170 km from north to south and 183 kilometres from east to west (106mi by 114mi). TM was unable to distinguish individual houses or trees, but it was able to record areas in which dwellings had been built or forests had been removed.

Table 2:Landset5 TM Spectral properties with important bands for this study"

Band Number	Band Name	Spectrum range (**µm**)	Resolution
1	visible blue	0.45-0.52	30m
2	visible green	0.52-0.60	30m
3	visible red	0.63-0.69	30m
4	NIR	0.76-0.90	30m
5	SWIR 1	1.55-1.75	30m
6	Thermal	10.40-12.50	120m
7	SWIR 2	2.08-2.35	30m

Source: (*Landsat 5 (TM) Characteristic*, 2022)

3.7 DATA ANALYSIS

Machine learning algorithms such as XGBOOST and Support Vector Machines (SVM) have shown great potential in remote sensing and GIS analysis. The use of Geographic Information Systems (GIS) for multi-source data analysis offers several advantages over more conventional approaches to change detection. GIS allows simultaneous integration of various data sources like satellite imagery, ground-based sensors, and social media data, enhancing comprehensive understanding of change processes and enabling more accurate change detection results. One key advantage is the ability to integrate and analyze disparate data sources simultaneously. GIS allows the integration of various data types, including satellite imagery, ground-based sensors, and social media data, among others, which enhances the comprehensive understanding of the change process. By considering multiple data sources, GIS enables more accurate and reliable change detection results. They are used for land cover classification and prediction, enabling accurate and efficient assessment of producers' and users' accuracy. These algorithms can handle large datasets and extract complex patterns and relationships in the data, leading to improved accuracy assessments. The two algorithms were used in the study area of this research.

3.7.1 Support Vector Machine

Support vector machines (SVMs) are one of the high-performing classifier that seeks to build an ideal hyperplane separating different classes with the minimum number of misclassified pixels during the training phase. SVM uses the kernel trick to transform non-linear datasets into linear ones by projecting them onto a higher dimensional feature space. One of the key limitations of SVM is that accuracy is very sensitive to the selection of kernels and other input parameters (Mountrakis, Im, & Ogole, 2011). Despite this, SVM is a common classifier used in remote sensing because it produces reliable results by using a limited number of support vectors derived from the data used for training (Giles M. Foody, Mathur, Sanchez-Hernandez, & Boyd, 2006). In the following research, XGBoost was compared against a method using SVM as a classifier.

3.7.2 Extreme gradient boosting classification

The ML library XGBoost is a gradient-boosting tool. An algorithm that is both very efficient and highly effective (Li et al., 2023). For this research, tree regression was utilized. Due to its exceptional computational power and model performance, XGBoost can overcome a wide range of constraints. Optimal for experimentation with user-defined prediction, regression, classification, ranking, and related tasks (Kavzoglu & Teke, 2022). To improve prediction accuracy (GBMs) use an addictive model of shallow decision trees that are learners (Friedman, 2001). XGboost is a newer version of the GBM that optimizes the loss function while building the additive model (Chen & Guestrin, 2016). One XGboost innovative features is an objective function that incorporates both the loss function and a regularization term to regulate the complexity of the model. This allows for parallel processing and the preservation of peak efficiency during computation.

3.7.3 Landscape Fragmentation Analysis

The term "landscape fragmentation" refers to the process through which large areas of naturally occurring land cover are broken up into smaller, more isolated regions. This method was utilized for land use land cover fragmentation. Landscape pattern is used to reflect landscape variability and human and environmental actions (Kamusoko & Aniya, 2007). In this study, the method examined the trend and changes in land use and land cover well the density changes. QGIS software was utilized with the plug in for landscape fragmentation analysis and results for patches were obtained.

Landscape fragmentation analysis, using tools like LANDSCAPE METRIX, is another important aspect of accuracy assessment. It quantifies the degree of fragmentation and connectivity in a landscape, providing insights into ecological processes and patterns. By assessing landscape fragmentation, one can evaluate the quality and suitability of habitats, identify areas at risk of habitat loss or fragmentation, and plan effective conservation strategies. Regardless of the sample size, Landscape Fragmentation Analysis made it simple to locate unique cell values in the error matrix.

3.7.4. Google Earth Engine

According to Andrea and Marco, (2020), Google Earth Engine is a powerful tool that can be utilized for Land Use and Land Cover (LULC) image classification. Andrea and Marco (2020) alluded that the tool allows for the analysis of satellite images and geospatial data, providing an efficient and accurate way to classify and understand the different land uses and covers in the municipality. With its vast collection of satellite imagery, Google Earth Engine enables researchers and urban planners to easily access and process large datasets, enabling them to classify and monitor changes in LULC over time. This can be particularly beneficial in the case of Polokwane Municipality, where urban expansion and land use changes occur rapidly. According to Mahdianpari et al., (2020) Google Earth Engine, gives valuable insights into the LULC patterns, such as the expansion of residential areas, agricultural land use, and changes in vegetation cover. This information can aid in various decision-making processes, such as urban planning, conservation efforts, and natural resource management, contributing to the sustainable development and management of the municipality.

Google Earth Engine is a powerful platform that allows for the visualization and analysis of geospatial data on a global scale. It provides access to a vast collection of satellite imagery, including Landsat 5 TM and Landsat 8 OLI. These datasets are widely used for land cover and land use change analysis, as well as monitoring urban growth. Their high spatial resolution and temporal coverage make them valuable tools for accuracy assessment.

3.8 MODEL VALIDITY AND RELIABILITY (ACCURACY ASSESSMENT)

The classification of land use and land cover was performed through the utilization of satellite data sets and then validated through the collection of ground truth data from a range of different sources. After that, the densification of each land use was used to analyse each pixel, and the landscape fragmentation analysis method was utilized to produce the land use and cover maps that span the years 2008 through 2022. This was done so that the maps would be more accurate. When calculating the level of accuracy, the Kappa statistics were taken into consideration. With the assistance of GEE coding, it was possible to evaluate the precision of both the users and the producers. A comparison was made between the land use/cover maps from 2008, 2013, 2017, and 2022 using reference data that was collected in the field.

Following the classification of the images, the total geographical area of each land use and land cover class was calculated in terms of hectares for each time. In addition, the degree to which the kind of land use shifted both over the course of a single time and between other time periods was analysed. To complete the calculations, we made use of the Landscape fragmentation analysis plugin in QGIS by inserting the classification map from GEE to the landscape Metrix plugin. This allowed us to determine the rate of change in hectares per year as well as the % share of each class across the time periods that were investigated.

The accuracy of the classification results obtained from each of the classifiers was evaluated making use of the confusion matrix method. To determine the accuracy rating, the code on GEE was utilized to first calculate the kappa coefficient (Kc), which was then combined with the aforementioned information (Congalton and Green 1999). The kappa coefficient was able to provide a measurement of the degree of variance that was there between the definite layouts that were being compared. The grouping was carried out by the location data and the classifier that is used to dealing with it. This was done so that it could be compared to the possibility of a covenant existing between the unsystematic classifier and the reference data (Song & Park, 2020)

3.8.1 McNemar's Test Statistics

McNemar's test is a paired nonparametric or distribution-free statistical hypothesis test (McNemar, 1947). It is less intuitive than some other statistical hypothesis tests. The test is used to compare the performance of the two classifiers using the error matrix of the two classifiers. It is simple yet powerful method to compare class-wise predictions between algorithms (Dietterich 1988; de Leeuw et al. 2006).

McNemar's test is calculated using the following formula:

$$Z= \frac{b-c}{b+c}$$

In the articles of Leeuw et al. (2006) and Foody (2004), indicated that symbols represented the numerical values where b represents the number of samples that were incorrectly classified by the first classification algorithm but correctly classified by the second classifier, and c represents the number of samples that were correctly classified by the first classification algorithm but incorrectly classified by the second classification algorithm. Therefore, if the cells have counts that are similar, it shows that both models make errors in much the same proportion, just on different instances of the test, and if the cells have counts that are not similar, it shows that both models not only make different errors, but in fact have a different relative proportion of errors on the test set (Food, 2004:

Leeuw et al., 2006). In addition, if the Z value is more than 1.96, there is a statistically significant difference in the classifiers' accuracy (Food, 2004: Leeuw et al., 2006)

3.8.2 Kappa Statistics

In this study, kappa statistics were used to measure how well two different classifiers worked. Adam et al. (2014) says that the error rates of the two classifiers are used in non-parametric tests like the kappa statistics. The data from spatial remote sensing were thoroughly studied to categorize different types of LULC (Stehman, 1996). We selected a set of places (pixels) to test for accuracy using a sample design that was numerically rigorous.

The value of each background cell was adjusted using the columns and rows that were adjacent to it. Both the producer and the user acknowledged its accuracy. To obtain a percentage, take the value of each row and column cell and multiply it by 100.

3.8.3 Assessment of Pixel-Based Accuracy

The overall accuracies were computed using GEE as well as producers and users' accuracies (Matarira et al., 2022). In this pixel-based accuracy, F-scores were also calculated to measure the rate of how successful classifier is. The F-score calculations was calculated using the producer's accuracy and the user's accuracy (Matarira et al., 2022). Zurgani et al (2010) indicated that by weighting the average of producer and user accuracies, the F1 score illustrates the classifiers performance in terms of both accuracies. The formula for F-scores is as follows:

$$F\ score = 2 * \frac{(UA)(PA)}{UA + PA}$$

The formula was used to evaluate the model in relation to its accuracy were the percentages of the F-score were determined.

3.8.4. Assessment of Producers and users Accuracy

One basic accuracy measure is the overall accuracy, which is calculated by dividing the correctly classified pixels (the sum of the values in the main diagonal) by the total number of pixels checked.

Besides the overall accuracy, the classification accuracy of individual classes can be calculated in a similar manner. Two approaches are possible:

• user’s accuracy, and
• producer’s accuracy

According to Congalton and Green (2008), the accuracy of the producer can be calculated by dividing the number of correct pixels in a single class by the total number of pixels that can be calculated using reference data (column total in the tables below). The accuracy of the producer is determined by how accurately a particular area has been categorized. It shows the error of omission, which is the number of things that can be seen on the ground but are not shown on the map. This error refers to the error of omission. The accuracy of the producer will suffer in direct proportion to the number of errors of omission that are present.

This statistic is known as the user's accuracy, and it is calculated by dividing the number of pixels in a class that were correctly classified by the total number of pixels that were classified in that class. As a result, the dependability of the map is directly proportional to the precision of the user. It gives the user an idea of how well the map portrays the situation that is occurring on the ground. It's possible for a single class to represent two different types of players on the ground (Congalton, & Green, 2008).. The "right" class indicates that the land-cover class on the map and on the ground correspond to one another, whereas the "wrong" class indicates that the land-cover class on the ground differs from what was anticipated by the map. Errors of commission are the term used to describe the later categories of mistakes. The accuracy of the user will suffer in direct proportion to the number of mistakes of commission that are present. Therefore, accuracy assessment is a crucial step in remote sensing and GIS analysis that allows the evaluation of the quality and reliability of the produced maps or data. It involves comparing the classified or interpreted results with ground truth data or reference data to measure the level of agreement or accuracy (Setiawan, et al.,2019).

The use of accurate and reliable assessment techniques is important because it provides quantitative measures of accuracy, which helps in evaluating the reliability and usability of the data. By conducting accuracy assessments, one can identify areas of improvement, limitations, and potential errors in the mapping or monitoring process. This allows for data refinement and enhancements, leading to more accurate and precise results (Lu, & Weng, 2007).

In conclusion, the use of accurate assessment techniques such as overall accuracy, users' accuracy, producers' accuracy, Kappa statistics, McNemar's Test Statistics, and F-scores, along with the utilization of reliable tools like Google Earth Engine, Landsat imagery, XGBOOST, SVM, LANDSCAPE METRIX, Planet GIS, and QGIS is of utmost importance. These tools and techniques enable accurate and comprehensive assessments for land cover and land use change, as well as the monitoring of urban growth in Polokwane municipality. They provide valuable insights into the quality and reliability of data, facilitating informed decision-making and effective planning for sustainable development and environmental management.

CHAPTER 4: RESULTS INTERPRETATION AND DISCUSSION

This chapter embodies the results and discussion of this study. The analysis of Landsat 5 TM and Landsat 8 OLI data gave quantification and prognosis results for two classifiers: SVM and gradient tree boosting, which delivered a comprehensive understanding of LULC detection over the period of 15 years from 2008 to 2022 in Polokwane area.

4.1. LAND COVER/ USE CLASSIFICATION

True colour composite (TCC) of Landsat was used to digitize many training sites, and imagery as well as attributes were generated from Google Earth Engine using expert knowledge as the basis.

Utilizing the South African classification system, 1996, all the Landsat images were separated into four primary categories: buildup, waterbody, business, baresoil and agriculture area. This was accomplished with the help of supervised gradient boost and SVM on Google Earth Engine (Intergraph, 2014).

Magidi & Ahmed, (2019:42) highlighted that, XGBoost is used in most research because it uses regression analysis to geocode Landsat TM and OLI satellite data, which was then geocoded to WGS 84 coordinates. It was also brought to WGS 84, and satellite images were clipped with Polokwane's administrative boundaries. Through Google Earth Engine, satellite data was used to figure out how the land is utilized in the area. Traditional boosting algorithms tend to overfit and have more parameters to use than ML algorithms like RF and SVM. SVM was also used, "wherein a decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements" (Magidi & Ahmed, 2019:42).

4.2. RESULTS FROM OBSERVATIONS AND STRUCTUAL MEETINGS AND DIALOGUES WITH THE PUBLIC

The purpose of this report is to present the observations and findings from the collection of ground truth data for land use and land cover change in the Polokwane area. The data was obtained from both public and formal meetings held in the region. The information gathered is significant for evaluating changes in land use and land cover over time, aiding in urban planning, and promoting sustainable land management practices.

public meetings were held at different locations in the Polokwane area. - Discussions and debates were conducted to gather expert opinions on the observations made and to identify potential challenges and opportunities.

4.2.1. Public meetings:

During the discussion, several participants drew attention to the substantial growth of urban areas over the past ten years, which has resulted in the encroachment upon previously untouched lands. This expansion of cities and towns posed a cause for concern for many, as it leads to the conversion of agricultural land into urban residential areas. Consequently, the region's potential for agriculture suffers as a result. Another issue that was emphasized by the attendees was the extensive development of infrastructure in the region. Projects such as the construction of new roads, shopping centers, and other similar ventures were cited as clear indicators of the rapid pace of urbanization. This rapid urbanization trend has undeniable consequences for the environment and society. The meeting also shed light on the alarming deforestation observed in certain areas. In particular, the expansion of commercial activities was highlighted as a driving force behind the evident destruction of forests. The attendees expressed concern over the impact this deforestation has on the region's biodiversity and overall environmental health.

In summary, important topics discussed during the meeting included the expansion of urban areas, the conversion of agricultural land, the development of infrastructure, and the concerning levels of deforestation. These issues underscore the need for thoughtful and sustainable planning to mitigate the negative consequences of urban growth on the environment and society..

4.2.2. Formal meetings:

During the discussions, experts emphasized the importance of implementing stricter land use regulations to address the issue of haphazard urban development and preserve valuable agricultural lands. They stressed that such regulations would have significant policy implications. Additionally, participants expressed concerns about the negative environmental consequences associated with changes in land use. They highlighted issues such as increased pollution and habitat fragmentation as significant environmental impacts that need to be addressed.

The discussions also shed light on the socio-economic implications of land use changes. It was emphasized that while urban development is necessary, it must be balanced with the preservation of cultural heritage and the livelihoods of existing communities. This balance would ensure that the socio-economic fabric of these communities is not disrupted.

Moreover, stakeholders raised concerns about the accuracy of data related to land use and land cover. They highlighted the inadequacy and outdated nature of such data, emphasizing the need for more comprehensive and frequent monitoring to enable informed decision-making. This issue of data accuracy holds crucial importance in ensuring effective land use planning and management. The ground truth data collected from public and formal meetings in the Polokwane area revealed key observations regarding land use and land cover change. The findings highlighted the rapid urban expansion, conversion of agricultural lands, infrastructure development, and concerns around deforestation. Expert opinions emphasized the need for better land use policies, environmental considerations, and improved data accuracy. These observations and recommendations are essential for future land management planning and sustainable development in the Polokwane area.

4.3. RESULTS FOR QUANTIFYING LAND USE LAND COVER CHANGE USING LANDSCAPE METRICS / ACCURACY ASSESSMENTS AND F-SCORES

Monitoring and Detection of Changes in Land Use Land cover were also evaluated using landscape metrics, and there are two primary categories of landscape metrics: compositional and configurational metrics (Gustafson, 1998; Herold et al., 2005). OCA, PD, SPA, GPA, MPA, and NP are the several types of landscape metrics that were measured for this research (Mcgarigal et al., 2002; McGarigal et al., 2009).

It is anticipated that the value of the "number of patches, or NUMP, will rise in the event of nucleated urban development, but it will drop significantly in the event that the urban patches combine into a single homogenous patch (Seto and Fragkias, 2005; McGarigal et al., 2009). Total Edge (TE) is a measurement that is used to determine the overall perimeter of all the urban patches (Seto and Fragkias, 2005). The "mean patch edge" (MPE) refers to the typical amount of edge that is found in each patch (McGarigal et al., 2009).

After the remote sensing classification and analysis, landscape metrics for each map from 2008 to 2022 were calculated. Six landscape metrics were computed using the Landscape Metrics Plugin on QGIS (ESRI, 2015), which is a product of the FRAGSTATS software (McGarigal and Marks, 1995; Mcgarigal *et al.*, 2002; McGarigal *et al.*, 2009). Landscape metrics were chosen to calculate six main characteristics of the land use land cover areas, namely their absolute size, relative size, and complexity. To calculate the absolute size of urban features, these patch densities The greatest patch area, smallest patch area, mean patch area, overall core area, and number of patches were calculated using QGIS software (Subramanian, 2019).

The LULC maps that were created through the interpretation of satellite data were applied to landscape analysis tools to evaluate the structural characteristics of the various cover types. Even though landscape metrics were calculated for each LULC class. Tables 1-4 provide a list of the parameters that were considered in this research endeavour. Every metric was computed, and then the results were analysed for every value.

4.3.1. Landscape Matrix Analysis / Accuracy assessments and F-scores for XGradient_ Boost

Table 3:Landscape Matrix Analysis for SVM Classifier (2008, 2013, 2018 and 2022)

The table below gives results using patch matrix model within Landscape matrix analysis which provides a key to understanding land use systems and changes by interpreting quantitative landscape indicators (Hoechstetter et al. 2008). According to Turner (2010), Landscape matrix analysis helps quantify landscape metrics like fragmentation, connectivity, and heterogeneity, providing insights into ecological integrity. In Polokwane Municipality, it helps to assess land use policies, identify areas needing interventions, and explore social and economic implications of land use change, such as impacts on livelihoods, resource availability, and urban-rural dynamics.

Landscape Matrix SVM Classification_2008						
Class	**Patch Density**	**Greatest Patch area**	**Mean Patch area**	**Number of Patches**	**landscape Proportions (%)**	**Landcover (HA)**
Waterbody	2.84163E+08	1179900	25120	1366	0%	397
Baresoil	8.98819E-06	2873366100	74476	85941	67%	372200
Business	1.08522E-06	2717100	4960	9082	1%	2993
Agriculture	2.634E-06	5454900	7402	45808	2%	10841
Buildup	8.70834E-06	1046690100	35024	93910	31%	169588

TOTAL					100.00%	
Landscape Matrix SVM Classification_2013						
Class	**Patch Density**	**Greatest Patch area**	**Mean Patch area**	**Number of Patches**	**landscape Proportions (%)**	**Landcover (HA)**
Waterbody	1.71217E-07	1460700	5748	952	0%	547
Baresoil	9.64860E-06	2251402200	61052	53648	59%	327532
Business	7.18681E-07	1413900	4685	3996	0%	1872
Buildup	5.623E-06	874072800	41113	31266	23%	128544
Agriculture	1.20741E-05	535820400	14527	67134	18%	97524
TOTAL					100%	
Landscape Matrix SVM Classification_2017						
Class	**Patch Density**	**Greatest Patch area**	**Mean Patch area**	**Number of Patches**	**landscape Proportions (%)**	**Landcover (HA)**
Waterbody	1.865045E-07	1057500	5139	1037	0%	533
Business	1.050864E-06	2202300	5168	5843	1%	3019
Buildp	6.733587E-06	681286500	32483	37440	22%	121617
Baresoil	9.202929E-06	1359333900	58745	51170	54%	300601
Agriculture	9.350406E-06	173853900	25053	51990	23%	130249
TOTAL					100.00%	
Landscape Matrix SVM Classification_2022						
Class	**Patch Density**	**Greatest Patch area**	**Mean Patch area**	**Number of Patches**	**landscape Proportions (%)**	**Landcover (HA)**
Waterbody	1.532323E-07	1305000	6542	852	0%	557
Business	1.123164E-06	2221200	4817	6245	1%	3008
Buildup	5.314930E-06	900565200	46611	29552	25%	137746
Baresoil	8.479032E-06	1416383100	59477	47145	50%	280406
Agriculture	7.731575E-06	617360400	31241	42989	24%	134301
TOTAL					100%	

Landscape metrics aided in the quantification of the landscape structures and conveyed the extent of the changes and their effects. In the present study, landscape metrics have been calculated only at the class level because the main objectives are to study the changes in LULC classes. Patch Density (PD), greatest patch area (GPA), mean patch area (MPA), number of patches (NP), landscape proportions (LP) and Landcover area calculated in hectors. The results for support vector classifier were calculated and in 2008 waterbody occupied GPA of 1179900, baresoil GPA of 2873366100, business GPA of 2717100, agriculture GPA of 5454900 and lastly buildup GPA of 1046690100. During 2013 waterbody occupied GPA of 1460700, baresoil GPA of 2251402200, business GPA of 1413900, agriculture GPA of 5454900 and lastly buildup GPA of 874072800. The results for 2017 were waterbody GPA 1057500, business GPA 2202300, buildup GPA of 681286500, baresoil yielded GPA of 1359333900 and agriculture yielded 173853900. Lastly, in 2022 waterbody have the results 1305000 GPA, business 2221200 GPA, buildup 900565200 GPA, baresoil 1416383100 GPA and agriculture had a GPA of 617360400.

Through the analysis under the greatest area from 2008 to 2022, baresoil has registered the largest patch in 2008 and decreased in 2013 while the buildup also had greatest patch area in 2008 and reduced in 2013. Baresoil was reducing from 2008 to 2022 with increase in business area from 2008 to 2022. Agriculture greatest patch was also increasing from 2008 (5454900) - 2022 to an amount of 617360400. In terms of landscape proportions, baresoil occupied more space in 2008 than other classes to an amount of 67% followed by the buildup with an amount of 31%. The least landscape proportions in 2008 was agriculture to an amount of 2% with business site to an amount of 1%. Baresoil also had large number of PD at 8.98819E-06 followed by buildup to 8.70834E-06. It shows that in 2008 the two classes baresoil and buildiup were dominant. During 2013 baresoil (59%)and buildup(23%) area were reducing in landscape proportions as a result of agriculture increasing to an amount of 18% . waterbody and business area did not increase in 2013, they remained at 0%.

In 2013, the GP was taken up by baresoil, but baresoil was also reducing from 2008 and agriculture was increasing to an amount of 535820400 from 5454900. A massive increase of agriculture in 2013. In 2017, there was cultivation of agriculture area were it reduced from 535820400 GP to 173853900. Baresoil was also reducing in GP to an amount of 1359333900, buildup was also reducing GP from 874072800 to 681286500. The large landscape proportion was still accounted to baresoil (54%) and in 2017 buildup area (22%) and agriculture(23%) were closer to each other. Even though they were closer to each other, it was an increase in agriculture by 5% and a decrease in buildup by 1%. Agriculture and baresoil occupied large number of PD and followed by the buildup area to 6.733587E-06. During 2022, baresoil reduced by 4% to 50% landscape proportion, while the LP of buildup increased by 3% to be 25%. Waterbody has increased the GPA to 1305000 from 1057500 in 2017, it shows that there has been some wetlands or river increase in Polokwane. There was a decrease in baresoil (50%) and as a result of 3% increase in buildup and as a result of an increase in agriculture by 1% from 23% in 2017. Baresoil had 1416383100 GPA and buildup to 900565200 GPA. The area for waterbody was increasing from 2008 with an amount of 397 HA and has increased 557HA in 2022. It shows an expansion of waterbody in Polokwane over the past 15 years. Between 2008 and 2013, the area of baresoil has increase from 372200HA to 327532HA. The area of business has reduced in 2008 from 2993 HA to 1872 as well as the decrease of area in buildup from 169588HA to 128544HA. Between 2013 to 2017, business area has increased from 1872HA to 3019HA as well as the number of buildup has reduced from 128544HA to 121617. In 2022 buildup area has increased from 121617HA to 137746HA, which shows an expansion of 16129HA.

The results for landscape composition was calculated using two landscape matrix namely mean patch area (MPA) and number of patches (NP). It is noted that the increase in number of patches decrease the mean patch area. The results for waterbody in 2008 has the higher number of MPA (25120) with the corresponding decrease in NP of 1366. The MPA of baresoil in 2008 account to 74476 and an increase number on patches of 85941. During 2008 buildup area has the more NP of 93910 than the other classes and followed by baresoil with an amount of 85941. Waterbody has the less NP of patches. Agriculture has more NP in 2013 to an amount of 67134 followed by the NP of baresoil. In 2022 baresoil and shows a fragmentation with an amount of 51170 in 2017 to 47145 in 2022. There was also a fragmentation of agriculture from 51990 to 42989 as a result of the increase in business area from 5843 in 2017 to 6245 in 2022.

Table 4:Landscape Matrix Analysis for XGBoost Classifier (2008, 2013, 2018 and 2022)

The table below gives results using patch matrix model within Landscape matrix analysis which provides a key to understanding land use systems and changes by interpreting quantitative landscape indicators (Hoechstetter et al. 2008). According to Turner (2010), Landscape matrix analysis helps quantify landscape metrics like fragmentation, connectivity, and heterogeneity, providing insights into ecological integrity. In Polokwane Municipality, it helps to assess land use policies, identify areas needing interventions, and explore social and economic implications of land use change, such as impacts on livelihoods, resource availability, and urban-rural dynamics.

Landscape Matrix XGboost Classification_2008						
Class	**Patch Density**	**Greatest Patch area**	**Mean Patch area**	**Number of Patches**	**landscape Proportions (%)**	**Landcover (HA)**
Waterbody	2.45675E-07	1206900	11378	1366	0%	1554
Baresoil	1.54565E-05	2094881400	35795	85941	55%	307626
Business	1.63340E-06	2493000	3343	9082	1%	3036
Agriculture	8.239E-06	1028766600	40778	45808	34%	186797
Buildup	1.68897E-05	34471800	6070	93910	10%	57005
TOTAL					100%	

Landscape Matrix XGboost Classification_2013						
Class	**Patch Density**	**Greatest Patch area**	**Mean Patch area**	**Number of Patches**	**landscape Proportions (%)**	**Landcover (HA)**
Waterbody	8.95294E-07	1360800	11378	4978	0%	1021
Baresoil	1.21652E-05	1851621300	35795	67641	55%	305414
Business	1.84580E-06	2062800	3343	10263	1%	4044
Buildup	1.103E-05	908986500	40778	61327	33%	181967
Agriculture	1.85098E-05	73776600	6070	102918	11%	63573
TOTAL					100%	

Landscape Matrix XGBoost Classification_2017						
Class	**Patch Density**	**Greatest Patch area**	**Mean Patch area**	**Number of Patches**	**landscape Proportions (%)**	**Landcover (HA)**
Waterbody	1.593472E-07	990000	3914	1366	0%	347
Business	2.073851E-06	1416600	4474	85941	1%	5159
Buildp	1.269544E-05	784619100	19737	9082	25%	139319
Baresoil	1.242009E-05	1489217400	35371	45808	44%	244268

Agriculture	1.577231E-05	348100200	19034	93910	30%	166925
TOTAL					100.00%	
Landscape Matrix XGboost Classification_2022						
Class	**Patch Density**	**Greatest Patch area**	**Mean Patch area**	**Number of Patches**	**landscape Proportions (%)**	**Landcover (HA)**
Waterbody	1.721165E-07	1075500	4314	957	0%	413
Business	1.723144E-06	2320200	4076	9581	1%	3905
Buildup	5.518160E-06	573962400	36688	30682	20%	112567
Baresoil	6.303206E-06	2153204100	74333	35047	47%	260516
Agriculture	1.163504E-05	783746100	27610	64693	32%	178617
TOTAL					100%	

The results for XGboost classifier were calculated and in 2008 waterbody occupied GPA of 1206900, baresoil GPA of 2094881400, business GPA of 2493000, agriculture GPA of 1028766600 and lastly buildup GPA of 34471800. During 2013 waterbody occupied GPA of 1360800, baresoil GPA of 1851621300, business GPA of 2062800, agriculture GPA of 73776600 and lastly buildup GPA of 908986500. The results for 2017 were waterbody GPA 990000, business GPA 1416600, buildup GPA of 784619100, baresoil yielded GPA of 1489217400 and agriculture yielded 348100200. Lastly, in 2022 waterbody have the results 1075500 GPA, business 2320200 GPA, buildup 573962400 GPA, baresoil 2153204100 GPA and agriculture had a GPA of 783746100. From 2008 to 2022 the waterbody has increased and the number of baresoil was decreasing from 2008-2013 and has increased once more in 2022. Agriculture has been decreasing from 2008 to 2022 due to fragmentation accruing and as a result buildup area was increasing from 2008 to 2022. The greatest landscape proportion was taken up by baresoil in 2008 with a proportion of 55% and agriculture was the second one to occupy the by 34%. Buildup has a least amount of 10% proportion and business occurred with only 1% of the landscape proportion. In 2013 baresoil remained with a proportion of 55% as well business remained with the proportion of 1%. Only buildup area has increased by 23% while agriculture reduced to 11%. There was fragmentation of agriculture area as a result of an increase in buildup area. Between the year 2017 and 2022 agriculture has increased from 30% to 32% and baresoil has increased from 44% to 47%. Buildup was also reducing from 25% to 20%. Buildup shows an increase in landcover from 2008 with an amount of 57005 HA to 112567 HA, this is also evident from the PD with a value of 5.518160E-06 from 1.68897E-05.

Through the analysis under the greatest area from 2008 to 2022, baresoil has registered the largest patch in 2008 and decreased in 2013 while the buildup also had greatest patch area in 2008 and also reduced in 2013. Baresoil was reducing from 2008 to 2022 with increase in business area from 2008 to 2022. Agriculture greatest patch was also increasing from 2008 (5454900) - 2022 to an amount of 617360400. In terms of landscape proportions, baresoil occupied more space in 2008 than other classes to an amount of 67% followed by the buildup with an amount of 31%. The least landscape proportions in 2008 was agriculture to an amount of 2% with business site to an amount of 1%. Baresoil also had large number of PD at 8.98819E-06 followed by buildup to 8.70834E-06. It shows that in 2008 the two classes baresoil and buildiup were dominant. During 2013 baresoil (59%)and buildup(23%) area were reducing in landscape proportions as a result of agriculture increasing to an amount of 18% . waterbody and business area did not increase in 2013, they remained at 0%.

The results for landscape composition was calculated using two landscape metrix namely mean patch area (MPA) and number of patches (NP). It is noted that the increase in number of patches decrease the mean patch area. The results for waterbody in 2008 has the higher number of MPA (11378) with the corresponding decrease in NP of 1366. The MPA of baresoil in 2008 came to 35795 and an increase number of patches to 85941. There was an increase in NP of business area between 2008 to 2013 to an amount of 9082 to 10263. Fragmentation occurred in agriculture between the year 2017 to 93910, the NP decreased from 93910 with 19034 MPA to 64693 NP with 27610 MPA. There was a huge fragmentation between 2017 and 2022 that has occurred in business area from NP of 85941 to 9581. The MPA of baresoil in 2022 has increased from 74333 as a result of decrease in NP to 35047

4.4. RESULTS AND DATA INTERPRETATION OF LULC CLASSIFICATION AND ACCURACIES

The land use Land cover maps of 2008 (Figure 1), 2013 (Figure 2), 2017 (Figure 3), and 2022 (Figure 4), based on supervised classification and visual interpretation, showed changes in urban areas between the years. Overall classification accuracy and landscape matrix, as well as the area of each land use and land cover, users' accuracy, and Producers accuracy.

4.4.1 Results for Gradient Boosting classification in 2008, 2013, 2017 and 2022

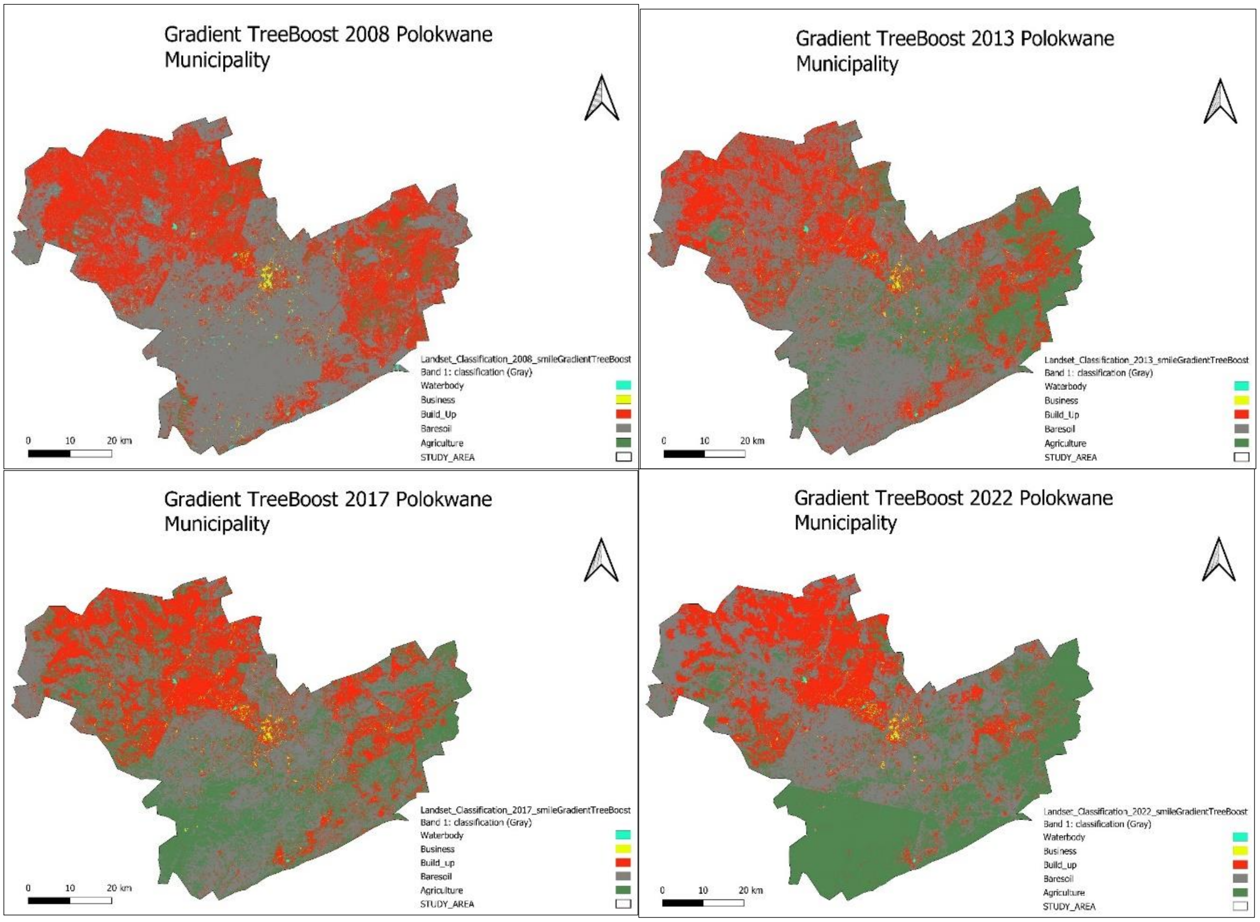


Figure 2:XG Boost classification results (2008, 2013, 2017, 2022)

4.4.2 Results for Support Vector Machine classification in 2008, 2013, 2017 and 2022

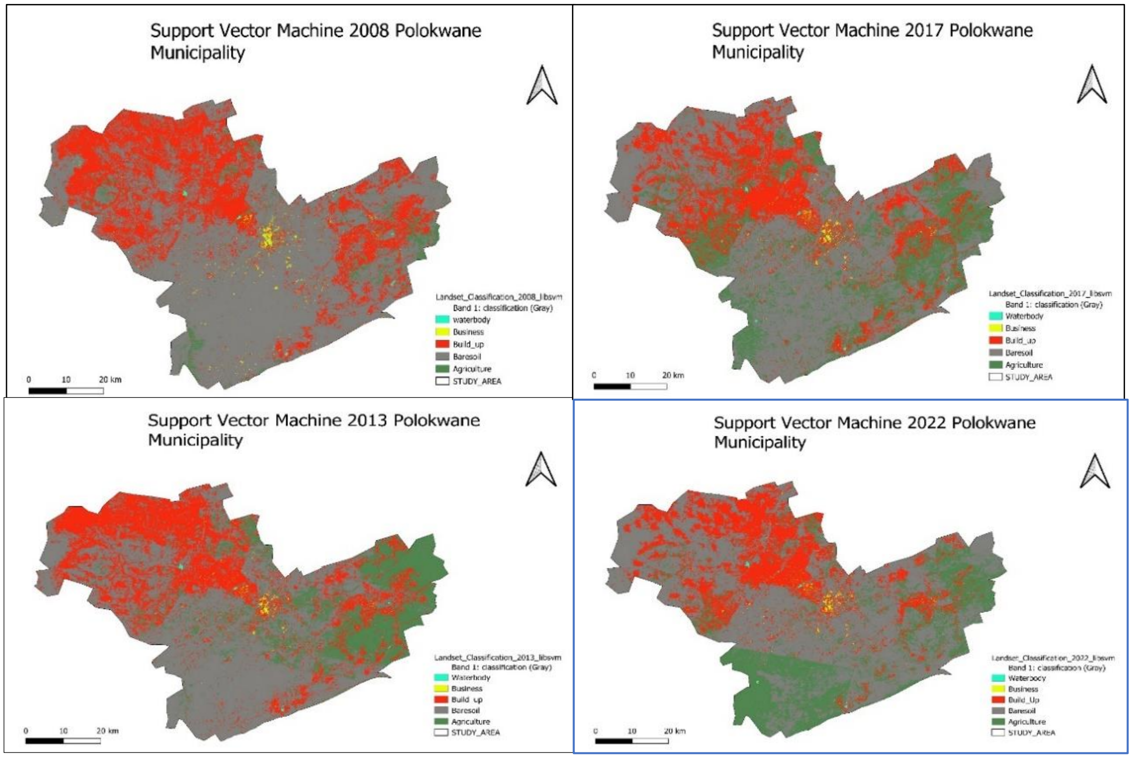


Figure 3:SVM classification results (2008, 2013, 2017, 2022)

4**.4.3 Results for Validation test of XGBoost classifier** (2008, 2013, 2018 and 2022)

Table 5:Accuracy assessments, Confusion Matrix and F-scores for the validation samples results of XGBoost classification (2008, 2013, 2018 and 2022)

Accuracy Assessment For XGBRFClassification Landsat (2008) for Training data 30%

Reference Data

	Waterbody	**Baresoil**	**Business**	**Buildup**	**Agricultiure**	**Total**	**User'Accuracy**
Waterbody	22	0	0	0	0	22	1,00
Baresoil	0	39	0	16	3	58	0,67
Business	1	0	13	0	2	16	0,81
Buildup	0	0	0	14	1	15	0,93
Agricultiure	0	6	3	2	17	28	0,61
Total	23	45	16	32	23	139	**Kappa Index**
Producer's Accuracy	0,96	0,87	0,81	0,44	0,74	3,81	0,68
F1	0,98	0,76	0,81	0,60	0,67		**Overall Accuracy**
							0,75

Accuracy Assessment For XGBoost Classification Landsat (2013) for Training data 30%

Reference Data

	Waterbody	Baresoil	Business	Agriculture	Buildup	Total	User'Accuracy
Waterbody	11	0	0	0	0	11	1,00
Baresoil	0	32	1	4	10	47	0,68
Business	0	0	49	5	0	54	0,91
Agriculture	1	8	3	37	7	56	0,66
Buildup	0	1	0	0	6	7	0,86
Total	12	41	53	46	23	175	**Kappa Index**
Producer's Accuracy	0,92	0,78	0,92	0,80	0,26	3,69	0,69
F1	0,96	0,73	0,92	0,73	0,40		**Overall Accuracy**
							0,77

Accuracy Assessment For XGBoost Classification Landsat (2017) for Test data 30%

Reference Data

	Waterbody	Business	Buildup	Baresoil	Agricultiure	Total	User'Accuracy
Waterbody	19	0	0	0	0	19	1,00
Business	1	26	1	0	0	28	0,93
Buildup	0	2	34	2	9	47	0,72
Baresoil	0	0	6	37	10	53	0,70
Agricultiure	0	0	9	10	37	56	0,66
Total	20	28	50	49	56	203	**Kappa Index**
Producer's Accuracy	0,95	0,93	0,68	0,76	0,66	3,974387755	0,68
F1	0,97	0,93	0,70	0,70	0,66		**Overall Accuracy**
							0,90

Accuracy Assessment For XGBoost Classification Landsat (2022) for Test data 30%

Reference Data

	Waterbody	Business	Buildup	Baresoil	Agricultiure	Total	User'Accuracy
Waterbody	3	0	0	0	0	3	1,00
Business	1	33	2	0	0	36	0,92
Buildup	0	3	21	1	3	28	0,75
Baresoil	0	0	5	22	7	34	0,65
Agricultiure	0	0	1	4	22	27	0,81
Total	4	36	29	27	32	128	**Kappa Index**
Producer's Accuracy	0,75	0,92	0,72	0,81	0,69	3,893119413	0,72
F1	0,86	0,92	0,74	0,81	0,75		**Overall Accuracy**
							0,89

The results above for validation test on land use and land cover (LULC) classification, XG Boost emerged as the top performer in terms of f-scores and accuracy assessments. XG Boost achieved significant success in accurately monitor/evaluate and classifying different land use categories based on the provided dataset. This algorithm outperformed SVM classification model and exhibited superior performance in terms of both precision and recall. Additionally, XG Boost proved to be highly accurate in correctly identifying and classifying various land cover types within the validation test

4.4.4. **Results for Validation test of SVM classifier (2008, 2013, 2018 and 2022)**

Table 6:Accuracy assessments, Confusion Matrix and F-scores for the validation samples results of SVM classification (2008, 2013, 2018 and 2022)

Accuracy Assessment For SVMClassification Landsat (2008) for Test data 30%

Reference Data

	Waterbody	**Baresoil**	**Business**	**Agriculture**	**Buildup**	**Total**	**User'Accuracy**
Waterbody	22	0	0	0	0	22	1,00
Baresoil	0	39	0	16	3	58	0,67
Business	1	0	13	0	2	16	0,81
Agriculture	0	0	0	14	1	15	0,93
Buildup	0	6	3	2	17	28	0,61
Total	23	45	16	32	23	139	**Kappa Index**
Producer's Accuracy	0,96	0,87	0,81	0,44	0,74	3,81	0,68
F1	0,98	0,76	0,81	0,60	0,67		**Overall Accuracy**
							0,75

Accuracy Assessment For SVM Classification Landsat (2013) for Test data 30%

Reference Data

	Waterbody	**Baresoil**	**Business**	**Buildup**	**Agricultiure**	**Total**	**User'Accuracy**
Waterbody	14	0	2	0	0	16	0,88
Baresoil	0	27	1	5	7	40	0,68
Business	3	0	35	3	0	41	0,85
Buildup	0	6	4	26	2	38	0,68
Agricultiure	0	0	0	7	23	30	0,77
Total	17	33	42	41	32	165	**Kappa Index**
Producer's Accuracy	0,82	0,82	0,83	0,63	0,72	3,827940905	0,69
F1	0,85	0,74	0,84	0,69	0,74		**Overall Accuracy**
							0,75

Accuracy Assessment For SVM Classification Landsat (2017) for Test data 30%

Reference Data

	Waterbody	**Business**	**Buildup**	**Baresoil**	**Agricultiure**	**Total**	**User'Accuracy**
Waterbody	19	0	0	0	0	19	1,00
Business	1	26	0	0	0	27	0,96
Buildup	0	2	36	4	8	50	0,72
Baresoil	0	0	7	35	13	55	0,64
Agricultiure	0	0	7	10	35	52	0,67
Total	20	28	50	49	56	203	**Kappa Index**
Producer's Accuracy	0,95	0,93	0,72	0,71	0,63	3,937857143	0,67
F1	0,97	0,95	0,72	0,69	0,65		**Overall Accuracy**
							0,74

Accuracy Assessment For SVM Classification Landsat (2022) for Test data 30%

Reference Data	Waterbody	Business	Buildup	Baresoil	Agricultiure	Total	User'Accuracy
Waterbody	3	0	0	0	3	6	0,50
Business	1	33	2	0	0	36	0,92
Buildup	0	3	22	2	4	31	0,71
Baresoil	0	0	5	22	3	30	0,73
Agricultiure	0	0	0	3	22	25	0,88
Total	4	36	29	27	32	128	**Kappa Index**
Producer's Accuracy	0,75	0,92	0,76	0,81	0,69	3,927602171	0,73
F1	0,60	0,92	0,73	0,85	0,77		**Overall Accuracy**
							0,79

The results above shows that validation test for land use and land cover (LULC) classification, the Support Vector Machine (SVM) algorithm was found to be outperformed by the XG Boost algorithm in terms of f-scores and accuracy assessments. The SVM algorithm, which is commonly used for classification tasks, did not achieve the same level of accuracy and predictive power as the XG Boost algorithm. This result suggests that XG Boost was more effective in accurately classifying different land use and cover categories in the validation test.

4.4.5 Accuracy assessment results for XGBoost (2008, 2013, 2018 and 2022)

Table 7:Accuracy assessments, Confusion Matrix and F-scores for the training samples results of XGBoost classification (2008, 2013, 2018 and 2022)

Accuracy Assessment for XGB Classification Landsat (2008) for Training data 70%

Classified Data / Reference Data	Waterbody	Baresoil	Business	Agriculture	Buildup	Total	User'Accuracy
Waterbody	37	0	0	0	0	37	1,00
Baresoil	0	82	6	19	9	116	0,71
Business	0	0	31	0	2	33	0,94
Agiculture	0	2	0	16	1	19	0,84
Buildup	0	5	2	3	46	56	0,82
Total	37	89	39	38	58	261	**Kappa Index**
Producer's Accuracy	1,00	0,92	0,79	0,42	0,79	3,14	0,68
F1	1,00	0,80	0,86	0,56	0,81		**Overall Accuracy**
							0,82

Accuracy Assessment for XGBoost Classification Landsat (2013) for Training data 70%

Reference Data

Classified Data

	Waterbody	Baresoil	Business	Agriculture	Buildup	Total	User'Accuracy
Waterbody	35	0	0	1	1	37	0,95
Baresoil	0	76	1	6	3	86	0,88
Business	0	0	99	3	0	102	0,97
Agriculture	0	4	4	82	7	97	0,85
Buildup	0	3	0	2	53	58	0,91
Total	35	83	104	94	64	380	**Kappa Index**
Producer's Accuracy	1,00	0,92	0,95	0,87	0,83	4,57	0,69
F1	0,97	0,90	0,96	0,86	0,87		**Overall Accuracy**
							0,91

Accuracy Assessment for XGBoost Classification Landsat (2017) for Test data 70%

Reference Data

Classified Data

	Waterbody	Business	Buildup	Baresoil	Agricultiure	Total	User'Accuracy
Waterbody	44	0	0	0	0	44	1,00
Business	0	61	3	0	0	64	0,95
Buildup	3	4	91	3	5	106	0,86
Baresoil	0	0	4	94	5	103	0,91
Agricultiure	0	0	6	5	84	95	0,88
Total	47	65	104	102	94	412	**Kappa Index**
Producer's Accuracy	0,94	0,94	0,88	0,92	0,89	4,564817	0,68
F1	0,97	0,95	0,87	0,90	0,89		**Overall Accuracy**
							0,90

Accuracy Assessment for XGBoost Classification Landsat (2022) for Test data 70%

Reference Data

Classified Data

	Waterbody	Business	Buildup	Baresoil	Agriculture	Total	User'Accuracy
Waterbody	17	0	0	1	0	18	0,94
Business	0	74	4	0	0	78	0,95
Buildup	0	3	71	0	4	78	0,91
Baresoil	0	0	7	80	9	96	0,83
Agriculture	0	0	2	6	56	64	0,88
Total	17	77	84	87	69	334	**Kappa Index**
Producer's Accuracy	1,00	0,96	0,85	0,92	0,81	4,537411	0,72
F1	0,97	0,95	0,88	0,90	0,84		**Overall Accuracy**
							0,89

Considering the categories accuracy, random method provided very high accuracy assessment for the classified classes. From the above table, the producers' accuracy of the classes in Gradient Boost 2008 shows that waterbodies were classified with 100% accuracy, while business classes were classified with 79% accuracy, which shows excellent classification. Build-up was also classified as having 79% good accuracy. Bare soil was classified at 92% excellent accuracy as well as agriculture at 42%, which is very low accuracy and not reliable. Business class obtained higher accuracy of the PA and waterbody obtained 100% PA from all the classes.

Users' accuracy of waterbodies was classified at 100% accuracy; this shows a good map and a reliable map. Business class shows 94% reliability of the map. Build-ups have an accuracy of 82% reliability of the map, while Bare soil shows a 71% moderate level of reliability of the map. Lastly, agriculture have the results of 84% reliability of the map, which is a good map to use for reference. UA accuracy obtained from all class shows that the classification results are reliable as UA shows the accuracy of the reality on ground.

The classification obtained the excellent accuracy of 82% which shows an excellent reliability of map and tells that only 18% of the data is not explained in the classification. Well the Kappa statistics measures the difference between the true agreement of classified map and chance agreement of random classifier compared to reference data (Lillesans et al., 2004). It is stated that Kappa values of more 0.80 indicate good classification performance. Kappa values between 0.40 to 0.80 indicate moderate classification performance and Kappa values of less than 0.40 indicate poor classification performance (Jensen, 2005, Lillesans et al., 2004). XGBoost classifier in 2008 produced a moderate classification map with a Kappa of 68%.

The producers accuracy were obtained in waterbody, baresoil, business, agriculture and buildup in 2013. The results were 100%, 92%, 95%, 87% and 83% respectively. The users accuracy were also obtained to 95%, 88%, 97%, 85% and 91% respectively. all UA accuracies obtained are above 85% USGS standard which means the classifier has produced 85% and more true ground data and the map is reliable. The overall accuracy in 2013 yielded to 91% excellent accuracy of the training samples. This explains that only 9% of the data is not explained by the classification. The Kappa statistics of the data obtained to 0.69 moderate classification performance.

The results obtained producers accuracy in 2017 of waterbody, business, buildup, baresoil and agriculture were 94%, 94%, 88%, 92% and 89%. The user accuracy obtained were 100%, 95%, 86%, 91% and 88% respectively. all UA accuracies obtained are above 85% USGS standard which means the classifier has produced 85% and more true data of ground and the map is reliable. The overall accuracy in 2013 yielded to 90% excellent accuracy of the training samples. This explains that only 10% of the data is not explained by the classification. The Kappa statistics of the data obtained to 0.68 moderate level of reliability.

The producers accuracy were obtained in waterbody, business, buildup, baresoil and agriculture in 2022. The results were 100%, 96%, 85%, 92% and 81% respectively. The users accuracy were also obtained to 94%, 95%, 91%, 83% and 88% respectively. The UA accuracies obtained were higher which shows that the classifier has produce more true ground data which is from 83% level of reliability. The overall accuracy in 2013 yielded to 89% excellent accuracy of the training samples. This explains that only 11% of the data is not explained by the classification. The Kappa statistics of the data obtained to 0.72 moderate classification performance.

The f-scores were also calculated to conclude the best classifier between XGBoost and SVM classifier. The results in 2008 with XGBoost shows an average of 0.56-1.00 , in 2013 an average of 0.86-0.97 excellent classification, 2017 an average of 0.87-097 and lastly in 2022 an average of 0.84-0.97 was obtained.

In conclusion, According to the accuracy assessments, confusion matrix, and F-scores, XGBoost algorithm significantly outperformed the Support Vector Machine (SVM) in monitoring land use and land cover (LULC) changes in Polokwane municipality. These evaluations provide crucial insights into the performance of different classification algorithms and can aid decision-makers in understanding the patterns and dynamics of LULC change.

According to Chen et al., (2016), the accuracy assessments provide an overall measure of how accurately the algorithms classified LULC changes in the study area. XGBoost demonstrated a higher accuracy rate compared to SVM, indicating its superior performance in correctly classifying different land use and land cover types. The confusion matrix highlights the specific errors made by each algorithm, allowing for a more detailed analysis of misclassifications. XGBoost obtained a lower number of misclassified classes, indicating its ability to correctly assign pixels to their respective LULC categories more effectively than SVM.

Furthermore, the F-scores, which combine precision and recall, also revealed the higher performance of XGBoost in capturing LULC changes. The F-scores measure the algorithms' ability to balance accuracy and completeness in classifying different land use and land cover types. XGBoost achieved higher F-scores, suggesting its superior performance in accurately identifying and mapping LULC changes in the municipality.

The findings of XGBoost performing better than SVM in monitoring LULC change in Polokwane municipality have significant implications for land management and planning. Understanding the dynamics of LULC change is crucial for sustainable development, resource management, and decision-making in urban and rural areas. The accuracy, precision, and completeness provided by XGBoost can help planners and policymakers to take informed actions for conserving natural resources, improving land use practices, and mitigating the negative impacts of LULC change on the environment.

Lastly, the accuracy assessments, confusion matrix, and F-scores demonstrate that XGBoost algorithm outperformed SVM in monitoring LULC changes in Polokwane municipality. These results highlight the potential of XGBoost as a powerful tool for analysing and mapping LULC change dynamics, providing valuable information for land management and planning purposes.

4.4.6 Accuracy assessment results for SVM (2008, 2013, 2018 and 2022)

Table 8:Accuracy assessments, Confusion Matrix and F-scores for the training samples results of SVM classification (2008, 2013, 2018 and 2022)

Accuracy Assessment for SVM Classification Landsat (2008) for Training data 70%

Reference Data

Classified Data

	Waterbody	**Baresoil**	**Business**	**Agriculture**	**Buildup**	**Total**	**User'Accuracy**
Waterbody	37	0	0	0	0	37	1,00
Baresoil	0	82	6	19	9	116	0,71
Business	0	0	31	0	2	33	0,94
Agriculture	0	0	0	16	1	17	0,94
Buildup	0	5	2	3	46	56	0,82
Total	37	87	39	38	58	259	**Kappa Index**
Producer's Accuracy	1,00	0,94	0,79	0,42	0,79	3,95	0,68
F1	1,00	0,81	0,86	0,58	0,81		**Overall Accuracy**
							0,82

Accuracy Assessment For SVM Classification Landsat (2013) for Test data 70%

Reference Data

Classified Data

	Waterbody	**Baresoil**	**Business**	**Buildup**	**Agricultiure**	**Total**	**User'Accuracy**
Waterbody	27	0	0	0	1	28	0,96
Baresoil	0	74	1	12	16	103	0,72
Business	1	0	104	5	0	110	0,95
Buildup	0	15	8	74	5	102	0,73
Agriculture	0	2	0	8	33	43	0,77
Total	28	91	113	99	55	386	**Kappa Index**
Producer's Accuracy	0,96	0,81	0,92	0,75	0,60	4,045301	0,69
F1	0,96	0,76	0,93	0,76	0,67		**Overall Accuracy**
							0,80

Accuracy Assessment for SVM Classification Landsat (2017) for Test data 70%

Reference Data

Classified Data

	Waterbody	**Business**	**Buildup**	**Baresoil**	**Agriculture**	**Total**	**User 'Accuracy**
Waterbody	46	0	1	0	0	47	0,98
Business	1	58	6	0	0	65	0,89
Buildup	0	7	74	9	17	107	0,69
Baresoil	0	0	11	81	20	112	0,72
Agriculture	0	0	12	12	57	81	0,70
Total	47	65	104	102	94	412	**Kappa Index**
Producer's Accuracy	0,98	0,89	0,71	0,79	0,61	3,98307	0,67
F1	0,98	0,89	0,70	0,75	0,65		**Overall Accuracy**
							0,76

Accuracy Assessment for SVM Classification Landsat (2022) for Test data 70%

Reference Data

Classified Data

	Waterbody	**Business**	**Buildup**	**Baresoil**	**Agriculture**	**Total**	**User'Accuracy**
Waterbody	17	0	0	0	1	18	0,94
Business	0	70	8	0	0	78	0,90
Buildup	0	6	68	1	7	82	0,83
Baresoil	0	1	4	74	13	92	0,80
agriculture	0	0	4	11	48	63	0,76
Total	17	77	84	86	69	333	**Kappa Index**
Producer's Accuracy	1,00	0,91	0,81	0,86	0,70	4,274732	0,73
F1	0,97	0,90	0,82	0,81	0,73		**Overall Accuracy**
							0,83

From the above table, the producers' accuracy of the classes in SVM 2008 shows that waterbodies were classified with 100% accuracy, while baresoil classes were classified with 94% accuracy, which shows excellent classification. Business was also classified as having 79% good accuracy. buildup was classified at moderate accuracy of 79% as well as agriculture at 42%, which is very low accuracy and not reliable. Waterbody obtained 100% accuracy sampled and baresoil show an excellent accuracy of 94%.

Users' accuracy of waterbodies was classified at 100% accuracy; this shows a good map and a reliable map. Baresoil class shows 72% moderate level of reliability of the map. Business has an accuracy of 94% excellent level of reliability of the map, while agriculture shows a 94% excellent level of reliability of the map. Lastly, buildup have the results of 82% reliability of the map, which is an excellent map to use for reference. The UA accuracies obtained were higher which shows that the classifier has produce more true ground data from the 71% level of reliability. The overall accuracy in 2008 yielded to 82% excellent accuracy of the training samples. This explains that only 18% of the data is not explained by the classification. The Kappa statistics of the data obtained to 0.68 moderate classification performance.

The producers accuracy were obtained in waterbody, baresoil, business, agriculture and buildup in 2013. The results were 96%, 81%, 92%, 75% and 60% respectively. The users accuracy were also obtained to 96%, 72%, 95%, 73% and 77% respectively. The UA accuracies obtained were higher which shows that the classifier has produce more true ground data from the 72% level of reliability. The overall accuracy in 2013 yielded to 80% excellent accuracy of the training samples. This explains that only 18% of the data is not explained by the classification. The Kappa statistics of the data obtained to 0.69 moderate classification performance.

The results obtained producers accuracy in 2017 waterbodies were classified with 98% accuracy, while business class were classified with 89% accuracy, which shows excellent classification. Buildup was also classified as having 71% good accuracy. Baresoil was classified at moderate accuracy of 79% as well as agriculture at 61%, which is moderate map. The user accuracy obtained were 94%, 90%, 83%, 80% and 76% respectively. The UA accuracies obtained were higher which shows that the classifier has produce more true ground data from the 76% level of reliability. The overall accuracy in 2017 yielded to 76% excellent accuracy of the training samples. This explains that only 24% of the data is not explained by the classification. The Kappa statistics of the data obtained to 0.67 moderate classification performance.

The producers accuracy obtained in 2022 were waterbody, business, buildup, baresoil and agriculture. The results were 100%, 91%, 81%, 86% and 70% respectively. The users accuracy were also obtained to 94%, 90%, 83%, 80% and 76% respectively. The UA accuracies obtained were higher which shows the classifier has produce more true ground data from 76% level of map is reliable. The overall accuracy in 2020 yielded to 83% excellent accuracy of the training samples. This explains that only 17% of the data is not explained by the classification. The Kappa statistics of the data obtained to 0.73 moderate classification performance.

The f-scores were also calculated to conclude the best classifier between XGBoost and SVM classifier. The results in 2008 with SVM shows an average of 0.58-1.00 , in 2013 an average of 0.60-0.96 excellent classification, 2017 an average of 0.65-0.98 and lastly in 2022 an average of 0.73-0.97 was obtained.

4.4.7 Table 9:McNemar's Test statistics between XG Boost and Support Vector Machine

Classifier	Support Vector Machine 2008			
XG Boost		Classified Incorrectly	Classified correctly	Total
	Classified Incorrectly	49	212	**261**
	Classified Correctly	47	212	**259**
	Total	**96**	**424**	**520**
	Outcome Value (Z)	**0,81**		
Classifier	Support Vector Machine 2013			
XG Boost		Classified Incorrectly	Classified correctly	Total
	Classified Incorrectly	35	312	**347**
	Classified Correctly	74	345	**419**
	Total	**109**		**766**
	Outcome Value (Z)	**0,79**		
Classifier	Support Vector Machine 2017			
XG Boost		Classified Incorrectly	Classified correctly	Total
	Classified Incorrectly	38	316	**354**
	Classified Correctly	96	374	**470**
	Total	**134**	**690**	**824**

	Outcome Value (Z)	**0,78**		
Classifier	Support Vector Machine 2022			
XG Boost		Classified Incorrectly	Classified correctly	Total
	Classified Incorrectly	36	277	**313**
	Classified Correctly	56	298	**354**
	Total	**92**	**575**	**667**
	Outcome Value (Z)	**0,76**		

From the above table it is noted that if the Z value is more than 1.96, there is a statistically significant difference in the classifiers' accuracy (Food, 2004: Leeuw et al., 2006). Therefore, during 2008 the z-score of the classifiers were 0.81 in 2013 was 0.79, during 2017 it came to 0.78 and lastly during 2022 it went to 0.76. The ratio for correctly classified classes in 2008 is represented as (**XGBoost : SVM)** therefore the results shows 212:212 which indicate that during 2008 SVM and XG boost performed the same way as they classified the same number of samples. During 2013 the ratio was 345:312 which shows that XGBoost performed better than SVM. The results of 2017 show 374:316 which shows that XGBoost performed better than SVM. Lastly, in 2022 the ratio was 298:277 which shows that XGBoost performed better than SVM. Overall z scores results shows that XGBoost performs better than SVM.

4.5 DISCUSSION

Remote sensing has the potential to deliver useful data on the detection and evaluation of LULC (Glanville et al., 2015). Multispectral image classification has extensively involved the courtesy of the remote-sensing society because classification grades are the origin for various ecological and socio-economic ideas (Kulkarni et al., 2016). Researchers and experts have made a lot of progress in developing new classification methods and systems (Kulkarni et al., 2016) that improve the accuracy of classification. Recently, the usage of diverse categories of spatial and spectral resolutions of visual sensors has attained diverse grades of accomplishment in land-use/ land cover mapping. Nevertheless, worries over accuracy levels continue (Adam ell al., 2014).

In this context, this research sought to make use of the Landsat 5TM and Landsat 8 OLI in mapping and detecting LULC using high-resolution images; map the changes in LULC over time using an advanced classification algorithm; and compare gradient boosting with SVM in mapping LULC. Classification outcomes in this research substantiated that Landsat 5TM and Landsat 8 OLI images with the use of gradient boosting and SVM are more appreciated in mapping and understanding changes in land-use/ land cover types. High classification accuracy was attained. The diverse technique, in comparison to a classifier confusion matrix method, was employed to improve the accuracy of the subsequent mapping of land-use/ land-cover types. The valuation was carried out to compute the enhancement in accuracy mapping between a gradient boost and SVM. Table 6-7 displays the user, producer accuracy and overall accuracy performance as well as the F-scores to test the performance of the classifiers of the Gradient Boosting and SVM algorithms for the various land use land cover types.

From the results obtained both in gradient boosting and XGBoost performed better than SVM both in overall accuracies and Kappa statistics as well as F-scores and the ratio of Z-score. This observation is in line with (Mahdavi et al., 2019), who noted that the extraction of various features from satellite data is one technique to increase the accuracy of image classification. This observation is compatible with their findings. The recent findings help shed light on the power of the XGBoost classifier when applied to the problem of dealing with high-dimensional feature sets. (Ruiz Hernandez & Shi, 2018)

F-scores were used as a supporting accuracy metric for the classification classes identification, and they were obtained from the confusion matrix values from user's accuracy (UA) and producer's accuracy (PA), and the z-scores were obtained using McNemar's statistical tests. The estimated F-scores served as a measure of how well each class was able to distinguish between the other classes. According to (Zurqani., Post., Mikhailova., & Allen, 2019), the F1 score is a measure of a classifier's overall performance by taking a weighted average of the producer and user accuracies. All values were not statistically significant as all values were below the z score of 1.96. The association was not shown by all classifiers.

GEE's integrative capacity and its excellent script writing with parallel processing analysis of a stack of input features, explains much of the improved accuracy in the current study. This allows for the possibility of combining different feature sets for improved mapping accuracy (Tassi &

Vizzari, 2020). GEE was able to capture the morphological features of LULC better than studies using traditional tools,

The study found that during the years 2008–2013, in gradient boosting, discussion was focused on the best algorithm, which performed better than SVM. Therefore, waterbody has remained the same while the business class has increased. Build-up increases and bare soil reduced. Finally, agriculture increased. This shows that during 2008–2013 Polokwane was dominated by agricultural and business land. From 2013–2017, business land has decreased while build-up land has increased, due to the increase in population. Bare soil was also decreasing, which shows that indeed there was an increase in population because people were filling up the vacant land for build-up. Agriculture land continues to increase. In 2022, bare soil decreased due to fragmentation and more build-up occupied on agriculture land. Business land decreased with the increase in built-up areas. However, comparing the classification performances, XGBoost classification within the GEE platform, performed in the current study, yielded significantly higher accuracies and higher classified classes than SVM classification implemented in the eCognition software (Kuffer, Pfeffer, Sliuzas, & Baud., 2016).

In this research, the benefits of utilizing geospatial tools were investigated for the purpose of analysing and evaluating the changes in LULC (2008–2022) that occurred at four different land use land cover classes. The findings were utilized to make an evaluation of the nature and pattern of landscape fragmentation in the area that was under investigation. The processes of rectification and classification were carried out on the data that had been extracted with the assistance of the coding in GEE. The growth and pattern of land use development in Polokwane between the years 2008 and 2022 indicated that there has been a continuous increase in the development of the built-up area, which was a by-product of a reduction in the amount of agriculture and bareland classes. The results of this growth and development was observed at this points to the predominance of urban growth in the area, as was observed in several earlier studies that noted the existence of an inverse link between the built-up and other LULC such as agriculture (Irwin & Bockstael, 2007; Olayiwola & Fakayode, 2019; Bertram et al., 2021). Fragmentation of the landscape is the result of the uncontrolled urbanization process that occurred in the area under consideration. Metrics analysis of the landscape structure in Polokwane indicated high fragmentation and less infill development in the study area. This was shown by increases in the PD and GPA, MPA, LP and

Landcover of the built-up class, both of which indicated that there were more patches that were the largest in size.

Therefore, the expansion of the built-up area is explained as one of outward expansion. Changes in GPA (greatest patch area) for bareland suggest that the land use class was more fragmented between 2008 and 2013 than in 2017 and 2022. This is like some previous studies that established outward urban growth as a major consequence of high fragmentation of landscape within an urban environment (Simons, 1988; Singh et al., 2017). An increase in the patch density of the built-up area can be attributed to an increase in the demand for more houses, urban infrastructure, and services. Consequently, new built-up areas were formed at the fringes that segmented the existing bare land, vegetation, and cultivated lands in the latter epoch. However, there were reductions in the PD of some land use classes to indicate that some patches are merging to form a homogeneous landscape. For instance, bare land, agricultural land, and business land were reduced in 2017 and 2022. This confirms the study done by Simons, 1988; Dewan et al., 2012; Singh et al., 2014; Akın & Erdogan, 2020; Ai et al., 2022 who alluded that these reductions are an indication of scattered cultivation in the area.

From the results obtained, it shows that gradient boosting performs better than SVM, therefore with will assist Polokwane municipality as using gradient boosting machine learning in LULC analysis is the best with its ability to handle complex and high-dimensional data sets (Lwin & Neelamani, 2012). LULC data often includes multiple variables such as vegetation cover, built-up areas, water bodies, and agricultural land. Traditional statistical methods may struggle to adequately capture the relationships between these variables. Zhang et al., (2019) highlighted that gradient boosting machine learning algorithms excel at identifying non-linear patterns and interactions in complex data, thus providing more accurate and precise LULC predictions. gradient boosting machine learning will assists Polokwane municipality with its ability to handle spatial data. LULC analysis involves understanding the spatial distribution of different land cover types. Gradient boosting algorithms can incorporate spatial attributes into their models, thus improving the precision of LULC predictions (Zhang et al., 2019). This is particularly useful in Polokwane Municipality, where there may be specific spatial patterns and dependencies in the land cover types.

According to Singh and Kaur (2020), gradient boosting machine learning is of great importance in monitoring and evaluating LULC in Polokwane Municipality. Its ability to handle complex and high-dimensional data, handle spatial attributes, and address imbalanced data sets makes it a valuable tool for accurate and precise land cover predictions. This information is critical for effective urban planning and environmental management in the municipality.

Furthermore, since gradient boosting machine learning can handle imbalanced data sets, which is a common challenge in LULC analysis, this is crucial in the urban growth monitoring of Polokwane municipality. Imbalanced data occurs when there is a significant disparity in the number of samples for different land cover types, therefore traditional statistical approaches may struggle to accurately represent minority classes, resulting in biased predictions. Gradient boosting algorithms have mechanisms to address imbalanced data, ensuring fair representation of all land cover types in the analysis.

CHAPTER 5: CONCLUSION AND RECOMMENDATION

5.1 CONCLUSION

The use of cloud computing by GEE was successfully used in the process of mapping Polokwane human settlements. GEE demonstrated a high degree of flexibility and adaptation because of its integrative skills and its effective platform for script writing. This work created and evaluated a pixel-based classification of a variety of input combinations within the GEE environment. By repeatedly putting the SVM ensemble classifier through a series of tests and XGBoost with a wide variety of feature inputs, we could focus on the variables whose results were the most favourable. The XGBoost model performed better than SVM in classifying LULC, producing high average accuracy than SVM. After performing the calculations, it was determined that the XGBoost classifier had a larger percentage of accurate predictions than the SVM classifier. The findings show how the GEE framework is changing the paradigm in built-up area mapping from a static, product-based approach to a more dynamic, application-specific one with reasonable accuracy and in no time at all. This shift is made possible by simplifying access to and processing of a large amount of satellite data.

Through the LULC changes, human activities have altered the climate all over the world. Since the pre-industrial era, the LULC changes, such as deforestation and urbanization, have been a significant contributor to climate change (Solomon et al., 2007). So, Polokwane municipality make great efforts to prevent unplanned urban growth on fertile land and have therefore initiated numerous projects to evaluate the spatiotemporal changes. This study used remote sensing, machine learning, and GIS technologies to monitor the spatiotemporal changes in land use and land cover (LULC) in Polokwane Municipality. The study also proposes a new method for identifying changes in LULC using GEE for image processing, Landscape matrix analysis and comparing the performance of multiple classifiers. The results show that the gradient boosting classification method achieved the highest accuracy in LULC classification as compared to SVM. Human activities have significantly altered the climate through LULC changes such as desertification and urbanization, and that these changes have had a significant impact on the agricultural land in the area. The LULCC analysis and landscape matrix analysis information have revealed a typical spatial dynamic for all Polokwane city. Over the past 15 years, urban or built-up areas have grown, followed by agricultural lands, which have shrunk over a specific period due to encroachments. Due to land reclamation and the construction of new areas there, the agriculture area has shrunk.

gradient boosting is a powerful machine learning technique that can effectively handle complex and non-linear relationships between variables. It has been widely used in various environmental and remote sensing applications, including LULC classification and urban growth modelling (Colombo et al., 2019; Müller et al., 2017). By utilizing gradient boosting, it is possible to accurately evaluate and monitor changes in LULC and urban growth patterns in the Polokwane Municipality. Polokwane Municipality is experiencing rapid urbanization and expansion, resulting in significant changes in land use. Monitoring and understanding the dynamics of LULC and urban growth is essential for effective urban planning, resource management, environmental conservation, and socio-economic development (Kleindorfer et al., 2010; Koomen et al., 2008). Gradient boosting can provide accurate and timely information on the patterns, drivers, and impacts of urban growth, thereby facilitating evidence-based decision-making and policy formulation in the municipality.

Furthermore, gradient boosting can handle large datasets with multiple input variables, such as satellite imagery, topographic data, and socio-economic indicators. By integrating these diverse datasets, it is possible to capture the multidimensional nature of LULC and urban growth processes in the Polokwane Municipality. This enables the identification of key drivers and patterns of change, as well as the development of reliable predictive models for future scenarios (Hansen et al., 2005; Seto et al., 2012).

5.2 RECOMMENDATION

The approach offers several advantages over previous methods, including faster and more efficient processing of large datasets, identification of the most accurate classifier for LULC change detection, and the use of GEE as well as landscape matrix analysis for LULC change detection in Polokwane municipality, Limpopo.

Nonetheless, remote sensing can monitor and assess the change in LULC. There has been insufficient research into the development and testing of methods for mapping and monitoring LULC during the past fifteen years (Glanville et al., 2015). There have been a significant number of scholarly studies published about LULC and the monitoring of it. More research needs to be done so that larger numbers of high-quality test and training samples can be accumulated in more datasets. This can be accomplished by conducting more research. This would make it possible to conduct an examination of the applications of gradient tree boost and SVM methods in the interrelated fields of land use mapping and land cover monitoring. In contrast to standard algorithms, Weih Jr. et al. (2010) offers object-based classification as the preferable strategy. As compared to both supervised and unsupervised pixel-based algorithms, its performance is better in every scenario. The findings of this study provide credibility to the findings of other research that has been evaluated in the same environment over the past few years.

In view of the preceding, the study advised inter-temporal monitoring and checking on the land use development and territorial extension of the built-up region. To accomplish this goal, stakeholders in land use development and management, such as town planners and local government authorities, should be thoroughly sensitized on the immense benefits of utilizing remote sensing and geographic information systems in decision-making and policy formulation for the study of land use and cover changes. This should be done so that the goal can be achieved. In addition, governments (at all different levels) need to determine the urban infrastructure and services that are exclusive to regions. In this situation, it is vital for the government to support public-private partnerships in the development of sustainable infrastructural amenities in the most disadvantaged communities to prevent unevenness in both the distribution of these facilities and their location. The expansion of organic and unplanned developments in the study area will also be slowed down because of this action.

An alternative to the practice of monitoring the growth of settlements is provided by the fact that the proposed method, image classification, was successful in identifying places where human habitation had changed. The detection of building structures enables detailed monitoring of the urban built environment, which can help to the monitoring of Sustainable Development Goal indicator 11 and other associated goals. This can also contribute towards the aim of sustainable urbanization.

5.3 LIMITATION OF THE STUDY

This study is only about mapping and evaluating the LULC in Polokwane Municipality for 2008, 2013, 2017, and 2022. It does not cover all South Africa. The study focuses only on Polokwane Municipality, which may not be representative of other urban areas. Therefore, the findings and conclusions of the study may not be generalizable to other regions. This study does not focus on LULC prediction since it would require collecting data over a period of 15 years and monitoring the change each year, both of which are labour- and resource-intensive activities that are extremely expensive. Another component that was not able to be evaluated in depth in this study was the spectral-based analysis of LULC. If it were carried out in its entirety, there was a possibility that the accuracy of the training samples would have improved.

The study was limited by the availability of historical data on urban land use and land cover change. Without long-term data, it may be difficult to accurately analyse the trends and patterns of change over time. The accuracy of machine learning classification techniques depends heavily on the quality and representativeness of the training data, therefore the training data was limited because of the size of the study area.

6.1 REFERENCE

abuelgasim, A. A., Ross, W. D., Gopal, S., & & Woodcock, C. E. (1999). Change detection using adaptive fuzzy neural networks. *environmental damage assessment after the Gulf War. Remote Sensing of Environment,*, 70, 208–223.

Abutaleb, K., & al., e. (2021). (2021) ‘Estimating urban greenness index using remote sensing data: A case study of an affluent vs poor suburbs in the city of Johannesburg'. *The Egyptian Journal of Remote Sensing and Space Science,* , 24(3), pp. 343–351.

Adams, J. B., Sabol, D., Kapos, V., Filho, R. A., Roberts, D. A., Smith, M. O., & & Gillespie, A. R. (1995). Classification of multispectral images based on fractions of endmembers. *application to land-cover change in the Brazilian Amazoˆn. Remote Sensing*, 52, 137–154.

Agouris, P., Stefanidis, A., & And Gyftakis, S. (2001). Differential snakes for change detection in road segments. *Photogrammetric Engineering and Remote Sensing*, 67, 1391–1399.

Allen, T. R., & & Kupfer, J. A. (2000). Application of spherical statistics to change vector analysis of Landsat data. *southern Appalachian spruce-fir forests. Remote Sensing of Environment,* , 74, 482–493.

Allum, J. A., & Dreisinger, R. (1987,). Remote sensing of vegetation change near Inco’s Sudbury mining complexes. *International Journal of Remote Sensing,*, 8, 399–416.

Alonzo, M., J. P. McFadden, D. J., Nowak, &. D., & Roberts. (2016). “Mapping urban forest structure and function using hyperspectral imagery and lidar data". *Urban Forestry and Urban Greening*, 17:135-147.

Alves, D. S. (2002). Space–time dynamics of deforestation in Brazilian Amazoˆ nia. *International Journal of Remote Sensing,*, 23, 2903–2908.

Alves, D. S., Pereira, J. L., De Sousa, C. L., Soares, J. V., & Yamaguchi, F. (1999). Characterizing landscape changes in central Rondonia using Landsat TM imagery. *International Journal of Remote Sensing,*, 20, 2877–2882.

Alwashe, M. A., & & Bokhari, A. Y. (1993,). Monitoring vegetation changes in Al Madinah, Saudi Arabia, using Thematic Mapper data. *International Journal of Remote Sensing,*, 14, 191–197.

Amani, M., Mahdavi, S., Afshar, M., Brisco, B., Huang, W., Mohammad Javad Mirzadeh, S., . . . Hopkinson, C. (2019). CanadianWetland Inventory using Google Earth Engine:The First Map and Preliminary Results. *Remote Sens.*, 11, 842.

Anderson JR, Hardy EE, & Roach JT & Witmer RE. (1976). A land use and land cover classification system for use with remote sensor dat. *US Geological Survey Professional Paper* , 964:28.

Angelici, G., Brynt, N., & & Friendman, S. (1977). Techniques for land use change detection using Landsat imagery. *(Bethesda, MD: American Society of Photogrammetry)* (pp. 217-228). USA: the, Proceedings of the 43rd Annual Meeting of American Society of Photogrammetry and Joint Symposium on Land Data Systems, Falls Church, VA.

Anselin, & L. (1995). “Local Indicators of Spatial association—LISA. *Geographical Analysis , 8*(1), 27 (2): 93–115.

Armour, B., Tanaka, A., Ohkura, H., & And Saito, G. (1998). Environmental Monitoring Methods and Applications, edited by R. S. Lunetta and C. D. Elvidge (Chelsea, MI: Ann Arbor Press). *Radar interferometry for environmental change detection. In Remote Sensing Change Detection*, 245–279.

Aronson, M. F., & et.al. (2017). “Biodiversity in the City: Key Challenges for Urban Green Space Management. *” Frontiers in Ecology and the Environment 15 (no. 4)*, 15 (no. 4).

Ashcroft, M. B., & et.al. (2014). “Creating Vegetation Density Profiles for a Diverse Range of Ecological Habitats Using Terrestrial Laser Scanning". *” Methods in Ecology and Evolution'* , 5 (3): 263–272.

Asner, G. P., Keller, M., Pereira, R., & And Zweede, J. C. (2002,). Remote sensing of selective logging in Amazoˆnia—assessing limitations based on detailed field observations, Landsat ETMz, and textural analysis. . *Remote Sensing of Environment,*, 80, 483–496.

Baines, O., & P. Wilkes, &. M. (2020.). “Quantifying Urban Forest Structure with Open-access Remote Sensing Data Sets". *Urban Forestry and Urban Greening* , 50: 126653. .

Berland, A., K. Schwarz, D. L., & Hopton., &. M. (2015). “How Environmental Justice Patterns are Shaped by Place:. *Terrain and Tree Canopy in Cincinnati, Ohio, USA.” Cities and the Environment (CATE):Cities and the Environment (CATE),* 8 (1): 1. .

Bierwagen, B. G. (2007.). “Connectivity in Urbanizing Landscapes: . *The Importance of Habitat Configuration, Urban Area Size, and Dispersal.” Urban Ecosystems* , 10 (1): 29–42. .

Bittencourt, H. R., & & Clarke, R. T. (2003). Use of classification and regression trees (CART) to classify remotely-sensed digital images. *In Geoscience and Remote Sensing Symposium, 2003. IGARSS’03. Proceedings. 2003 IEEE International*, Vol. 6, pp. 3751–3753).

Boentje, J. P., & Blinnikov., &. M. (2007.). A Remote Sensing Perspective.”. *“Post-Soviet Forest Fragmentation and Loss in the Green Belt around Moscow, Russia : Landscape and Urban Planning (1991–2001):*, 82 (4): 208–221.

Bolund, P., & Hunhammar, &. S. (1999). “Ecosystem Services in Urban Areas. *”Ecological Economics*, 29 (2): 293-301.

Boyle, S. A., C. M. Kennedy, J. Torres, K., Colman, P. E., & Pérez-Estigarribia, &. U. (2014). “High-resolution Satellite Imagery Is an Important yet Underutilized Resource in Conservation Biology. . *” PloS One*, 9 (1): e86908. .

Breiman, L., Friedman, J. H., Olshen, R. A., & & Stone, C. J. (n.d.). Classification And Regression Trees. Boca Raton: Chapman & Hall/CRC.

Corbane, C., & et.al. ((2019)). ‘Automated global delineation of human settlements from 40 years of Landsat satellite data archives’,. *Big Earth Data*,, 3(2),pp. 140–169.

de la Barrera, F., S. Reyes-Paecke, & Banzhaf, &. E. (2016). “Indicators for Green Spaces in Contrasting Urban Settings.” . *Ecological Indicators* , 62: 212–219.

Development Core, R., & Team., &. (2013). “A Language and Environment for Statistical Computing.” . *Vienna, Austria: R Foundation for Statistical Computing* , 55: 275–286. .

Dikananda, A., Jumini, S., Tarihoran, N., Christinawati, S., & Trimastuti, W. &. (2022.). Comparison of Decision Tree Classification Methods and Gradient Boosted Trees. *R.EM Journal.*, (February, 28):316–322. .

Dobbs, C., & C. R. Nitschke, &. D. (2014. "). Global Drivers and Tradeoffs of Three Urban Vegetation Ecosystem Services." . *PLoS One* , 9 (11): e113000.

Dobbs, C., & D. Kendal, &. C. (2014.). "Multiple Ecosystem Services and Disservices of the Urban Forest Establishing Their Connections with Landscape Structure and Sociodemographic.". *Ecological Indicators* , 43: 44–55.

Dobbs, C., Á. Hernández-Moreno, S., Reyes-Paecke, & Miranda., &. M. (2018.). "Exploring Temporal Dynamics of Urban Ecosystem Services in Latin America: . *The Case of Bogota (Colombia) and Santiago (Chile)." Ecological Indicators* , 85: 1068–1080. .

Dobbs, C., C. Nitschke, & Kendal., &. D. (. 2017.). "Assessing the Drivers Shaping Global Patterns of Urban Vegetation Landscape Structure.". *Science of the Total Environment*, 592: 171–177.

Dolean., Bilasco., & at al. (2020). Evaluation of the Built-Up Area Dynamics in the First Ring of Cluj-Napoca Metropolitan Area, Romania by Semi-Automatic GIS Analysis of Landsat Satellite Images. *Appl.Sci.*, 10, 7722.

Dong, J., Xiao, X., Menarguez, M., Zhang, G., Qin, Y., Thau, D., . . . Moore, B. I. (2016). Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sens. Environ.*, 185, 142–154.

Foody, G. M., Mathur, A., Sanchez-Hernandez, C., & & Boyd, D. S. (2006). Training set size requirements for the classification of a specific class. . *Remote Sensing of Environment*,, 104(1), 1–14.

Foody, G., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(6), 1335-1343.

Friedl, M. .., McIver, D. .., Hodges, J. C., Zhang, X. .., Muchoney, D., Strahler, A. .., & … Schaaf, C. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment,*, 83(1–2), 287–302.

Gorelick, N. M. (2017). Google Earth Engine: Planetary-scale Geospatial Analysis for Everyone. *Remote Sensing of Environment*, 202: 18–27.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.*, 202, 18–27.

Kelley, L., Pitcher, L., & Bacon, C. (2018). Bacon, C. Using Google Earth Engine to Map Complex Shade-Grown Coffee Landscapes in Northern Nicaragua. *Remote Sens.*, 10, 952.

Kowe, P., & al., e. ((2020) ‘). A quantitative framework for analysing long term spatial clustering and vegetation fragmentation in an urban landscape using multi-temporal landsat data. *, International Journal of Applied Earth Observation and Geoinformation,*, 88, p. 102057. .

Kuffer, M., Pfeffer, K., Sliuzas, R., & Baud. (2016). Extraction of Slum Areas From VHR Imagery Using GLCM Variance. IEEE. *Sel. Top.Appl. Earth Obs. Remote Sens.*, 9, 1830–1840.

Lawrence, R. L., & & Wrlght, A. (2001). Rule-Based Classification Systems Using Classification and Regression Tree (CART) Analysis. *Photogrammetric Engineering and Remote Sensing,*, 67(10), 1137–1142.

Liu, Q., Huang, C., & Li, H. (2020). Quality Assessment by Region and Land Cover of Sharpening Approaches Applied to GF-2 Imagery. *Appl. Sci.*, 10, 3673.

Liu, X., Hu, G., Chen, Y., Li, X., Xu, X., Li, S., . . . Wang, S. (2018). High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. *Remote Sens. Environ.*, 209, 227–239.

Mahdavi, S., Afshar, M., Brisco, B., Huang, W., Mohammad Javad Mirzadeh, S., White, L., . . . Montgomery, J. (2019). CanadianWetland Inventory using Google Earth Engine: The First Map and Preliminary Results. *Remote Sens.*, 11, 842.

Mananze, S., Pôças, I., & Cunha, M. (2020). Mapping and Assessing the Dynamics of Shifting Agricultural Landscapes Using Google Earth Engine Cloud Computing, a Case Study in Mozambique. *Remote Sens*, 12, 1279.

Mishra, V., & et.al. ((2019)). 'Assessment of Spatio-Temporal Changes in Land Use/Land Cover Over a Decade (2000–2014) Using Earth Observation Datasets: A Case Study of Varanasi District, India'. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 43(S1), pp. 383–401.

Mountrakis, G., Im, J., & & Ogole, C. (2011). Support vector machines in remote sensing. *A review. ISPRS Journal of Photogrammetry and Remote Sensing,*, 66(3), 247–259.

Mudau, N., & al., e. ((2020)). 'Towards development of a national human settlement layer using high resolution imagery. *: a contribution to SDG reporting', South African Journal of Geomatics,* , 9(1), p. 12.

Nery, T., Sadler, R., Solis-Aulestia, M., White, B., Polyakov, M., & & Chalak, M. (2016). Comparing supervised algorithms in Land Use and Land Cover classification of a Landsat time-series. *In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* , (pp. 5165–5168).

Patel, N., Angiuli, E., Gamba, P., Gaughan, A., Lisini, G., Stevens, F., . . . Trianni, G. (2015). Multitemporal settlement and population mapping from Landsat using Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf. ITC J.*, 35, 199–208.

Pesaresi, M., & et.al. ((2013)). 'A Global Human Settlement Layer From Optical HR/VHR RS Data:. *Concept and First Results', IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,*, 6(5), pp. 2102–2131.

Pesaresi, M., & et.al. (. (2013)). 'A Global Human Settlement Layer From Optical HR/VHR RS Data:. *Concept and First Results', IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,*, 6(5), pp. 2102–2131. .

Ridwan, M., & et.al. (2018). ‘Applications of Landsat-8 Data: a Survey’,. *International Journal of Engineering & Technology*, 7(4.35), p. 436.

Ruiz Hernandez, I., & Shi, W. A. (2018). Random Forests classification method for urban land-use mapping integrating spatial metrics and texture analysis. *Int. J. Remote Sens.*, 1175–1198.

Shafizadeh-Moghadam, H., Khazaei, M., Alavipanah, S., & Weng, Q. (2021). Google Earth Engine for large-scale land use and land cover mapping: An object-based classification approach using spectral, textural and topographical factors. GISci. *Remote Sens*, 58, 914–928.

Tassi A., V. M. (2020). Object-Oriented LULC Classification in Google Earth Engine Combining SNIC, GLCM, and Machine Learning Algorithms. *Remote Sens*, 12.3776.

Teluguntla, P., Thenkabail, P., Oliphant, A., Xiong, J., Gumma, M., Congalton, R., . . . Huete, A. (2018). A 30-m landsat derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *Photogramm. Remote Sens. ISPRS J.+*, 144, 325–340.

Tingzon, I., Dejito, N., Flores, R., De Guzman, R., Carvajal, L., Erazo, K., . . . Ghani, R. (21–25 September 2020.). Mapping New Informal Settlements Using Machine Learning and Time Series Satellite Images: An Application in the Venezuelan. Geneva,Switzerland,.

Tso, B., & & Mather, P. M. (2009). Classification Methods for Remotely Sensed Data (Second). *CRC Press.* Boca Raton:.

Wekesa, B., Steyn, G., & Otieno, F. (2011). A review of physical and socio-economic characteristics and intervention approaches of informal settlements. *Habitat Int.*, 35, 238–245.

Wiweka, Suwarsono, Jalu T Nugroho, & Samsul. (2014). "Performance test parameters of remote sensing for identitifocation burned area using Landsat-8". *2014 International Conference on ICT for Smart Society (ICISS).*

Xiong, J. P. (2017). “Automated Cropland Mapping of Continental Africa Using Google Earth Engine Cloud Computing. ” *ISPRS Journal of Photogrammetry and Remote Sensing*, 126: 225–244.

Xiong, J., Thenkabail, P., Tilton, J., Gumma, M., Teluguntla, P., Oliphant, A., . . . Gorelick, N. (2017). Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on Google Earth Engine. *Remote Sens.*, 9, 1065.

Ye, Z., & et.al. ((2020) ,). ‘Mapping and Discriminating Rural Settlements Using Gaofen-2 Images and a Fully Convolutional Network’. *Sensors,*, 20(21), p. 6062.

Zhang, K., Dong, X., Liu, Z., Gao, W., Hu, Z., & Wu, G. (2019). Mapping Tidal Flats with Landsat 8 Images and Google Earth Engine: A Case Study of the China’s Eastern Coastal Zone circa 2015. *Remote Sens.*, 11, 924.

Zurqani., Post., Mikhailova., & Allen. (2019). Mapping Urbanization Trends in a Forested Landscape Using Google Earth. *Remote Sens.Earth.Sci*, 2, 173-182.

Twisa, S. & Buchroithner, M.F. 2019. Land-Use and Land-Cover (LULC) Change Detection in Wami River Basin, Tanzania. *Land*. 8(9):136.

Wang, L., Marzahn, P., Bernier, M. & Ludwig, R. 2017. Mapping permafrost landscape features using object-based image classification of multi-temporal SAR images. *ISPRS Journal of Photogrammetry and Remote Sensing*. 141:10–29.

Wekesa, B., Steyn, G., & Otieno, F. (2011). A review of physical and socio-economic characteristics and intervention approaches of informal settlements. *Habitat Int.*, 35, 238–245.

Wiweka, Suwarsono, Jalu T Nugroho, & Samsul. (2014). "Performance test parameters of remote sensing for identitifocation burned area using Landsat-8". *2014 International Conference on ICT for Smart Society (ICISS).*

Xiong, J. P. (2017). “Automated Cropland Mapping of Continental Africa Using Google Earth Engine Cloud Computing. ” *ISPRS Journal of Photogrammetry and Remote Sensing*, 126: 225–244.

Xiong, J., Thenkabail, P., Tilton, J., Gumma, M., Teluguntla, P., Oliphant, A., . . . Gorelick, N. (2017). Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on Google Earth Engine. *Remote Sens.*, 9, 1065.

Yang, D., Zhang, P., Jiang, L., Zhang, Y., Liu, Z. & Rong, T. 2022. Spatial change and scale dependence of built-up land expansion and landscape pattern evolution—Case study of affected area of the lower Yellow River. *Ecological Indicators*. 141:109123.

Ye, Z., & et.al. ((2020) ,). ‘Mapping and Discriminating Rural Settlements Using Gaofen-2 Images and a Fully Convolutional Network’. *Sensors,*, 20(21), p. 6062.

Zhang, Q., Wang, J., Peng, X., Gong, P. & Shi, P. 2002. Urban built-up land change detection with road density and spectral information from multi-temporal Landsat TM data. *International Journal of Remote Sensing* 23 (15): 3057-3078

Zhang, K., Dong, X., Liu, Z., Gao, W., Hu, Z., & Wu, G. (2019). Mapping Tidal Flats with Landsat 8 Images and Google Earth Engine: A Case Study of the China’s Eastern Coastal Zone circa 2015. *Remote Sens.*, 11, 924.

Zhou, Q., Wang, S. & Liu, Y. 2022. Exploring the accuracy and completeness patterns of global land-cover/land-use data in OpenStreetMap. *Applied Geography*. 145:102742.

Zurqani, H.A., Post, C.J., Mikhailova, E.A. and Allen, J.S. (2019). Mapping Urbanization Trends in a Forested Landscape Using Google Earth Engine. Remote Sensing Earth Systems Science, 2, 173–182.

Zurqani., Post., Mikhailova., & Allen. (2019). Mapping Urbanization Trends in a Forested Landscape Using Google Earth. *Remote Sens.Earth.Sci*, 2, 173-182.