

Regine_Essay_afterCorrection.pdf

by Regine Ines Piyou Kamadeu

Submission date: 09-Jun-2024 04:34PM (UTC+0300)

Submission ID: 2398686578

File name: Regine_Essay_afterCorrection.pdf (3.8M)

Word count: 16213

Character count: 88282



AIMS

African Institute for
Mathematical Sciences
CAMEROON

Assessing the Impact of Urban Growth Dynamics on Quality and Access to Service Delivery in Uganda using Supervised Machine Learning

Regine Ines Piyou Kamadeu (regine.piyou@aims-cameroon.org)
African Institute for Mathematical Sciences (AIMS)
Cameroon

Supervised by: Prof. Nara Monkam
Pretoria University, South Africa (Cape Town)

18 May 2024

Submitted in Partial Fulfillment of a Co-op Master's Degree at AIMS-Cameroon

Abstract

This study investigates the relationship between urban growth and resident satisfaction with service delivery in Uganda. We employed supervised machine learning techniques and Landsat 7 satellite data to map and quantify urban expansion between 2011 and 2015. Our analysis identified significant urban growth across the country, with some districts experiencing rapid expansion while others remained stagnant. Using this district-level urbanization data, we analyzed resident satisfaction through survey data. Our findings reveal that, while urban areas are growing rapidly, the quality and access to basic government services like education, healthcare, and electricity haven't synchronized. This highlights a critical need for improved service delivery strategies to meet the demands of Uganda's growing urban population. The study emphasizes the importance of incorporating both urbanization data and resident feedback when designing and implementing service delivery plans. By understanding the specific needs of each district, policymakers can allocate resources more effectively and ensure sustainable development in Uganda's urbanizing landscape.

Keys words: Urbanization, Landsat 7 satellite data, supervised machine learning, survey data.

Declaration

I, the undersigned, hereby declare that the work contained in this essay is my original work and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

A handwritten signature in black ink, appearing to read "Regine Ines Piyou Kamadeu".

Regine Ines Piyou Kamadeu, 18 May 2024.

Contents

Abstract	i
List of Figures	v
1 General Introduction	1
1.1 Context of the Study	1
1.2 Problem Statement	2
1.3 Research Objectives	2
1.4 Thesis Organisation	3
2 Literature Review	4
2.1 Definition of keys words	4
2.2 Change Detection (CD)	5
2.3 Land Cover Classification (LCC)	8
2.4 Urbanization and service delivery	9
2.5 Our Contributions	10
3 Methodology	11
3.1 Research Design	11
3.2 Classification and Regression Tree (CART)	12
3.3 Random Forest	14
3.4 Support Vector Machine (SVM)	15
3.5 Performance Evaluation Metric	19
4 Implementation and Results	20
4.1 Data Description	20
4.2 Preprocessing of Afrobarometer data	22
4.3 Google Earth Engine (GEE)	23
4.4 Land Cover Classification	24
4.5 Change Detection	29
4.6 District-Level Urban Change and Satisfaction with Government Service Delivery	30

5 Discussions	38
5.1 Our findings	38
5.2 Recommandations to Local Government	38
5.3 Challenges	39
6 Conclusion	41
A Some additional data	42
Acknowledgements	43
References	46

List of Figures

2.1	Graphical illustration of the change detection problem [19].	5
2.2	Traditional Change Detection[30].	6
2.3	General workflow of change detection [30].	7
2.4	LULCC general framework diagram [39].	9
3.1	Flow Chart of our Methodology Research	11
3.2	Illustration of the random Forest classification structure [13].	14
3.3	Maximum margin-minimum norm classifier in support vector machine with optimal hyperplane for linearly non-separable classes[27].	16
4.1	Composite RGB Landsat 7 2011	25
4.2	Composite RGB Landsat 7 2015	25
4.3	CART 2011	26
4.4	RF 2011	26
4.5	SVM 2011	26
4.6	CART 2015	26
4.7	RF 2015	26
4.8	SVM 2015	26
4.9	Confusion Matrix classifier 2011	27
4.10	Confusion Matrix classifier 2015	28
4.11	All change between 2011-2015	29
4.12	Urban Change between 2011-2015	29
4.13	Dist Urban growth with CART	31
4.14	Dist Urban growth with MODIS	31
4.15	Access to Basic Needs in Mbarara, 2011 and 2015	33
4.16	Trends in Public Opinion in Mbarara on Local Government Handling of Basic Needs, 2011-2015	33
4.17	Access to Basic Needs in Kabarole, 2011 and 2015	34
4.18	Trends in Public Opinion in Mbarara on Local Government Handling of Basic Needs, 2011-2015	34
4.19	Access to Basic Needs in Mubende, 2011 and 2015	35

4.20 Trends in Public Opinion in Mubende on Local Government Handling of Basic Needs, 2011-2015	35
4.21 Comparative analysis: Urbanization Vs Satisfaction to Access to Service Delivery	36
4.22 Comparative analysis: Urbanization Vs Satisfaction of the Improvement of the Quality of the Service Delivery	37

1. General Introduction

Africa is rapidly urbanizing, with its urbanization rate climbing from 15% in 1960 to 40.43% in 2015. Projections from the World Urbanization Prospects (2014) indicate that this trend will persist, with urbanization in Africa expected to reach 49.3% by 2035 and 60% by 2050 [26]. Africa boasts the highest urban growth rate globally, at 3.4%, surpassing Asia's 2.10% and Latin America and the Caribbean's 1.28%. Africa's urban population exploded, doubling between 1995 and 2015, may double again in 2035, and is predicted to triple by 2050. Africa and Asia together are studied to account for almost 90% of the world's urban population growth. If this upward trend continues, African cities will host almost a quarter of the world's urban population, totaling around 1.2 billion people [34].

This substantial urbanization trend is consistent with Tiebout's 1956 "feet-voting theory", which holds that urbanization is frequently caused by an inadequate supply of public goods and services in areas[26]. Furthermore, as more people in rural areas move to cities to join a growing middle class, it makes cities seem like the place to be for success and a better life. This leads to more money being invested in cities, with better roads, buildings, and fewer people out of work.

In the middle of this continent-wide urban phenomenon, we zoom in on specific places like Uganda. In Uganda, where nearly 20% of the population lives in cities, understanding the dynamics of urban expansion and its implications for service delivery is critical. By exploring Uganda's urban landscape, we can get insights into the difficulties of urbanization in the context of a growing African country.

1.1 Context of the Study

Uganda, nicknamed "The Pearl of Africa," is a country in East Africa. Its neighbors include Kenya to the east, Tanzania to the south, Rwanda and the Democratic Republic of Congo to the west, and Sudan to the north. Despite being landlocked, Uganda is renowned for its natural beauty and cultural diversity. With an estimated population of 35 million, Uganda has a notably young demographic, with a median age of 15 years [1].

Against this context, urbanization emerges as a major trend in Uganda, with an annual growth rate of 4.5%. According to projections, the urban population will nearly double between 2010 and 2030, with 26% of the population now living in cities [5, 21]. Rapid urbanization is changing the country's urban landscape and land use patterns.

Remarkably, Uganda ranks eleventh among African countries with a strong economic performance in 2024, boasting a Gross Domestic Product (GDP) growth rate of 6% [6]. This economic vitality, coupled with Uganda's youthful demographic, offers a distinct opportunity for sustainable development.

In Uganda, the challenges of urbanization and demographic changes make it necessary to coordinate activities with the seventeen Sustainable Development Goals (SDGs). These aspirations cover a wide range of themes, such as the elimination of poverty, the reduction of hunger, the provision of high-quality education, the advancement of health, the establishment of gender equality, and the development of sustainable urban areas and communities [4]. Uganda can make great progress toward equitable and sustainable development by using its youthful population and its economic momentum. Providing the above services in an efficient manner, especially in cities that are growing quickly, becomes essential to this study. Uganda may use the process of urbanization as a catalyst for positive transformation by attending to the diverse needs of its urban population.

1.2 Problem Statement

Uganda's growing urbanization creates an extensive number of challenges that limit long-term development in its cities. These challenges include uncoordinated planning and development which results in uncontrolled urban sprawl, a youth rise in urban areas, the proliferation of poor neighborhoods accompanied by deteriorating housing and inadequate sanitation, privatization of urban development without corresponding provision of public space and connectivity, administrative weaknesses, institutional deficiencies and inadequate coverage of basic infrastructure services, particularly in water, energy, and sanitation [3].

To address these challenges caused by urban growth and aid the achievement of sustainable development goals in Uganda's cities, our study looks at how this rapid growth is affecting the quality and availability of services in Uganda cities. To achieve this, the study proposes using satellite imagery of urban areas to compare land conditions across two different years(2011 and 2015), providing insights into the patterns and dynamics of urban growth. In addition, we will use data from Afrobarometer questionnaires to better understand citizens' opinions about the government's service delivery and propose areas for improvement. Finally, our study combines satellite imagery analysis with Afrobarometer data. This combined approach aims to provide insights into Uganda's current state of urban development and service provision, informing evidence-based strategies for policymakers and local governments to improve urban sustainability and quality of life for all residents.

1.3 Research Objectives

Uganda's national administrative system is hierarchical, with six tiers according to the last census 2024: cities (10), districts (146), counties (322), municipalities (39), town councils or sub-counties (1,488), parishes councils (7,553), and villages or wards councils (58,197). Since we lack sufficient satellite and survey data for the period 2011 and 2015 in the lower administrative tiers to assess effective urban growth, our approach involves examining the growth of Uganda's cities through district-level analysis.

Our study aims to:

1. Analyze Afrobarometer data to gain valuable insights into service delivery, with a focus on health-care access, education access, security concerns, water and sanitation infrastructure, energy availability, and transportation infrastructure (specifically roads).
2. Use machine learning, including Classification Regression Tree (CART), Random Forest (RF), and Support Vector Machine to conduct land cover classification and change detection using Landsat 7 satellite data.
3. Integrate the findings from both Afrobarometer data analysis and spatial analysis using supervised machine learning techniques to conduct a comparative analysis. This analysis will identify the link between community needs and urban growth patterns discovered through land cover classification and change detection. The goal is to identify areas with high needs and make recommendations to local government authorities for better urban planning and sustainable development activities in Uganda's cities.

1.4 Thesis Organisation

This chapter has provided a comprehensive overview of the research context, problem statement, and research objectives. The remaining structure of the work is as follows.

In Chapter 2, we will start by defining some key terms. Then we will go over recent research on applying machine learning techniques for land cover classification and change detection. We will talk about what previous studies have accomplished, the problems they faced, and their limits. Finally, we will present our contributions.

In Chapter 3, we will explore the methodology of our research. We will begin by outlining the research design, followed by a detailed explanation of the supervised machine learning methods employed in this study. This includes Classification and Regression Trees (CART), Random Forest (RF), and Support Vector Machines (SVM). Finally, we will conclude by presenting the metrics used to evaluate the performance of these models.

In Chapter 4, we will examine closely the implementation details and present the results. We will present our work done in Google Earth Engine, including the results for Land Cover Classification and the Change Detection (transition matrix). We will start by describing the satellite imagery and the Afrobarometer survey data used. Following that, we will analyze the survey data, assess urban growth at the district level, and conduct a comparative analysis between district-level urban growth and residents' access to and quality of basic services.

In Chapter 5, We will present our key findings, along with recommendations for local governments to address identified issues. We will also discuss the challenges encountered during the research process.

The last part of the work will be the conclusion, which will summarize our work, and suggest areas for future research.

2. Literature Review

In this chapter, we start by explaining important terms. Then we discuss "change detection" and its mathematical description, along with some work done with machine learning and deep learning in remote sensing. Furthermore, we will define the task of classifying land cover and review some studies that use machine learning/deep learning techniques and/or Geographic Information System (GIS) software for this purpose. In addition, we will present a review of related research on urbanization and service delivery in Africa. We wrap up by giving our contribution to this study.

2.1 Definition of keys words

For a better understanding of urban development in Uganda, it is essential to apprehend some basic concepts outlined here. Additionally, the analysis of image data introduces new terminology, which we will define here as well.

2.1.1 Definition (Urban growth). Clark (1982) defines urban growth as a process where towns and cities become more important [11]. This means more and more people are moving to these areas, making them a bigger part of the overall economy and society.

2.1.2 Definition (Urbanisation). Urbanization is also a social process, even though it does not directly involve physical space (as opposed to urban growth). It refers to the changes in behavior and social relationships that people experience when living in close proximity within urban environments [11].

2.1.3 Definition (Urban sprawl). According to Sudhira and Ramachandra (2007) [11], urban sprawl describes how cities grow in a messy, unplanned way. This can happen along the boundaries of existing cities or along highways and roads leading out of them. These sprawling areas often lack basic necessities like clean water, toilets, and healthcare because planners (policymakers) could not predict how much the city would grow.

2.1.4 Definition (Land use). It describes how humans interact with the land, such as for housing (residential), businesses (industrial), farming (agricultural), or leisure (recreational).

2.1.5 Definition (Land cover). It refers to what covers the Earth's surface, like forests, grasslands, fields used for crops, and bodies of water [11].

2.1.6 Definition (Urban Analysis). It is the most important component in urban planning that utilizes spatial and statistical analysis to understand the patterns, processes, and complexities within urban areas. It involves analyzing datasets from sources like censuses, surveys, satellite imagery, and social media feeds for a better understanding of urban phenomena. In our study, we will use surveys and satellite imagery to evaluate urban patterns in Uganda [11].

2.1.7 Definition(Remote Sensing). Jensen (2006) explains that remote sensing involves gathering information about Earth from a distance, without physically touching it [11]. This is achieved using instruments like cameras, scanners, and lasers mounted on airplanes or satellites. These instruments capture data across different parts of the light spectrum, including invisible light like ultraviolet and infrared, as well as the visible light we can see. Scientists then analyze this collected data using both visual inspection and computer processing techniques.

2.1.8 Definition(Geospatial dataset). It is an image dataset accompanied by geographic metadata that describes the geometries of each polygon, point, or line in the image [2].

2.1.9 Definition(Non geospatial image). It is a image dataset without metadata [2].

2.1.10 Definition(Semantic segmentation). It is a computer vision task that involves predicting labels for each pixel in an image such that each class has a unique label [2].

2.1.11 Definition(Coordinate Reference System(CRS)/ Spatial Reference). It is a system that defines how to locate geographic entities on a projected surface (a geometric process to show the curved surface 3D of the Earth on a flat map 2D) [2].

2.1.12 Definition(Band). It is a specific spectral range or channel of the electromagnetic spectrum acquired by the imaging sensor. Red, green, and blue are examples of bands, with each representing a specific wavelength of light [2].

2.1.13 Definition(Tile/scene). It is a single image file captured by a remote sensor, such as a satellite, that covers a specified area of the Earth [2].

2.1.14 Definition(chip/patch). It is a smaller image sampled from a larger tile [2].

2.2 Change Detection (CD)

2.2.1 Definition (Change Detection [12, 19]). Change detection is like finding differences in pictures of the same place taken at different times. These pictures are usually satellite images that capture how the land cover (forests, buildings, etc.) has changed over time, whether due to natural causes or human activity. Figure 2.1 illustrates this concept.

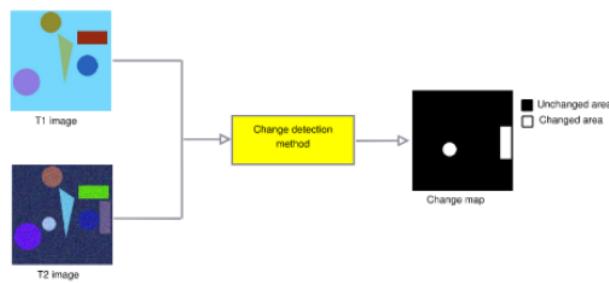


Figure 2.1: Graphical illustration of the change detection problem [19].

2.2.2 Definition (Mathematical formulation of CD [19]). Let I_1 and I_2 be two co-registered images of size $W \times L$ representing the same geographical region. These images are captured at times t_1 and t_2 , respectively, using the same sensor (assuming a classic monomodal case). We can represent them mathematically as:

$$I_1 = \{I_1(x, y) \mid 1 \leq x \leq W, 1 \leq y \leq L\} \quad (2.2.1)$$

$$I_2 = \{I_2(x, y) \mid 1 \leq x \leq W, 1 \leq y \leq L\}, \quad (2.2.2)$$

where (x, y) represents the pixel coordinates.

The goal of a change detection system is to generate a binary change map (CM) that identifies areas with changes between the two images:

$$CM = CM(x, y) \in \{0, 1\} \mid 1 \leq x \leq W, 1 \leq y \leq L. \quad (2.2.3)$$

Here, $CM(x, y)$ represents the change information for a specific pixel at (x, y) . Traditionally, the change map can be obtained through a difference image (DI) operation using functions like subtraction or log-ratioring:

$$DI = |I_1 - I_2|. \quad (2.2.4)$$

2.2.3 Machine learning for Change Detection in remote sensing.

In this section, we present some related work of change detection in remote sensing images using deep learning techniques.

Ayesha et al. [30] reviewed Deep Learning approaches for Change Detection (CD) in remote sensing imagery. They highlighted the use of various image types for CD, including Synthetic Aperture Radar (SAR), Multispectral (MS), and Hyperspectral (HS) acquired from active (SAR) and passive (MS) sensors. The authors discussed traditional CD methods, categorized as Pixel-Based (PBCD) and Object-Based (OBCD) based on the analysis unit (Figure 2.2). PBCD, the older approach, compares individual pixels for change but struggles with variations in light and misalignment (the image comes from different CRS) between images. This makes PBCD less effective for Very High Resolution (VHR) imagery where object variability is higher. OBCD overcomes these limitations, improving CD accuracy for VHR images. Here, objects are first segmented from the image, and then compared and analyzed between periods. Data availability and accuracy also impact change detection. Deep learning approaches offer a powerful framework for CD in remote sensing, as shown in the general workflow of Figure 2.3.

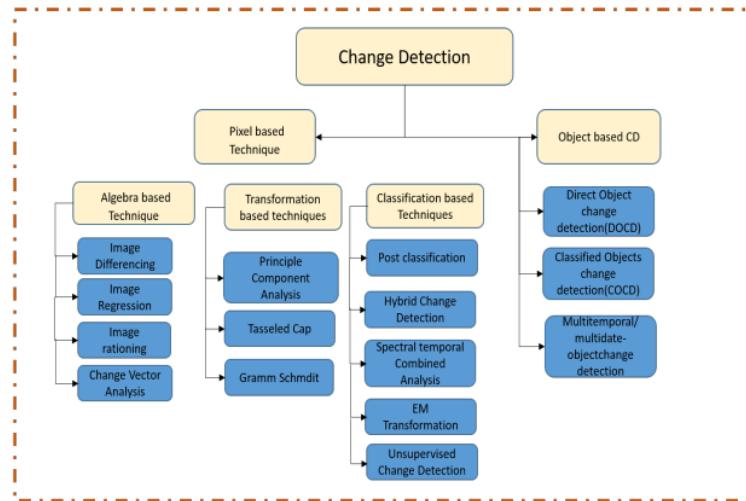


Figure 2.2: Traditional Change Detection [30].

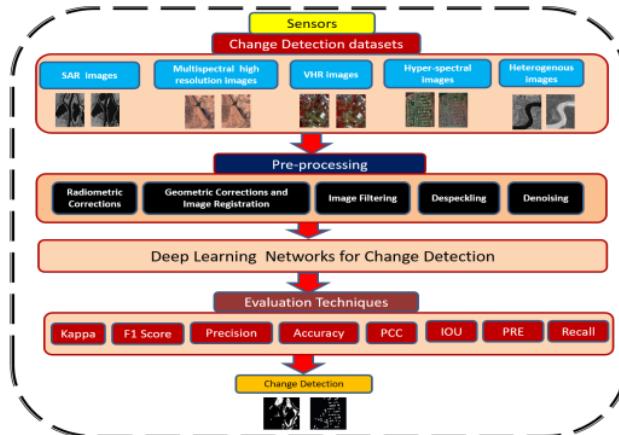


Figure 2.3: General workflow of change detection [30].

Sefin et al. [29] explored land cover classification and change detection using Sentinel-2 satellite images. These images capture Earth's surface across multiple time points (multitemporal) and various wavelengths of light (multispectral, around 11 bands). They tested two deep learning architectures: a fully convolutional neural network (FCN) and a FCN combined with a long short-term memory (LSTM) network. Their findings showed that the LSTM approach, which considers information from multiple periods, outperformed the FCN using only a single image. However, they limited the number of images input into the LSTM network. For future research, they proposed investigating the impact of using a variable number of input images for the LSTM network, potentially leading to even better results.

Another study of Sefin et al. [40], proposes a new approach called Single Temporal Supervised Learning (STAR) for change detection in unpaired images. Unlike traditional methods that require paired "before and after" images with labels, STAR utilizes unpaired images and explores a different perspective: detecting changes at the object level. To demonstrate STAR's effectiveness, the authors introduce ChangeStar, a simple change detector. ChangeStar uses existing deep learning models designed for semantic segmentation (understanding image content) and incorporates a new module called ChangeMixin. This ChangeMixin module empowers any existing semantic segmentation model to detect object-level changes in unpaired images. This is significant because it allows ChangeStar to use the power of pre-trained segmentation models for change detection tasks without the need to design a completely new architecture. This approach bridges the gap between semantic segmentation and change detection. The study reports that STAR outperforms existing methods when using only single-timepoint data (unpaired images). It also achieves superior performance even when compared to methods utilizing paired images for training (bitemporal supervision). This suggests that STAR is a promising approach for change detection tasks.

Furthermore, Wele Gedare et al. [9] proposed a new architecture for CD called ChangeFormer. Unlike most existing methods that depend on convolutional neural networks (ConvNets), ChangeFormer uses a transformer-based Siamese network architecture. This approach combines a hierarchical transformer encoder with a Multi-Layer Perception (MLP) decoder within a Siamese network. This design allows ChangeFormer to efficiently capture long-range details across different image scales, which is crucial for accurate change detection. Experiments on two different datasets showed that ChangeFormer outperformed previous methods like U-Net, U-Net++, and ResNet in terms of CD performance. This suggests

that ChangeFormer is also a promising new approach for change detection tasks.

2.3 Land Cover Classification (LCC)

This section explains land cover classification by presenting some related work using machine learning techniques and/or Geography Information Systems (GIS)

Land use and land cover classification help understand what is on the ground, like trees, water, or buildings by using remote sensing techniques to classify those categories of ground cover. This helps us know more about the Earth's surface and how land is being used. It is useful for protecting the environment, planning cities, and managing land resources better [38].

Boipuso et al. [35] utilizes the findings of various studies, affirming the effectiveness of supervised machine learning classifiers in Land Use Land Cover (LULC) mapping. The authors implemented three traditional classifiers namely Gradient Tree Boosting (GTB), Random Forest (RF), and Support Vector Machine (SVM), and compared them with a simple Multilayer Perceptron Artificial Neural Network (MLP-ANN). Employing Landsat data from 1984 to 2020 at five-year intervals for the Greater Gaborone Planning Area (GGPA) in Botswana, RF emerged as the top-performing classifier with an overall average accuracy of 92.8%, followed by MLP-ANN (91.2%), SVM (90.9%), and GTB (87.8%). To enhance the ML classifiers' potential, the authors implemented a post-classification feature in-feature-out (FEI-FEO) fusion approach. This technique significantly improved urban LULC class mapping accuracy by leveraging the feature detection and classification capabilities inherent in machine learning classifiers. The study findings underscore the nuanced nature of urban LULC class detection and mapping, highlighting dependencies on factors such as sensor spectral resolutions, as well as temporal, atmospheric, illumination, and geometric variations. The authors concluded that MLP-ANN was the preferred classifier for mapping urban built-up areas and water bodies over 35 years. Additionally, RF was identified as the optimal classifier for extracting vegetation classes, with MLP-ANN suitable for grass, and RF for both woody vegetation and forested areas.

Pall et al. [15] implemented RF for land cover classification of a multisource remote sensing and geographic data set, which is both a challenging and important classification problem in remote sensing. They highlight that RF is a bagging(bootstrap aggregating) ensemble method that is composed of tree-type classifiers as decision trees (Classification and Regression Tree (CART)). Bagging (or bootstrap aggregating) is based on training many classifiers on bootstrapped samples from a training set and then aggregating the result of each classifier either average (for regression task) or voting (for classification task) to obtain a final prediction. Their result shows that RF outperformed the single CART classifier and is much faster in training compared to the boosting ensemble method (it uses iterative re-training, where the incorrectly classified samples from the previous weak learner are given as input to the next learner while increasing weighting as the iterations progress. The final prediction is determined by the weighted performance of each classifier).

Shengyu et al. [39] provided an overview of deep learning in land use land cover classification (LULCC). They explained five deep learning methods: convolutional neural networks (CNNs), autoencoders (AEs), generative adversarial networks (GANs), and recurrent neural networks (RNNs), along with two baseline datasets: pixel-level and patch-level for LULCC. The authors emphasized the unique advantage of deep learning compared to classical machine learning. Unlike traditional methods that require manually designed features, deep learning automatically learns and extracts relevant features for the task, enhancing robustness and ease of model migration across different datasets. Deep learning algorithms can handle

large-scale data and discover potential patterns and regularities, thus improving classification accuracy and effectiveness. However, they also discussed the limitations of deep learning in LULCC, such as data diversity, category imbalance, labeling issues, model generalization, and contextual modeling challenges due to spatial correlations in remote sensing data. To address these limitations, they proposed advanced techniques, as illustrated in the general workflow diagram shown in Figure 2.4.

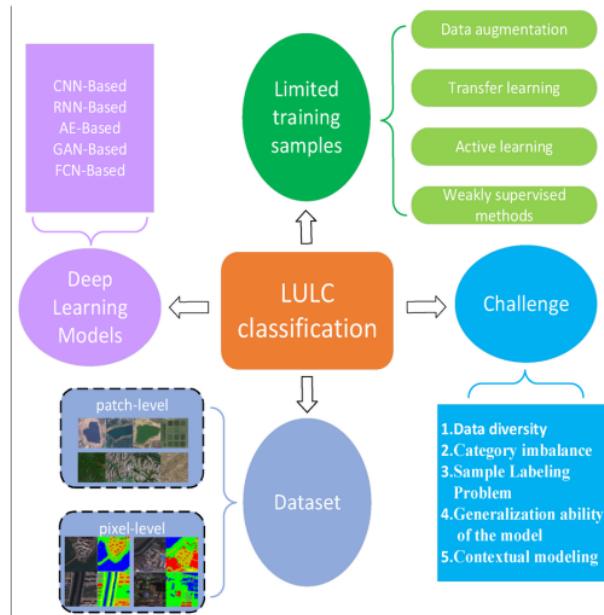


Figure 2.4: LULCC general framework diagram [39].

Islam Atef et al. [7] implemented the maximum likelihood algorithm (MLH) and machine learning techniques, including the random forest algorithm (RF) and support vector machine (SVM), for image classification to monitor spatiotemporal land use changes in El-Fayoum governorate, Egypt. They processed the images using the Google Earth Engine (GEE). The machine learning techniques were applied in GEE using an inbuilt classifier package, while the maximum likelihood algorithm was applied in GIS using the processed images exported from GEE. Land use and land cover (LULC) changes were assessed over the last 20 years in three different periods: 2000-2012, 2012-2016, and 2016-2020. Their results showed that SVM provided the most accurate maps in terms of the kappa coefficient (0.916) compared to MLH (0.878) and RF (0.909).

2.4 Urbanization and service delivery

In this section, we present some related work on urbanization leading to either poor or good service delivery in African countries. For the last several decades, cities in sub-Saharan Africa (SSA) have been growing at historic rates, as presented in Chapter 1. At the same time, since most of them are very poor, they have not been able to keep up with the need to provide the most essential services to a large proportion of their population.

Richard Stren et al. [33] in their work discusses the role of international assistance in urban service delivery in Africa. Their goal is to provide insights into the challenges and opportunities for international donors in supporting urban service delivery in Africa. To do that, they highlight the difficulties urban governments and service providers have encountered in delivering services, particularly in the context of rapid urbanization and limited resources. Also, they emphasize the need for international assistance to support capacity building, institutional reforms, and innovative financing mechanisms to improve urban service delivery in Africa.

Furthermore, the work of Danielle Resnick introduces a special issue focusing on key governance challenges related to addressing gaps in urban service delivery in Africa. He aims to highlight the role of politics and policies in urban governance and service delivery in African cities. He discusses two major trends affecting urban service delivery in Africa: 1) the transfer of service provision responsibilities to sub-national authorities due to decentralization policies, and 2) the increasing control of opposition parties over many African cities, leading to "vertically-divided authority" where central governments lack incentives to help municipal governments improve their performance. Case studies of Senegal, South Africa, and Uganda illustrate how these dynamics can become problematic for urban service delivery.

Finally, the work of Somik Lall discusses the challenges of service delivery in Africa's rapidly growing cities. He highlights the need for African policymakers to address issues of urban service delivery to improve the quality of life for the urban population. He shows that as Africa has rapidly urbanized, the number of people living in underserved neighborhoods doubled from 100 million in 1990 to 200 million in 2014. Citizens living in shantytowns (an area in or near a city in which poor people live in small, badly built houses) face worse conditions than those living elsewhere in cities, with lower access to piped water, flush toilets, and electricity. The authors argue that to make cities work for development, African policymakers need to raise agricultural productivity, improve the rural sector, adopt well-functioning land markets, enhance land-use planning, and embrace new technologies.

2.5 Our Contributions

In this work, we will use the pre-built algorithms and tools provided by Google Earth Engine, a cloud-based platform, for geospatial analysis tasks like image classification and change detection. This platform will assist us in preprocessing, analyzing, and visualizing Landsat 7 data collected from 2011 to 2015 in Uganda. We will implement three machine learning techniques including CART, random forest, and support vector machine for land cover classification. After identifying the best-performing model using the kappa coefficient as a key metric for mapping urban changes, we will use Google Earth Engine's pre-built algorithms to generate transition change maps between those two years.

3. Methodology

In this chapter, we begin by outlining our research design and detailing the methodology employed in our study. Next, we delve into the workings of three supervised machine learning algorithms, explaining their underlying principles and algorithms. Finally, we conclude with a discussion of the metrics we will use to assess the performance of our models.

3.1 Research Design

Inspired by the general workflow of change detection present in Figure 2.3, we design the flowchart of our study present in Figure 3.1.

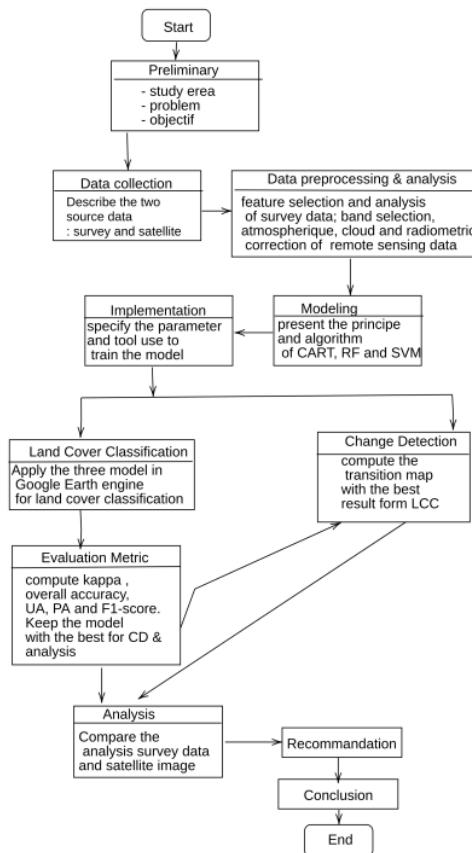


Figure 3.1: Flow Chart of our Methodology Research

Following the diagram presented in Figure 3.1, the preliminary aspects are already covered in the in-

introduction of our work. However, the description of the data, implementation of the model, analysis, and conclusion will be presented in Chapter 4. For the next section of this chapter, we will explore the modeling part alongside the mathematical principles behind each method used and the definition of the evaluation metric.

3.2 Classification and Regression Tree (CART)

CART is a variation of the decision tree algorithm produced for the first time by Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone in 1984. It is a supervised learning algorithm that learns from labeled data to predict unseen data by drawing a flowchart (tree) mapping a decision-making process. This tree is organized into nodes and branches. The nodes represent different decision points, and the branches (or edges) represent the possible outcomes of those decisions (usually True/False).

3.2.1 Definition (Nodes [32]). There are three types of nodes.

- **The root node:** Also called the decision node, is the start of the decision process (the first test performed on the whole training dataset) and represents a choice of features or attributes that will result in the subdivision of all the observed data points for the training set into two exclusive subsets.
- **The internal nodes:** Also known as chance nodes, act like decision points in the tree. They represent different possibilities or choices you can make at that step. Each internal node has a single line coming from above (its parent node) and can have multiple lines branching down below (its child nodes or leaf nodes).
- **The leaf nodes:** Also called the end nodes, are the final destinations on the tree. They represent the final answer or prediction for each branch of the decision process.

A decision tree works by asking a series of questions, like a branching path on a map. Each question leads to different possibilities, represented by branches in the tree. As you answer the questions (follow the branches), you eventually reach a final answer, which is like a leaf on the tree. Another way to think about it is like a set of "if-then" rules. For example, "if condition 1 and condition 2 and condition ... and condition k occur, then outcome j occurs."

3.2.2 CART algorithm [31]. The CART algorithm involves finding the best features that divide the input space into two subsets using the Greedy approach, which is known as a recursive binary node splitting. This best attribute is the one that best reduces the impurity of the resulting subsets. The impurity at each node is determined by Gini's diversity index, which evaluates the probability of misclassifying a randomly chosen element from a subset labeled according to the majority class, based on the specific variable used for node splitting.

The Gini impurity at each node i is defined as :

$$G_i = 1 - \sum_{k=1}^{N_{\text{classes}}} \left(\frac{N_{k,i}}{N_i} \right)^2 = 1 - \sum_{k=1}^{N_{\text{classes}}} (p_k)^2, \quad (3.2.1)$$

with $N_{k,i}$ the number of data samples of class k in node i and N_i the total number of data samples in node i . The terms p_k in the sum are equivalent to the probability of getting a sample of class k in the node i .

5

The Gini's impurity index ranges from 0 (100% pure node) to 1 (very impure node) and a Gini's index of 0.5 means the data points are uniformly distributed into some classes.

The recursive process of the Greedy approach has two main steps as presented the Algorithm 1, where the cost function is defined as :

$$J(j, t_j) = \frac{N_{\text{left}}}{N_{\text{node}}} G_{\text{left}} + \frac{N_{\text{right}}}{N_{\text{node}}} G_{\text{right}}, \quad (3.2.2)$$

with N_{left} the number of data points whose feature x_j is such that $x_j < t_j$, N_{right} the number of data points whose feature x_j is such that $x_j \geq t_j$, N_{node} the number of data points in the node considered, and $G_{\text{left}}, G_{\text{right}}$ Gini's indices for each subset on both side of t_j .

The splitting criteria also known as hyperparameter tuning help the Decision Tree to avoid overfitting including; max depth (maximum depth of the decision tree), minimum sample split (minimum number of samples required to split an internal node), minimum sample leaf(minimum number of samples required to be at a leaf node), and max features (number of features to consider when looking for the best split)[18].

Algorithm 1: Decision Tree pseudo code

- 1 **Initialization:** at the root node with depth equal = 0 ;
 - 2 **STEP 1: THRESHOLD COMPUTATION;**
 - 3 **for** j from 1 to n **do**
 - 4 **for** t_j scanning the range x_j^{\min} to x_j^{\max} **do**
 - 5 | Compute the cost function $J(j, t_j)$;
 - 6 **end**
 - 7 Find the pair (j, t_j) that minimizes $J(j, t_j)$;
 - 8 **end**
 - 9 **STEP 2: BRANCHING;**
 - 10 Split the dataset along the feature j using the threshold t_j into two subsequent new nodes;
 - 11 depth = depth + 1 Repeat STEP 1 at each new node up to the stopping rules ;
 - 12 **Exit conditions:**
 - 13 At least one condition on one of the hyperparameters defined by the splitting criteria is fulfilled;
-

In some situations, the stopping rules defined by the splitting criteria do not work well to prevent overfitting in the decision tree model. Thus, we use pruning, a technique that aims to remove nodes that contribute little to the model's accuracy. There are two approaches for pruning: the pre-pruning (forward pruning) approach, in which the splitting or partition of the tree is stopped at a particular node judged to have non-significant accuracy; whereas in the post-pruning(backward) approach, subtrees are removed from the full tree. Nodes are removed based mostly on cost complexity (nodes with negative cost) and information gain (nodes with low information). The Information Gain at a specific node related to a certain feature measures how important this feature is for discriminating between the classes. It is defined as the difference between the entropy of the parent node and the weighted average entropy of the child node[16].

$$\text{Information gain} = \text{Entropy}(i) - \sum_{j=1}^m \frac{N_j}{N_i} \times \text{Entropy}(j), \quad (3.2.3)$$

where N_j is the number of data points n the j^{th} child node, N_i the total number of data points in the

parent node, m is the number of child nodes and the Entropy(i) is defined by

$$\text{Entropy}(i) = - \sum_{k=1}^{N_{\text{classes}}} p_k \log_2(p_k), \quad (3.2.4)$$

with p_k representing the probability of class k , computed as the proportion of class k in the dataset at node i .

It's crucial to note that this information gain is highly effective for feature selection, as implemented by random forest algorithms (ensemble of decision trees).

3.3 Random Forest

In this section, we explore the details of Random Forest (RF) and its application, focusing particularly on its use in classification tasks.

As defined in Section 2.3, where we presented the work of Pall et al. [15], RF is composed of multiple decision trees combined, as presented in Figure 3.2. These trees are trained on random subsets of labeled samples from the training set. Then, the predictions of each tree are aggregated (using either majority voting for classification or averaging for regression) to obtain the final prediction [13].

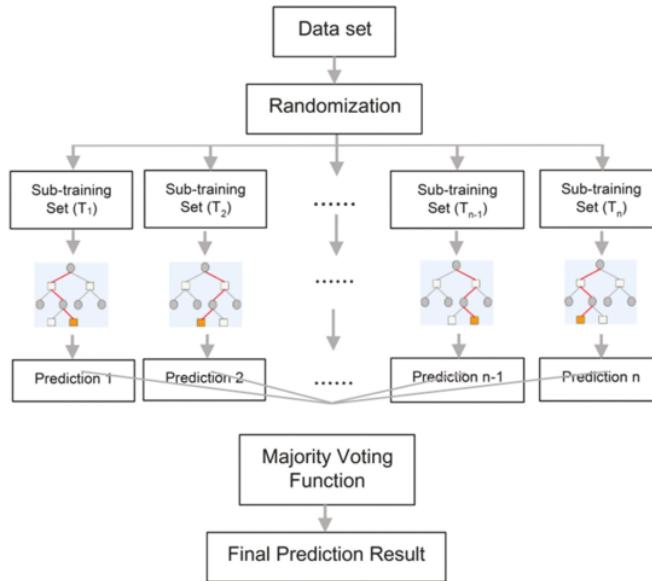


Figure 3.2: Illustration of the random Forest classification structure [13].

The majority voting function selects the class predicted by the most trees[13]. This function is given as

:

$$\hat{y}(x) = \arg \max_c \left(\sum_{i=1}^n 1((p_i(x, c) > 0.5)) \right), \quad (3.3.1)$$

where $1((p_i(x, c) > 0.5))$ is the indicator function that returns 1 if the predicted probability of the input x belongs to a class c by the tree i denoted $p_i(x, c)$ is greater than 0.5 and 0 otherwise. When this indicator function is 1, it contributes 1 to the vote of the input x to belongs to the class c , then the $\arg \max_c$ returns the class c that receives the maximum votes for all the tree n (n is the number of the tree)

3.3.1 RF algorithm. The Random Forest algorithm builds on the idea of decision trees. It creates multiple decision trees (a forest), each using a slightly different approach. Here is how it works: for each tree, it randomly selects a subset of data points (bootstrapping) from the original data set. Then, when building the tree, instead of considering all possible features F at each branching point, it randomly selects a smaller subset of features $f \subseteq F$. The tree then chooses the best split among those features to make the decision. This randomness in data selection and feature selection helps prevent the trees from becoming excessively dependent on specific patterns in the data, ultimately leading to a more robust and accurate forest for classification or prediction tasks [17].

Algorithm 2: Random Forest

```

1 Precondition: A training set  $S = (x_1, y_1), \dots, (x_n, y_n)$ , features  $F$ , and number of trees in forest
    $B$ .
2 function RandomForest( $S, F$ ):
3    $H \leftarrow \emptyset$ ;
4   for  $i$  from 1 to  $B$  do
5      $| S_i \leftarrow$  A bootstrap sample from  $S$ ;
6      $| h_i \leftarrow$  RandomizedTreeLearn( $S_i, F$ );
7      $| H \leftarrow H \cup \{h_i\}$ ;
8   end
9 end function
10 Function RandomizedTreeLearn( $S, F$ ):
11 At each node:
12    $f \leftarrow$  a very small subset of  $F$ ;
13   for  $j$  from 1 to  $f$  do
14      $|$  for  $t_j$  scanning the range  $x_j^{\min}$  to  $x_j^{\max}$  do
15        $|$  Compute the cost function  $J(j, t_j)$ ; Find the pair  $(j, t_j)$  that minimizes  $J(j, t_j)$ ;
16      $|$  end
17   end
18 split on best feature in  $f$  return The learned tree;
19 end function

```

3.4 Support Vector Machine (SVM)

Support Vector Machines (SVMs) are a powerful tool in machine learning, particularly for categorizing things (classification)[25]. It is as if we have data points representing different categories, like apples

and oranges. A SVM aims to draw a clear dividing line (hyperplane) in this space to separate the apples from the oranges. This line is positioned strategically to maximize the distance between the closest apples and the closest oranges (support vectors). These support vectors are like the training wheels for the SVM, helping it define the best possible boundary(optimal hyperplane) for future classifications.

As a stand-out widely utilized technique, SVM is implemented in binary classification(only two possible outcomes, eg: Yes or No) and multiclassification (more than two outcomes, eg: Land use classification). Consider a binary classification as defined by Yashon et al.[36], and a labeled training data set:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}; \quad x_i \in \mathbb{R}^d \text{ and } y_i \in \{1, +1\}, \quad (3.4.1)$$

where x_i is a feature vector representation, y_i is the class label of training and n are the elements in the training data sets. The optimal hyperplane is defined by

$$w^T x + b = 0, \quad (3.4.2)$$

where w is the weight vector, x is the input feature vector and b is the bias. w and b , respectively, satisfy the inequalities for all elements of the training set defined in Equation (3.4.3-3.4.4)

$$w^T x_i + b \geq 1, \text{ if } y_i = +1 \quad (3.4.3)$$

$$w^T x_i + b \leq -1, \text{ if } y_i = -1. \quad (3.4.4)$$

The vectors x_i for which Equation 3.4.5 is satisfied, are support vectors.

$$y_i(w^T x_i + b) = 1. \quad (3.4.5)$$

The decision function is defined;

$$f(x) = \text{sign}(w^T x + b). \quad (3.4.6)$$

We distinguished two main types of SVM: linear SVM and Non-linear SVM.

3.4.1 Linear Support Vector Machine. Linear SVM is best suited for data that can be clearly divided into two groups using a straight line, with no data points falling exactly on the line itself (perfectly separable). This is called a hard margin SVM. Here, the objective of SVM is to maximize the margin under the constraint that all data points must lie on the correct side of the hyperplane. This leads to a constrained optimization problem that can be solved by Lagrangian[14].

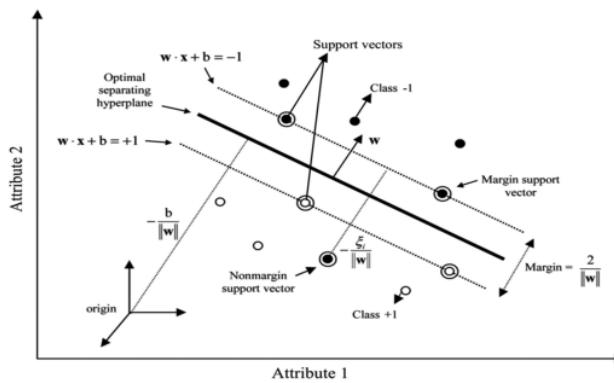


Figure 3.3: Maximum margin-minimum norm classifier in support vector machine with optimal hyperplane for linearly non-separable classes[27].

As present in Figure 3.3 the total margin is equal to $\frac{2}{\|w\|_2}$. This is because the distance from one margin (the distance between the hyperplane and the observations closest to the hyperplane which are the support vectors) to the optimal hyperplane is defined as

$$\frac{|w^T x + b|}{\|w\|_2} = \frac{1}{\sqrt{w^T w}} = \frac{1}{\|w\|_2}. \quad (3.4.7)$$

To maximize this margin, we thus need to minimize $\|w\|$ such that discrimination boundary is obeyed, i.e, under the combined constraints from Equations 3.4.3 and 3.4.4

$$y_i(w^T x_i + b) \geq 1, i = 1, \dots, n. \quad (3.4.8)$$

Therefore, we get the constrained optimization problem

$$\min_{w,b} \frac{1}{2} w^T w \text{ such that } y_i(w^T x_i + b) \geq 1. \quad (3.4.9)$$

We solve this by introducing Lagrange multipliers (α_i) and get the Lagrangian function of SVM

$$\min L(w, b, \alpha_i) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i (y_i(w^T x_i + b) - 1), \text{ such that } \forall i, \alpha_i \geq 0; \quad (3.4.10)$$

which is differentiable with respect to w and b .

$$\nabla_w L(w, b, \alpha) = 0 \Rightarrow w = \sum_{i=1}^n \alpha_i y_i x_i \quad (3.4.11)$$

$$\nabla_b L(w, b, \alpha) = 0 \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0. \quad (3.4.12)$$

w , derived in Equation (3.4.11) expresses that the optimal hyperplane solution can be written as a linear combination of training vectors. Therefore the decision function 3.4.6 is write as

$$f(x) = \operatorname{sgn} \left(\sum_{i=1}^n \alpha_i y_i x_i x + b \right). \quad (3.4.13)$$

We can note that only vectors x_i with $\alpha_i > 0$ have an effective contribution to the sum (3.4.11). Substituting (3.4.11) and (3.4.12) in (3.4.10), we get the dual problem of SVM: instead of minimizing over w , b , subject to constraints involving α 's, we can maximize over α (the dual variable) subject to the relations obtained previously for w and b from Equation (3.4.11) and (3.4.12);

$$\max L(\alpha_i) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j, \text{ such that } \sum_{i=1}^n \alpha_i y_i = 0 \text{ and } \alpha_i \geq 0. \quad (3.4.14)$$

This dual problem helps to deal with non-linear separate data.

3.4.2 Non-linear Support vector Machine. In non-linear SVM the process is very similar to linear SVM, we just need to take into account a very useful tool, the kernel trick which is a kernel function to map the data into a higher dimensional space where it is more likely to be linearly separable[10]. The kernels are similarity functions that return inner products between the images of data points.

$$K : X \times X \rightarrow \mathbb{R}, (\vec{x}, \vec{z}) \mapsto \langle \phi(\vec{x}), \phi(\vec{y}) \rangle. \quad (3.4.15)$$

As we define the dual function in Equation 3.4.16, we replace the dot product with the kernel function which can often be computed efficiently even for very high-dimensional spaces. So, the function we end up optimizing is:

$$\max L(\alpha_i) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i^T x_j), \text{ such that } \sum_{i=1}^n \alpha_i y_i = 0 \text{ and } \alpha_i \geq 0; \quad (3.4.16)$$

In this case, the decision function is written as :

$$f(x) = \operatorname{sgn} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right). \quad (3.4.17)$$

The common kernel functions defined for this purpose are :

- Polynomial: $K(x, y) = (x \cdot y + 1)^p$ with p the parameter turning (it is the degree of the polynomial).
- Radial Basis Function (RBF)(Gaussians): $K(x, y) = \exp \left(-\frac{\|x-y\|^2}{2\sigma^2} \right) = \exp(-\gamma \|x-y\|^2)$ with γ the degree of smoothing and shape of the hyperplane dividing the class (Increasing the γ parameter can alter the shape of the class-dividing hyperplane, which may consequently influence the accuracy of the classification results [23])
- Sigmoid : $K(x, y) = \tanh(kx \cdot y - \delta)$, where k is a gain and δ is a offset.

In the case of misclassification(soft margin SVM [22]), we need to minimize the misclassification error, meaning we'll have to deal with one more constraint. The loss of misclassification is called a slack variable and is added to the problem that we have for hard margin SVM. Thus, the optimization problem for soft margin SVM is;

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \zeta_i \text{ such that } y_i(w^T x_i + b) \geq 1 - \zeta_i \quad \forall i = 1, \dots, n, \quad \zeta_i \geq 0, \quad (3.4.18)$$

where C is the cost parameter that controls the trade-off between maximizing the margin and minimizing the loss(larger C values are preferred to minimize model over-fitting), ζ_i as shown in Figure 3.3 is a slack variable that adds flexibility for misclassifications of the model. The Lagrangian function derived from Equation (3.4.18) ultimately leads to the dual problem, with the only distinction being that $0 \leq \alpha_i \leq C$.

Our work employs the RBF kernel function, which has been shown by Knorn et al. [20] and Yu et al. [37] to be particularly effective for land cover classification due to its accuracy and reliability. As [8, 28] demonstrated in their work, we need to choose carefully the cost parameter C and the parameter γ to get good results.

3.5 Performance Evaluation Metric

To evaluate the performance of the classifier we have used the metric present in Equation (3.5.1-3.5.5) [24].

$$UA = \frac{K_{ii}}{K_{i+}} \quad (3.5.1)$$

$$PA = \frac{K_{jj}}{K_{+j}} \quad (3.5.2)$$

$$OA = \frac{\sum_{i=1}^n K_{ii}}{T} \quad (3.5.3)$$

$$\kappa = \frac{2 \cdot UA \cdot PA}{UA + PA}, \quad (3.5.4)$$

$$F1\text{-score} = \frac{2 \cdot UA \cdot PA}{UA + PA}, \quad (3.5.5)$$

where PA is the producer's accuracy; UA is the user's accuracy; OA is overall accuracy; κ is the Kappa coefficient; n is the number of classes; K_{ii} is the number of pixels correctly classified in class i ; K_{i+} and K_{+j} represent the total number of pixels in row i and column j of the confusion matrix, respectively; and T is the total number of pixels used for evaluation(total number of pixel in test set).

1. **Confusion Matrix:** is a table that helps us understand how well a classification model performs. It shows how many things the model classified correctly (diagonally) and how many were classified incorrectly (off-diagonally).
2. **User's Accuracy (UA):** Imagine we are looking at a map classified by the model as "forest." UA tells us the proportion of pixels truly belonging to the "forest" class out of all pixels the model classified as "forest." In simpler terms, it measures how good the model is at not misclassifying other classes as "forest." (Equation 3.5.1)
3. **Producer's Accuracy (PA):** This metric flips the perspective of UA . Let's say a reference map shows a specific area as a "forest." PA tells us the proportion of pixels the model classified as "forest" that actually belong to the "forest" class on the reference map. Put simply, it measures how good the model is at identifying actual "forest" areas. (Equation 3.5.2)
4. **Overall Accuracy (OA):** This is the most straightforward metric. It simply calculates the percentage of pixels the model classified correctly across all classes. It ranges between 0 and 1, with values close to 1 indicating good accuracy. (Equation 3.5.3)
5. **Kappa Coefficient (κ):** While OA is easy to understand, it can be misleading if there's a high agreement between classes by chance. Kappa considers this by comparing the model's agreement with the reference map to what would be expected by pure chance. A value closer to 1 indicates better agreement beyond random chance. (Equation 3.5.4)
6. **F1-score:** This metric combines both UA and PA into a single score, addressing their limitations. It provides a balance between how good the model is at avoiding false positives (UA) and how good it is at finding true positives (PA). A higher $F1\text{-score}$ indicates better overall performance. (Equation 3.5.5)

4. Implementation and Results

This chapter follows the workflow established in Chapter 3. We'll begin by introducing the datasets used in this study and analyzing the Afrobarometer data. Next, we'll provide a brief overview of Google Earth Engine (GEE), a cloud platform for geospatial analysis. Following that, we'll delve into the implementation of three supervised machine learning algorithms (explained in Chapter 3) for land cover classification within GEE. We'll present the results of these classifiers, compare their performance, and choose the best based on the evaluation metric. The best-performing classifier will be compared with existing literature. We'll then utilize this classifier's output to perform change detection and calculate the transition matrix. Additionally, we'll leverage the results from the best classifier to identify the district with the most significant urban growth. Focusing on this district with positive urban growth, we'll filter the Afrobarometer dataset to analyze the residents' access to essential services. We'll also explore their perceptions regarding local government actions aimed at improving service availability.

4.1 Data Description

In this study, we combine survey data and satellite imagery to examine trends in public opinion and land cover changes.

1. Survey data (Afrobarometer)

- (a) **Source:** The Afrobarometer dataset (<https://www.afrobarometer.org/data/>), is a valuable resource for understanding public opinion across Africa. The platform offers a collection of yearly surveys conducted in over 38 African countries recorded in separate rounds for analysis.
- (b) **Focus:** Uganda, data from rounds conducted in 2011 and 2015.
- (c) **Content:**
 - 1624 observations and 333 features (rounds 2011 and 2015)
 - Covers a range of critical issues like:
 - Democracy and Governance
 - Economic Conditions
 - Access to Basic Services
 - Civil Society and Quality of Life
- (d) **Purpose:** Tracking changes in public opinion over time, particularly regarding:
 - **Monitoring Progress towards SDGs:** Analyzing progress towards achieving Sustainable Development Goals (SDGs) related to basic service access and government performance in improving it such as no poverty, zero hunger, good health and well-being, quality education, etc.
 - **Identifying Policy Gaps:** Highlighting areas where there is a clear need for reform, aiding policymakers in making informed decisions.
- (e) **Feature Selection:** From over 333 variables, we selected 16 relevant features for each year:
 - Access to essential services like health, education, water/sanitation, electricity, and food.
 - Presence of security and infrastructure (tarred roads).
 - Public perception of government's handling of service access (8 remaining features).

2. Satellite imagery (Landsat 7)

- (a) **Data Source:** The data used for training the classifier is a collection of level-1 raw scenes from Landsat 7's Enhanced Thematic Mapper Plus (ETM+) sensor, acquired by the United States Geological Survey (USGS). USGS website (<https://www.usgs.gov/landsat-missions/landsat-data-access>) from <https://www.usgs.gov/landsat-missions> (DOI: 10.5066/P9TU80IG) and be downloaded in GEE using <https://developers.google.com/earth-engine/datasets>.
- (b) **Data Format:** Level-1 raw scenes (multispectral images with 16 bands)
- (c) **Preprocessing:** The data comes pre-processed with :
- terrain correction (T1TP) calibrated,
 - scaled digital numbers (DN),
 - Geometric and radiometric improvements
- (d) **Band Selection:** Eight bands (**B1**(Blue), **B2**(Green), **B3**(Red), **B4**(Near-infrared), **B5**(Shortwave infrared 1), **B6_VCID_1**(Low-gain Thermal Infrared 1), **B7**(Shortwave infrared 2)) with 30m resolution (except **B6** with 60m) are used for training the classifier. We all these bands to improve classification performance.
- (e) **Visualization:** A true-color composite image (RGB - red, green, blue) is created from these bands for visualization purposes.
- (f) **Region of Interest (ROI):** The data is used to extract training samples from a specific area (Uganda) defined by a polygon from the "Large Scale International Boundary Polygons, Simplified" (LSIB 2017)collection.
- (g) **Time Period:** The entire Landsat archive (1999-2021) is available, but specifically, data for two years (source year: 2011, target year: 2015) is extracted for the Ugandan ROI.

3. Satelite imagery(LSIB 2017)

- (a) **Data Source:** The data can be accessed from: <https://geodata.state.gov/geonetwork/>
- (b) **Data Format:** The data comes in a compressed format containing two datasets:
- LSIB line vector file: This file defines country boundaries.
 - World Vector Shorelines (WVS): This dataset provides shorelines for countries around the world.
- (c) **Data Content:** The LSIB data includes four attributes for each boundary line:
- Abbreviation of the country (abbreviat)
 - Country code (country_co)
 - Country name (country_na)
 - Continent region (wld_rgn)
- (d) **Our Usage:** We use this data specifically to extract the polygon representing Uganda's borders. This polygon serves as our Region of Interest (ROI) for further analysis.

4. Satellite imagery (FAO GAUL 500m Simplified: Global Administrative Unit Layers (2015, Second-Level Administrative Units))

- (a) **Source:** The data is from the Food and Agriculture Organization's (FAO) Global Administrative Unit Layers (GAUL) initiative(<http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691>). GAUL aims to compile and distribute the most reliable information on administrative boundaries for every country globally. This dataset specifically uses the simplified

version with a 500-meter resolution, focusing on second-level administrative units for the year 2015.

- (b) **Data Content:** While GAUL offers various administrative levels, we only utilize two relevant variables out of 12 variables in their tables schema:
- Country Name (ADM0_NAME)
 - Name of the First Administrative Level (ADM1_NAME), likely representing districts or provinces.
- (c) **Our Usage:** We employ this data to calculate the urban class area within each district using our best classifier method. Subsequently, we extract the results into CSV files for further analysis.

5. Satellite imagery (MCD12Q1.061 MODIS Land Cover Type Yearly Global 500m)

- (a) **Source:** The data originates from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument aboard the Terra and Aqua satellites. The specific product is the MCD12Q1 Version 6.1, which provides global land cover classifications at yearly intervals spanning 2001 to 2021. We access this data from the following source: https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MCD12Q1.
- (b) **Content:** MCD12Q1 offers five distinct land cover classification schemes done by using the supervised method as we did in our study, but we will specifically focus on the fourth scheme, which categorizes land cover into nine classes. These classes can be further grouped into broader categories like water, vegetation (including five vegetation-related classes), barren land, and urban areas.
- (c) **Our Usage:** We use the MCD12Q1 data to validate the land cover classifications generated by our most effective classifier. This validated land cover data, along with the GAUL data (providing district information), will be used to calculate the urban area within each district. This comparison between the urban area classified by our best classifier and the urban area classified by MODIS (using MCD12Q1) will help us assess the accuracy of our classification model.

4.2 Preprocessing of Afrobarometer data

4.2.1 Data preparation. Following the description of the survey data in the previous section, we will provide more information about this data in this section.

The survey data lacked a "district" column for geographical analysis. To address this, we used the available household latitude and longitude coordinates. By mapping these coordinates to the Uganda district shapefile, we created a "district" column, which will help with district-level analysis.

4.2.2 Variable selection. After the preparation of our data, we focused on variables related to basic needs. The selection process for these variables was documented directly within the notebook, providing references to the Afrobarometer website. This documentation serves as metadata, providing clear explanations for each variable chosen and ensuring transparency in the analysis (the document is added at the link in the appendix).

The specific chosen variables are listed below with a small overview of the meaning.

Table 4.1: Variable selection

N°	Access to the service		Performance of the Local Government	
	Variable	Description	Variable	Description
0	dep_food	How often gone without food	hand_enoughfood	Handling: enough food
1	psu_elecgrid	PSU Services: electric grid	hand_electricity	Handling: electricity supply
2	psu_pipedwater	PSU services: piped water	hand_hhwater	Handling: household water
3	psu_sewage	PSU services: sewage	hand_education	Handling: educational services
4	psu_school	PSU Services: school	hand_crime	Handling: reducing crime
5	psu_policestation	PSU Services: police station	hand_health	Handling: health services
6	psu_healthclinic	PSU services: health clinic	locgov_locmarkets	LocGov: maintaining local markets
7	psu_market	PSU services: market		

The response to access to services is either 'Yes' or 'NO', while the performance of the government is graded as 'fairly well', 'very well', 'fairly badly', and 'very badly'. PSUs are Public Sector Undertakings that specify a given district.

The exploration of this dataset will be done within the context of urbanization trends in the section [4.6.2](#). We will select three districts based on their distinct urbanization patterns. Following this selection, we will delve deeper into the responses and public opinion trends over the five-year period covered by the analysis.

4.3 Google Earth Engine (GEE)

We use Google Earth Engine (GEE), a cloud-based platform, to analyze vast amounts of satellite imagery and classify land cover. GEE eliminates the need for local storage and offers a user-friendly JavaScript API. This interface also helps to implement classifiers like CART, Random Forest, and SVM directly and compute the performance described in Chapter [3](#). GEE offers a wide range of functionalities, we'll focus on a few key functions:

- **reduceRegion:** Calculates the area of each land cover class within specific districts.
- **subtract/and:** Identifies changes in land cover over time by subtracting corresponding bands from two images.
- **ee.Image.pixelArea():** compute the area of each pixel size in an image and convert from square meters to square kilometers(or hectares), providing a more user-friendly scale for area calculations.
- **ee.Dictionary:** Creates a container to store key-value pairs of information, allowing us to organize classification results for each district and export in CSV file.
- **addMap:** Visualizes images or feature collections directly on the GEE map interface for clear presentation.

4.4 Land Cover Classification

4.4.1 Workflow of Land Cover Classification in GEE.

The following steps outline the process of land cover classification map creation in Google Earth Engine (GEE):

1. Select Region of Interest (ROI): This involves using the LSIB 2017 data to define our area of interest.
2. Choosing and Filtering Image Collection: We filter the Landsat 7 image collection for the years encompassing our ROI. This typically involves selecting two time periods, such as '2015-01-01' to '2015-12-31' and '2011-01-01' to '2011-12-31'.
3. Collecting Training Points: Point collections representing each land cover class are created. Here, common classes might include water (0), barren land (1), vegetation (2), and urban areas (3).
4. Merging Training Points: All the individual training point collections are merged into a single variable.
5. Creating Training Dataset by Sampling Regions: The merged training point collection is used to sample regions at the pixel level. This sampled data is then split into a training set (typically 80%) and a testing set (20%).
6. Create the Classifier and Classify the Image: A classifier is created based on the chosen algorithm (e.g., Random Forest) and trained using the training dataset. This trained classifier is then used to classify the entire image.
7. Create the Classified Image: The classified image resulting from step 6 is used for subsequent change detection analysis.
8. Display the Results: The 'addMap' visualization function is employed to display the classified image on the GEE map interface. These results are typically presented in a figure.
9. Evaluate Performance with Confusion Matrix: A confusion matrix is generated to assess the performance of the classification model. Various performance metrics, explained in Chapter 3, are calculated using the confusion matrix.
10. Calculate Area for Each Class: Focusing on the urban class (class 3), we calculate the urban area for each district and store the results in a CSV file for further analysis.

4.4.2 Implementation of those classifiers in GEE.

For better performance of our classifiers, we tuned the hyperparameters of each model according to GEE, and we came out with:

1. **CART:** The inbuilt function `ee.Classifier.smileCart(maxNodes, minLeafPopulation)` has two parameters: `maxNodes`, the number of nodes in the tree (set this to 10), and `minLeafPopulation`, the minimum number of pixels to form a leaf node (set this to 1).
2. **RF:** The inbuilt function `ee.Classifier.smileRandomForest(numberOfTrees, variablesPerSplit, minLeafPopulation, bagFraction, maxNodes, seed)` has 6 parameters but we just focus on the first and use a forest of 10 trees, leaf others with default values.

3. **SVM:** The inbuilt function `ee.Classifier.libsvm(decisionProcedure, svmType, kernelType, shrinking, degree, gamma, coef0, cost, nu, terminationEpsilon, lossEpsilon, oneClass)` with 12 parameter. We choose the RBF kernel as the best one according to the methodology in Chapter 3, with a cost parameter set to 35 and the gamma parameter set to 0.5 (we have observed that a large cost and a small gamma give better performance).

4.4.3 Result of the three classifiers.

We display the composite satellite image RGB Landsat 7 of Uganda for both years in Figure 4.1 and 4.2.



Figure 4.1: Composite RGB Landsat 7 2011



Figure 4.2: Composite RGB Landsat 7 2015

Based on the composite image, we defined a classification scheme with four classes:

- Water (0): Represents water body pixels.
- Barren (1): Represents ground areas with no vegetation or urban structures.
- Vegetation (2): Encompasses all types of plant cover, from small crops to large woody areas.
- Urban Area (3): Represents essential building areas.

Classification maps were generated over two years (2011 and 2015) using three algorithms: CART, Random Forest, and SVM. The maps for 2011 are shown in Figures 4.3, 4.4, and 4.5, and the maps for 2015 are shown in Figures 4.6, 4.7, and 4.8. We observed that while CART and Random Forest produced highly similar maps across both years, the SVM classifications diverged significantly.

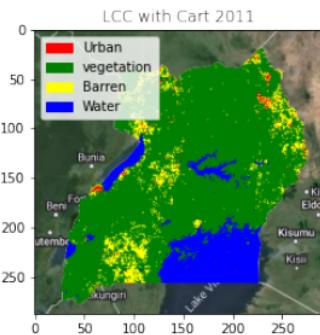


Figure 4.3: CART 2011

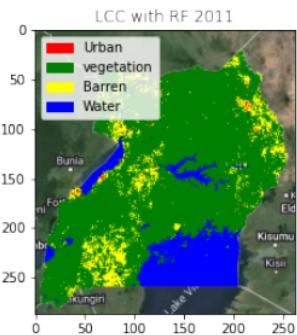


Figure 4.4: RF 2011

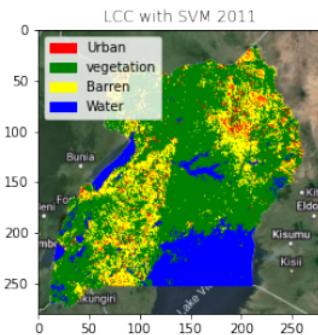


Figure 4.5: SVM 2011

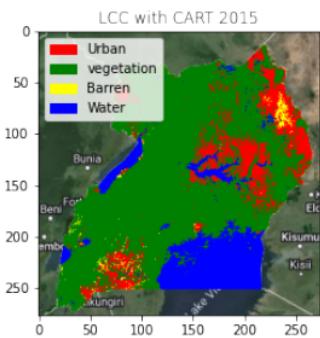


Figure 4.6: CART 2015

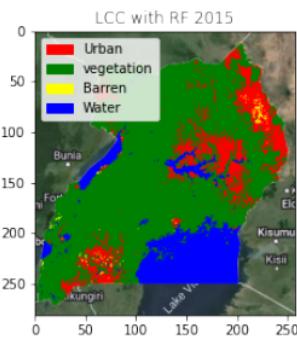


Figure 4.7: RF 2015

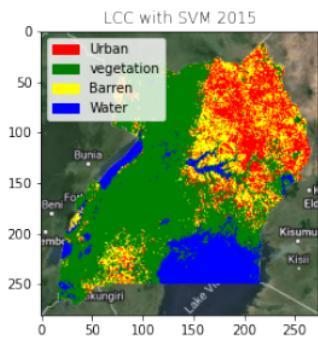


Figure 4.8: SVM 2015

To facilitate a quantitative comparison between classifiers, we calculated the area occupied by each class within the classification results. These areas are summarized in Table 4.2 and 4.3, along with the corresponding percentage for each class relative to the total area of Uganda City by all the models ($241,481.251 \text{ km}^2$). This value is consistent with estimates from both the actual shape of Uganda, which yields an area of $242,561 \text{ km}^2$, and literature sources, which indicate an area of $241,551 \text{ km}^2$ for the city (Nation Online Website).

Table 4.2: Summary of Class Area Percentages for CART, RF, and SVM (2011)

Class	CART		RF		SVM	
	Percentage (%)	Area (km ²)	Percentage (%)	Area(km ²)	Percentage (%)	Area(km ²)
0	14.79	35,704.84	15.27	36,874.57	17.55	42,375.73
1	6.86	16,565.54	7.99	19,284.24	13.51	32,616.99
2	77.68	187,583.31	76.44	184,584.73	65.60	158,418.08
3	0.67	1627.56	0.31	737.71	3.34	8070.44

Table 4.3: Summary of Class Area Percentages for CART, RF, and SVM (2015)

Class	CART		RF		SVM	
	Percentage (%)	Area(km ²)	Percentage (%)	Area(km ²)	Percentage (%)	Area(km ²)
0	15.12	36,520.37	15.38	37,140.98	15.23	36,784.58
1	0.98	2,362.65	0.33	807.03	12.02	29,022.76
2	75.76	182,935.39	75.57	182,494.18	64.13	154,857.30
3	8.14	19,662.83	8.71	21,039.06	8.62	20,816.61

With these quantitative values for each class over the years, we observe that the classification of our focal class, the urban class (Class 3), remains consistent. The only exception was in 2011 when SVM exhibited a higher percentage compared to the other classifiers.

4.4.4 Evaluation metric.

In Chapter 3, we established specific metrics to assess the performance of our classifiers. Primarily, we will emphasize two metrics: the overall accuracy and the kappa coefficient. These metrics will be used to evaluate the effectiveness of our classifiers. Additionally, we will examine the confusion matrix, which provides valuable insights into the classification results. The confusion matrices for the years 2011 and 2015 are shown in Figure 4.9 and Figure 4.10, respectively.



Figure 4.9: Confusion Matrix classifier 2011



Figure 4.10: Confusion Matrix classifier 2015

Table 4.4: Classification Results of Overall accuracy and Kappa coefficient

Year	CART_OA	CART_Kappa	RF_OA	RF_Kappa	SVM_OA	SVM_Kappa
2011	0.923	0.893	0.923	0.892	0.871	0.821
2015	0.921	0.892	0.902	0.859	0.902	0.859

We observe that all the classifiers exhibit good accuracy, Kappa coefficients, and nearly diagonal confusion matrices. Among the classifiers, CART and RF demonstrate better performance compared to SVM. For our further analysis, we opt to utilize CART due to its simplicity and slightly superior performance when compared to RF as we can see in Table 4.4.

4.4.5 Comparison of CART with the MODIS LC_type4.

Table 4.5: Summary of Class Area Percentages for CART and MODIS (2011)

Class	CART		MODIS	
	%	Area (km ²)	%	Area (km ²)
0	14.79	35,704.84	14.26	34,435.14
1	6.86	16,565.54	0.05	112.70
2	77.68	187,583.31	85.50	206,469.01
3	0.67	1,627.56	0.19	463.77

Table 4.6: Summary of Class Area Percentages for CART and MODIS (2015)

Class	CART		MODIS	
	%	Area (km ²)	%	Area (km ²)
0	15.12	36,520.37	14.19	34,268.31
1	0.98	2,362.65	0.04	97.73
2	75.76	182,935.39	85.56	206,614.09
3	8.14	19,662.83	0.21	497.05

Since we lack access to the classifier used by MODIS, we must compare the results based on the area of each class within the classification maps. The total area of the MODIS classification map for both 2011 and 2015 is very similar to our own model's area ($241,480.628\text{km}^2$ and $241,477.178\text{km}^2$ for MODIS compared to $241,481.251\text{km}^2$ for our model). However, a key difference lies in the urban class change over time. While MODIS doesn't show significant urban growth over a five-year period, our CART model indicates substantial urban expansion, which aligns better with our understanding of real-world trends. This suggests that our model might be more accurate in capturing urban change. However, to ensure a less biased evaluation (validation purpose), we will still utilize the MODIS data for comparison.

4.5 Change Detection

Building upon the completed land cover classifications for 2011 and 2015, this section focuses on identifying areas of change during this five-year period. We achieve this by subtracting the 2011 classified map from the 2015 classified map on a pixel-by-pixel basis. Figure 4.11 visualizes the overall changes across all classes during the timeframe. Additionally, Figure 4.12 specifically highlights the changes within the urban class identified by our best-performing classifier (CART).

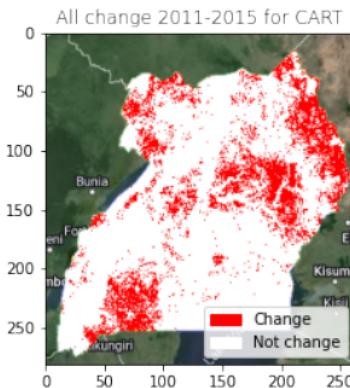


Figure 4.11: All change between 2011-2015

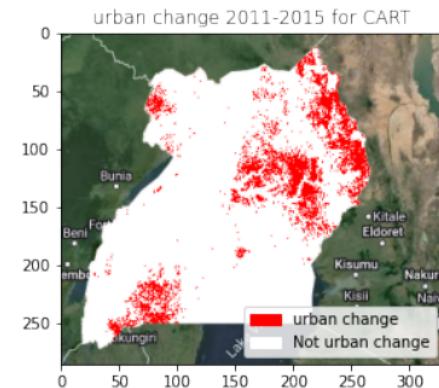


Figure 4.12: Urban Change between 2011-2015

To quantify the land cover changes between 2011 and 2015, we calculated a transition matrix presented in Table 4.7 and 4.8. This table records the area (in km^2) and percentage of each class in 2011 that transitioned to other classes in 2015.

- Class Remapping: For clarity in the table, the classes have been assigned new labels: Water = 1, Barren = 2, Vegetation = 3, Urban = 4
- Interpretation: Each cell in the table represents a specific transition between classes. For example:
 - "101": This cell represents the area (and percentage) of water in 2011 that remained as water in 2015 (no change).

- "104": This cell represents the area (and percentage) of water in 2011 that transitioned to an urban area in 2015.
- "204": This cell represents the area (and percentage) of barren land in 2011 that transitioned to an urban area in 2015.

Table 4.7: Summary of Transition land cover

Transition 1				Transition 2			
Index	Transition	(%)	Area	Index	Transition	(%)	Area
0	101	14.76	7.38×10^5	0	201	0.009220	4.61×10^2
1	102	0.014112	7.06×10^2	5	202	0.549470	2.75×10^4
2	103	0.083385	4.17×10^3	6	203	3.870839	1.93×10^5
3	104	0.005571	2.79×10^2	7	204	2.391276	1.20×10^5

Table 4.8: Summary of Transition land cover

Transition 3				Transition 4			
Index	Transition	(%)	Area	Index	Transition	(%)	Area
8	301	0.241428	1.207135×10^4	12	401	0.211694	1.058465×10^4
9	302	0.524187	2.620927×10^4	13	402	0.123802	6.190075×10^3
10	303	73.206353	3.660306×10^6	14	403	0.282221	1.411099×10^4
11	304	3.305529	1.652759×10^5	15	404	0.419372	2.096852×10^4

By analyzing this transition matrix, we can gain valuable insights into the dynamics of land cover change within the study area. Focusing on the urban class, we have the following transitions:

- Transition 104 suggests that 0.005% of water turns to urban area.
- Transition 204 means 2.39% of barren land turns to urban area.
- Transition 304 indicates that 3.30% of vegetation turns to urban area.
- Transition 404 indicates that 0.45% of urban area remains.

we summed the percentages of 104, 204, 304, and 404, which resulted in a total of 6.12%, close to the total 6.14% for urban class recorded in the CSV file where we computed the area of each district. However, this value is slightly different from the 8.14% percentage of the urban class in 2015 as shown in Table 4.6.

4.6 District-Level Urban Change and Satisfaction with Government Service Delivery

Applying CART, our best-performing classifier, we calculated the urban area for each of the 111 districts for both 2011 and 2015. These results were recorded in separate CSV files, which were then merged.

Finally, to assess urban growth, we calculated the difference in urban area for each district between the two years. Figure 4.13 illustrates the distribution of urban growth across the districts. The figure reveals several key observations:

- **Negative Growth:** A small number of districts experienced a negative growth rate, indicating a decrease in urban areas.
- **Moderate Growth:** The majority of districts exhibited urban growth, with many showing a difference of less than 500 km^2 .
- **High Growth:** A small number of districts experienced significant growth, exceeding 1000 km^2 of urban area increase.

For comparison, Figure 4.14 presents the distribution of urban growth obtained using the MODIS classifier. We can see that the majority of districts in the MODIS data exhibit growth of less than 5 km^2 and the density(line) shows that close to zero.

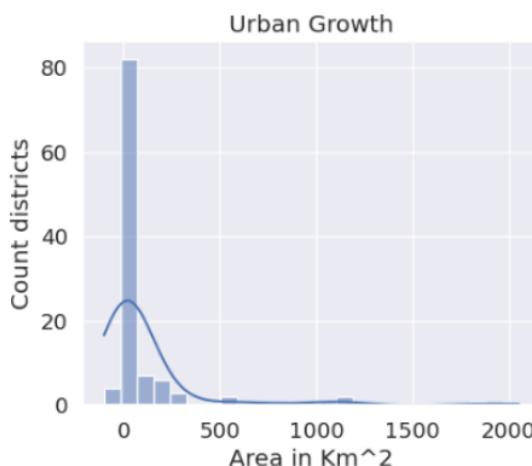


Figure 4.13: Dist Urban growth with CART

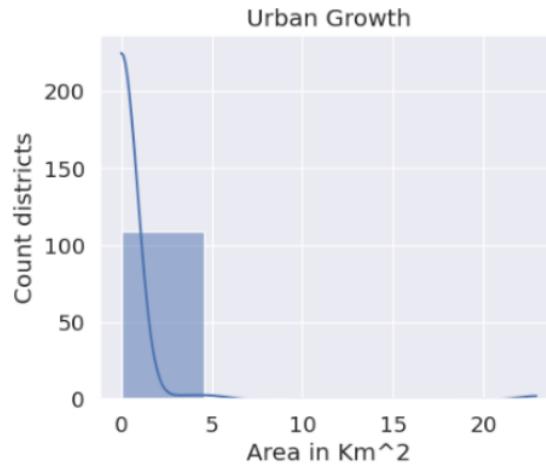


Figure 4.14: Dist Urban growth with MODIS

4.6.1 Investigating the Link Between Urbanization and Population Satisfaction.

In this section, we will explore the relationship between urban growth and the satisfaction of the population within specific districts. We will particularly focus on understanding whether positive urbanization (expansion that meets the needs of the population) is associated with higher satisfaction levels. Additionally, we will identify the needs of residents in districts that show a decrease in urban areas or no change. Finally, we will assess the opinion of residents on the efforts made by the local government to improve well-being in each district. We will achieve this through the following process:

1. **Urban Growth Patterns:** We will use data from both CART and MODIS classifiers to identify districts experiencing positive urban growth rates (as shown in Table 4.9, 6 out of 8). Districts with

growth indicated by both classifiers will be considered strong candidates for positive urbanization. We will also identify districts with negative growth, which seem to be identified by CART but show no change according to MODIS (as shown in Table 4.9, 6 out of 15). There are also districts identified as having no change by both classifiers, as shown in Table 4.9.

2. **Survey Data Integration:** We will use a survey dataset to assess the basic needs of the population across various dimensions, including food access, healthcare access, education, water and sanitation, electricity availability, roads, and security. We will also evaluate how the local government handles access to these needs.
3. **Comparative Analysis:** By comparing the urbanization patterns (positive, negative, or stagnant) with the satisfaction levels derived from the survey data, we can gain insights into the potential link between urbanization and well being of the population.

Table 4.9: Differences urban area 2015-2011 by District

District	CART	MODIS	District	CART	MODIS	District	CART	MODIS
Bundibugyo	0.0	0.0	Mbarara	224.48	0.50	Mubende	-2.30	0.0
Kabarole	0.0	0.0	Arua	93.16	0.25	Ntoroko	-74.77	0.00
Kisoro	0.0	0.0	Kampala	86.76	4.99	Kalangala	-60.95	0.00
Kyenjojo	0.0	0.0	Wakiso	60.12	22.89	Namayingo	-32.60	0.00
Mitooma	0.0	0.0	Mukono	19.42	3.39	Masindi	-12.46	0.00
Otuke	0.0	0.0	Mbale	16.30	0.74	Kyankwanzi	-10.90	0.00

4.6.2 Survey dataset.

This section focuses on the results for three specific districts, denoted as PSUs (Public Sector Undertakings) in survey data.

1. **District with high positive growth:** Based on the high growth rate identified by our CART model (which outperforms the MODIS classifier), we selected Mbarara district for further analysis (see Table 4.9). Figure 4.15 illustrates access to basic needs in Mbarara over time. We can observe that satisfaction with water and sanitation access has declined. While access to electricity and security has not improved, satisfaction with education and healthcare, which was initially high in 2011, has decreased by 2015. Interestingly, despite a decrease in markets, food access appears to have improved. Alongside, Figure 4.16 shows public perception of the Local Government's efforts to improve well-being. On average, residents felt the government made a fair effort.

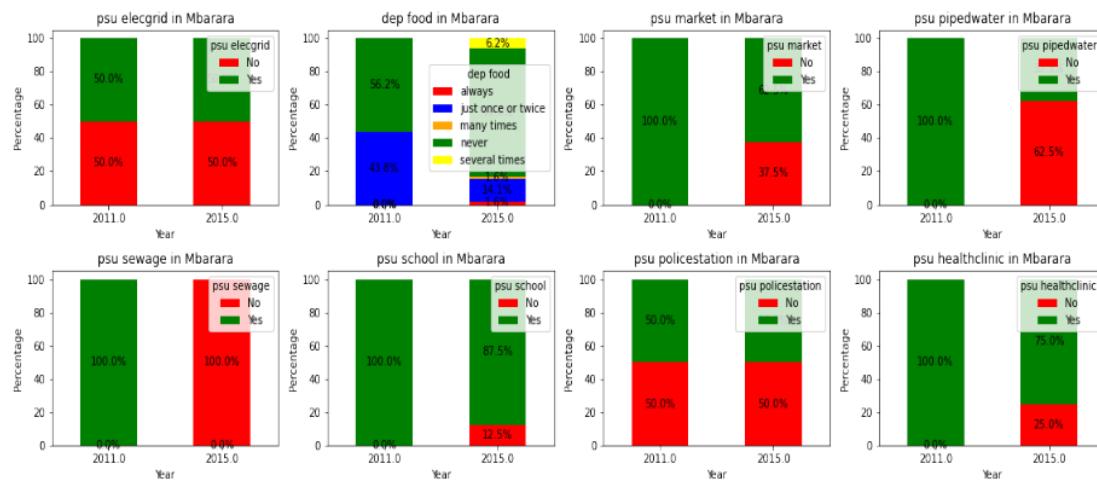


Figure 4.15: Access to Basic Needs in Mbarara, 2011 and 2015

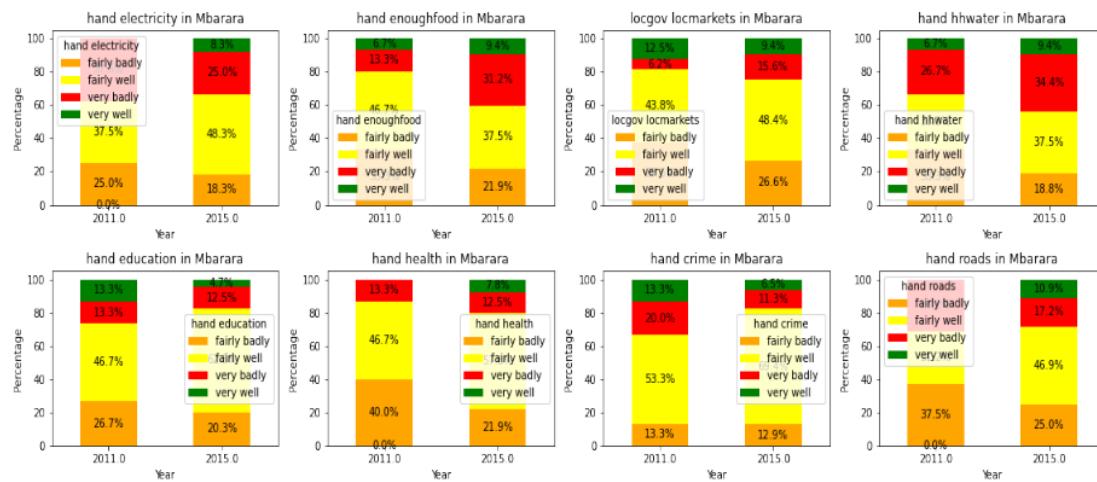


Figure 4.16: Trends in Public Opinion in Mbarara on Local Government Handling of Basic Needs, 2011-2015

2. **District with no change:** Table 4.9 shows that both the CART and MODIS classifiers identified no urban change in Kabarole district. Figure 4.17 shows that satisfaction with most basic needs in Kabarole was bad in 2015 (data not available for 2011) with exceptions for food and water where satisfaction appears to be around 50%. Regarding government efforts, Figure 4.18 likely illustrates public perception. While residents may perceive some fairly improvement in healthcare, security, markets, education, roads, and water access, satisfaction with the government's efforts to improve electricity and food availability is low.

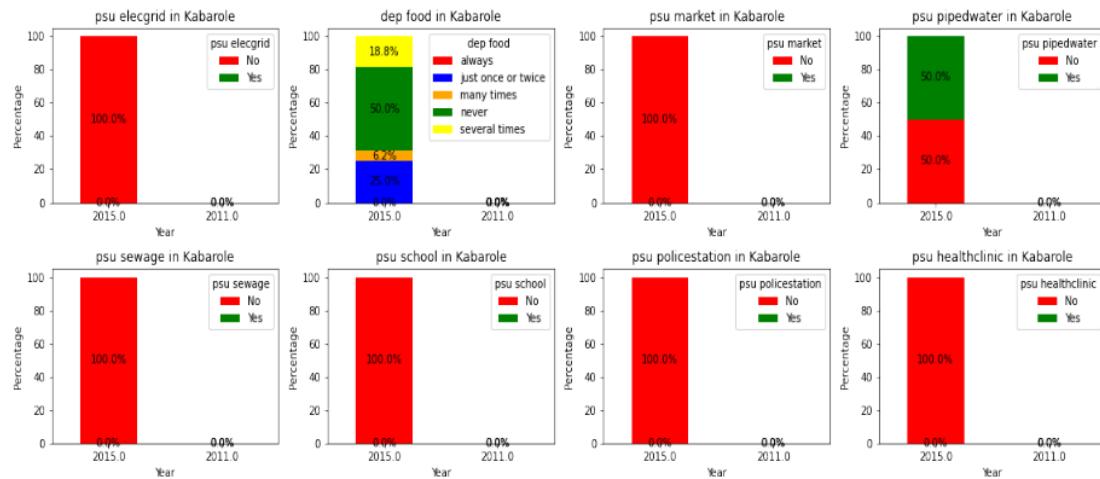


Figure 4.17: Access to Basic Needs in Kabarole, 2011 and 2015

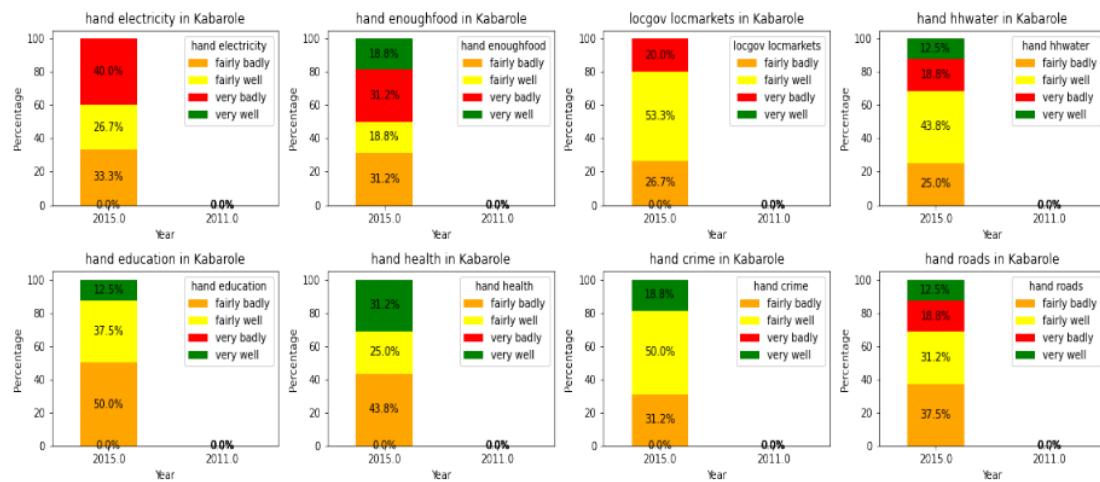


Figure 4.18: Trends in Public Opinion in Mbarara on Local Government Handling of Basic Needs, 2011-2015

3. District with negative growth : Table 4.9 highlights Mubende district, identified by the CART classifier as having a decreasing urban area and by the MODIS classifier as having no change. As shown in Figure 4.19, residents report high satisfaction with access to food and schools. Satisfaction with markets, security, and healthcare is moderate (around 40%) while access to electricity, water, and sanitation is poor. Additionally, Figure 4.20 shows that Mubende's residents report poor efforts by the government to handle access to these services, with an average of more than 50% rating the efforts as very bad or fairly bad for all services.

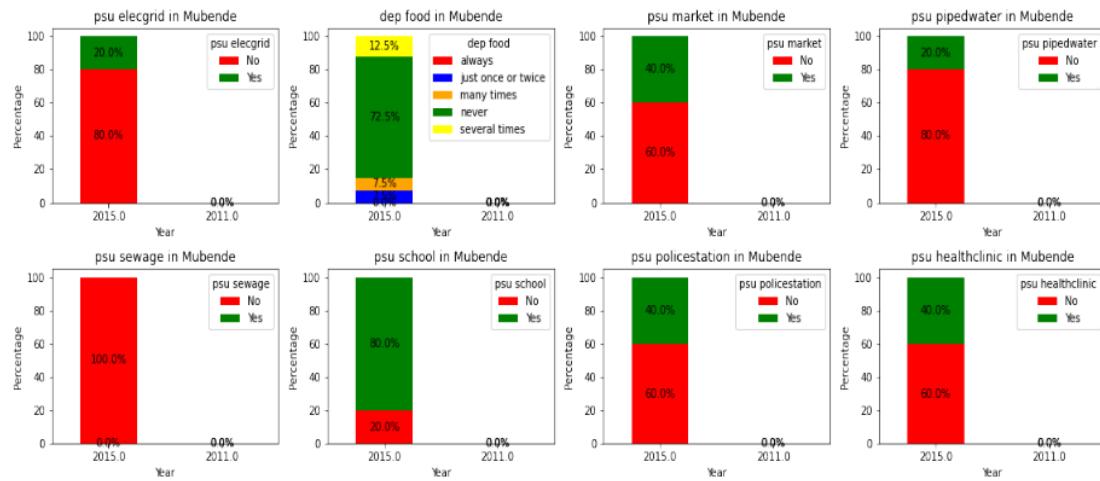


Figure 4.19: Access to Basic Needs in Mubende, 2011 and 2015

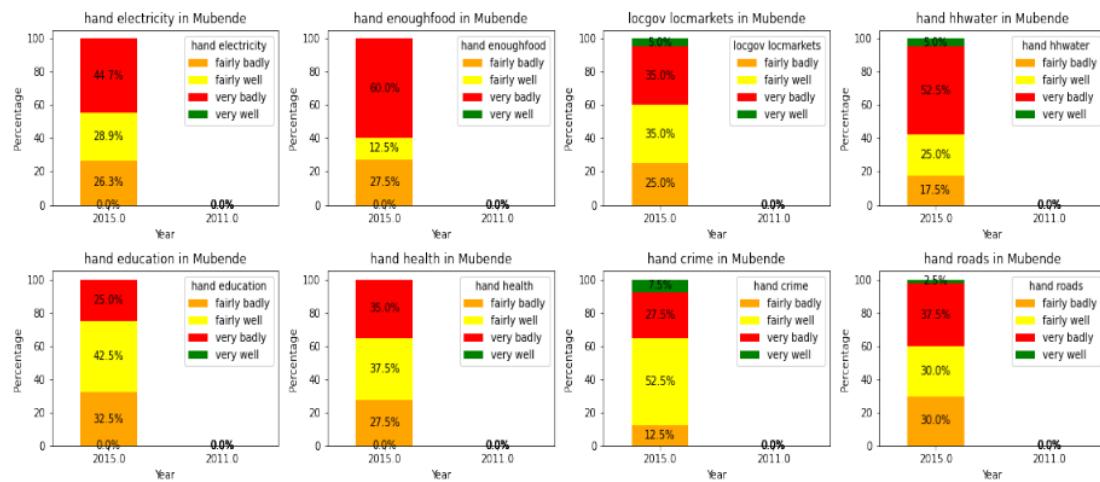


Figure 4.20: Trends in Public Opinion in Mubende on Local Government Handling of Basic Needs, 2011-2015

Following the analysis of the selected districts, we hypothesize that a link might exist between the level of urbanization and residents' satisfaction with essential service delivery by the government. To investigate this hypothesis, the next section will delve into a comparative analysis examining the relationship between urbanization trends and public satisfaction.

4.6.3 Comparative Analysis.

To analyze the potential correlation between urbanization and satisfaction with government services, we

need both variables on a comparable scale. Here is the process we followed to achieve that:

1. **Urbanization Percentage:** We calculated the urban percentage for each district in the provided CSV dataset (extracted during classification) for both 2011 and 2015.
2. **Percentage of Satisfaction:** Using the survey data for consistent districts (having data for both years), we created a new dataset. Each instance in this dataset represents a district with its corresponding percentage of satisfaction for each variable. These satisfaction percentages were calculated as:
 - Access to Services: "Yes" responses were used to represent satisfaction (we computed the percentage of "Yes" responses for each variable related to access to services).
 - Government Performance: "Very well" responses were used to represent performance satisfaction (percentage calculated similarly to access to services variables).
3. **Data Filtering:** We filtered the survey data to include only seven districts with consistent data across both years (2011 and 2015). Then, we extracted these specific districts from the urbanization dataset to ensure a valid comparison.
4. **Visualization:** Finally, for a chosen service and the seven consistent districts, we created a plot that visualizes the trends of both urbanization percentage and percentage of satisfaction for the two years. The Figures 4.21 and 4.22 display the output of those plots.

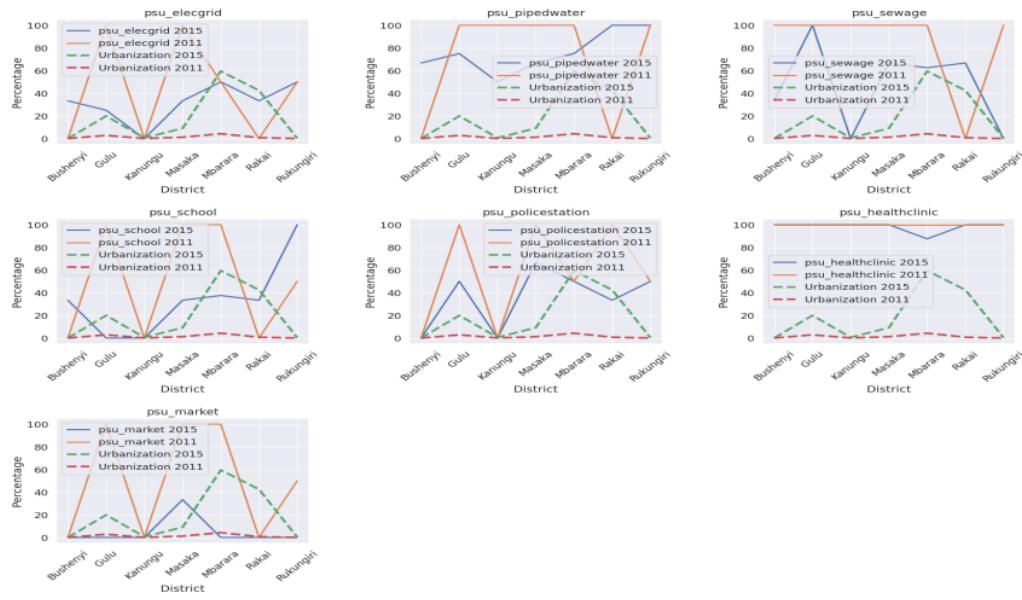


Figure 4.21: Comparative analysis: Urbanization Vs Satisfaction to Access to Service Delivery

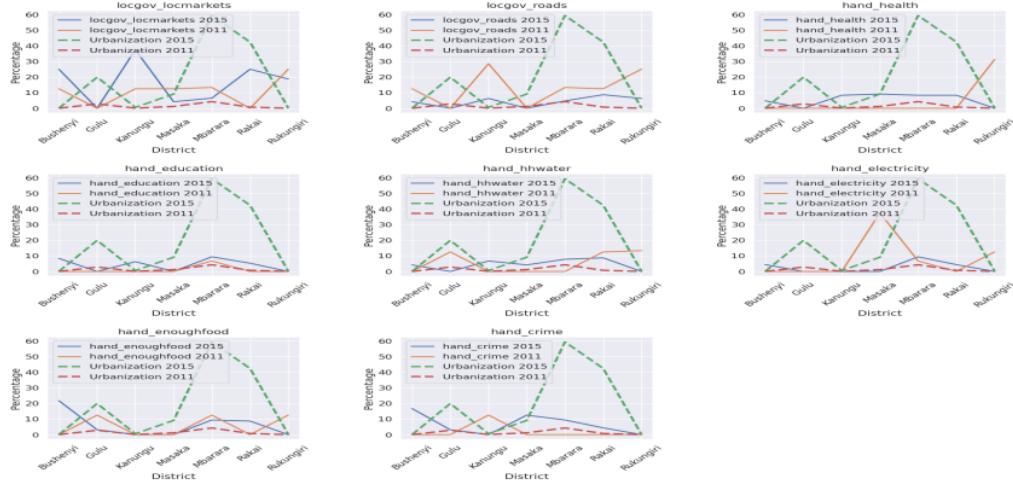


Figure 4.22: Comparative analysis: Urbanization Vs Satisfaction of the Improvement of the Quality of the Service Delivery

Figure 4.21 reveals a complex relationship between urbanization and satisfaction with service access. While some districts experience increases in urbanization over time, their satisfaction with service access actually declines. This is evident in the consistently lower satisfaction rates in 2015 compared to 2011 across all services.

- **Gulu:** A first example is Gulu, a district with significant urbanization growth. Despite a 100% satisfaction rating for several services (electricity, school, market, security) in 2011, these numbers drop dramatically in 2015 (22%, 0%, 0%, 45%, respectively).
- **Mbarara:** Similarly, Mbarara, the district with the highest urbanization rate, shows a decline in satisfaction for schools, markets, sewage, and healthcare. Ratings dropped from 100% in 2011 to 40%, 0%, 80%, and 60% in 2015, respectively.
- **Stagnant Urbanization:** Interestingly, districts with stagnant urbanization, like Kanungu, also exhibit consistently low satisfaction levels (0%) for electricity, security, markets, and schools in both 2011 and 2015.

These observations challenge our initial expectation. We anticipated that satisfaction with access to basic needs would improve alongside district growth. We also expected stagnant districts to show at least some improvement in satisfaction over time. Given these trends, questions arise regarding the efforts of local government in enhancing service quality. Figure 4.21 provides a concerning answer and present that the satisfaction with local government actions remains low across all services and districts, consistently below 20%. This suggests a potential need for a closer examination of local government initiatives and their impact on service delivery.

5. Discussions

This chapter presents our key findings, highlighting their relevance to the literature and how they contribute to a more comprehensive understanding of "Assessing the Impact of Urban Growth Dynamics on Quality and Access to Service Delivery in Uganda." We offer recommendations to policymakers for improving service quality, ultimately promoting the country's sustainable development. Finally, we report the challenges encountered throughout our research.

5.1 Our findings

5.1.1 Comparison with Existing Classification. Chapter 4 compares our classification results with the existing classification based on MODIS satellite imagery. We focused on urban areas and found that our best model, CART, yielded a more realistic and reasonable growth estimate. Here are the findings:

- **CART Model (Ours):** Urban area increased from 0.67% in 2011 to 8.14% in 2015 (6.14% increase aligns with the total urbanization percentage in the exported CSV file).
- **MODIS Classification:** Urban area increased 0.19% in 2011 to 0.21% in 2015.

The significantly higher growth rate detected by our model suggests a more accurate representation of urban expansion(capture the actual urban expansion that occurred during the study period.)

5.1.2 Validation and Performance. Furthermore, the total area of Uganda calculated using our classified map (summing the area of all classes across all districts) is $241,592.35 Km^2$, which closely matches the country's actual area of $241,551 Km^2$. This close correspondence provides further validation of our classification's accuracy. Our study also reinforces the effectiveness of supervised machine learning techniques in urban area classification, as evidenced by the high Kappa coefficient of 0.89 achieved by our model. This strong performance aligns with findings from the existing literature.

5.1.3 Implications for District-Level Analysis. The accurate district-level urban growth information obtained through our classification allows us to assess how urbanization impacts public service needs with greater confidence. By identifying urbanized districts, we can effectively evaluate and address the evolving public service requirements in these areas.

5.2 Recommendations to Local Government

This section outlines recommendations for local governments based on the findings of our study. Our goal is to assist in identifying areas with the most urgent public needs, going beyond simply relying on relative satisfaction of survey data and helping local governments with service improvement. So here are some recommendations for policymakers based on our findings:

1. Prioritizing Education Access

- **Problem:** The study identified districts like Bushenyi, Gulu, Kanungu, Masake, Mbarara, Rakai, and Rukungiri as having significant challenges in accessing education. The limited

access to education in these areas makes it harder for development to happen, as it affects innovation and technological progress.

- **Recommendation:** We recommend that local governments in these districts:

- Conduct a deeper needs assessment to understand the specific barriers to education access.
- Invest in building new schools or expanding existing ones, Ensuring services remain accessible and budget-friendly for residents (low price).
- Explore partnerships with non-governmental organizations (NGOs) or private institutions to increase educational opportunities.
- Implement low-fee or scholarship programs to increase educational opportunities, particularly in districts with stagnant or slow growth.

2. Ensuring Food Security :

- **Problem:** The study suggests potential food insecurity concerns in the previously mentioned districts.

- **recommendation:** We recommend local governments in these districts:

- Farmer Support Programs: Develop and implement programs to support farmers, such as providing seeds, fertilizers, and training on improved agricultural practices.
- Local Market Development: Facilitate the creation of local markets where farmers can sell their products at affordable prices to residents.

3. Enhancing Electricity Access

- **Probem:** Electricity is sometimes considered a development factor and a potential indicator of urbanization. However, the availability of electricity decreased over the years despite the urbanization as we have observed in Gulu and Masaka.

- **Recommandation:** Local governments can:

- Distribution Optimization: Analyze current electricity distribution patterns and identify areas with underutilization or inefficient allocation. Then develop plans for optimizing electricity distribution to reach underserved areas within districts identified as experiencing rapid urbanization or low initial electrification. This involves rerouting existing lines or exploring renewable energy solutions.
- Strategic Development: Plan for future electricity infrastructure development, ensuring efficient and equitable distribution in line with projected urbanization trends.

5.3 Challenges

The big challenges encountered in the work are :

- The availability of satellite imagery: ideally, we wanted to use more advanced methods(CNN, Unets, Imagenet, ResNet) that rely on analyzing satellite images. Unfortunately, there were not enough high-quality images available, so we had to use a different approach(supervised machine Learning).

- We used a free online platform (GEE) to do our analysis. While it is great, it can take a long time to process information(train the model).
- Our study looked at how urban areas are growing, but it is important to remember that population growth also plays a role in urban growth.

6. Conclusion

The goal of our study was to investigate the link between urban growth and service delivery in Uganda between 2011 and 2015. We reached this objective by conducting an urban analysis using both satellite images and survey data. Firstly, we have employed supervised machine learning with Landsat 7 satellite data to map urban changes across Uganda between 2011 and 2015, analyzing five-year periods. Using the best-performing classifier (CART) with a high Kappa score of 0.89 for both years (compared to Random Forest at 0.89 in 2011 and 0.86 in 2015 and Support Vector Machines at 0.82 in 2011 and 0.85 in 2015), we focused on urban classes to quantify urban growth. Notably, the urban area expanded significantly from 0.67% in 2011 to 6.14% in 2015. This detailed analysis allowed us to assess urban growth at the district level. We observed rapid growth rates in Mbarara (59%), Kampala (45%), and Rakai (42%) over five years, while districts like Kanungu, Bushenyi, and Rukungiri saw no growth. Secondly, we have leveraged this district-level urbanization data to analyze resident satisfaction through surveys conducted at the beginning and end of the study period. Finally, our findings revealed a concerning mismatch between urbanization and access/quality of basic government services. We identified critical resident needs such as food, education, and electricity, and based on these findings, formulated recommendations for local governments to address these service gaps.

However, to enhance this work, further studies could explore:

- **Incorporating population growth:** This would provide a more comprehensive understanding of residents' needs and enable more targeted recommendations for addressing them.
- **Longitudinal analysis:** Assessing urbanization and service delivery satisfaction over 20-30 years would reveal trends and strengthen the understanding of their relationship. High-resolution satellite imagery for Uganda would be necessary.
- **Object-level deep learning classification:** Shifting from pixel-level (our approach) to object-level classification using deep learning could provide a more detailed analysis of urban growth.
- **Predictive modeling:** Developing a model to predict district-level urban growth rates would facilitate resource management for sustainable development.

Appendix A. Some additional data

The link to the script in GEE and Python Code:

- https://code.earthengine.google.com/?scriptPath=users%2Freginepiyou%2FLULC_classification%3AUganda_discritlevel_ALL_classifier
- https://code.earthengine.google.com/?scriptPath=users%2Freginepiyou%2FLULC_classification%3AAnalysis_districtLevel
- https://code.earthengine.google.com/?scriptPath=users%2Freginepiyou%2FLULC_classification%3AUgandaRFLCClassifier
- https://drive.google.com/drive/folders/1iO4azzN4A9p5SGFNBYq891-Lo1EgI7UI?usp=drive_link

Acknowledgements

I would like to express my deepest gratitude to all those who have supported me in the completion of this thesis.

First and foremost, I extend my sincere thanks to God, the almighty, for granting me the strength, courage, health, inspiration, and intellectual capacity that I benefit from throughout this journey.

My sincere appreciation is also extended to my supervisory team, including Professor Nara Monkam, Dr. Maonei Mangwanya, Dr. Adedayo, and Mr. Rockfeller. Their guidance and support were invaluable as I embarked on this engaging real-world project, culminating in my Master's studies at AIMS.

Furthermore, I would like to acknowledge the invaluable contributions of my tutor, Mr. Domini Jocema Moutouo Leko, whose advice and insights proved highly beneficial.

I am also grateful to the entire AIMS Cameroon administration, particularly Professor Mama FOPOUAG-NIGNI, the Center President, Dr. Daniel Duviol Tcheutia, the Academic Director, and the exceptional staff of AIMS-Cameroon. Their dedication and support significantly enhanced my academic experience.

Finally, I extend my heartfelt gratitude to my family for their unwavering love and support, and to the AIMS community for the warm and welcoming environment they encouraged.

References

- [1] Information on uganda. <https://www.itu.int/en/ITU-D/Regional-Presence/Africa/Documents/Information> Accessed on: 23 April 2024.
- [2] TorchGeo glossary. Accessed on April 27, 2024.
- [3] Un-habitat uganda.
- [4] United nations sustainable development goals. <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>.
- [5] A better quality of life for all in an urbanizing world: Promoting socially and environmentally sustainable human settlements development and the achievement of adequate shelter for all. https://unhabitat.org/sites/default/files/2023/07/uganda_country_brief_final_en_1.pdf, 2023. Accessed on: 23 April 2024.
- [6] African Development Bank. Africa dominates ranking of top twenty countries with strongest economic growth in the world in 2024, according to a macro-economic report by the african development bank. *African Development Bank Journal*, 2024.
- [7] Islam Atef, Wael Ahmed, and Ramadan H. Abdel-Maguid. Modelling of land use land cover changes using machine learning and gis techniques: a case study in el-fayoum governorate, egypt. *Environmental Monitoring and Assessment*, 195(6):637, 2023.
- [8] L Ballanti, L Blesius, E Hines, and B Kruse. Tree species classification using hyperspectral imagery: A comparison of two classifiers. *remote sens.* 2016, 8, 445, 2016.
- [9] Wele Gedara Chaminda Bandara and Vishal M. Patel. A transformer-based siamese network for change detection, 2022.
- [10] R. Berwick and Village Idiot. *An Idiot's Guide to Support Vector Machines (SVMs)*. 2001.
- [11] Basudeb Bhatta. Analysis of urban growth and sprawl from remote sensing data. 01 2010.
- [12] Gouri Bhunia, Uday Chatterjee, Anil Kashyap, and Pravat Shit. *Land Reclamation and Restoration Strategies for Sustainable Development (Elsevier)*. 11 2021.
- [13] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.
- [14] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995. © 1995 Kluwer Academic Publishers, Boston. Manufactured in The Netherlands. Editor: Lorenza Saitta.
- [15] Pall Oskar Gislason, Jon Atli Benediktsson, and Johannes R. Sveinsson. Random forests for land cover classification. *Pattern Recognition Letters*, 27(4):294–300, 2006. Pattern Recognition in Remote Sensing (PRRS 2004).
- [16] Pooja Gulati, Amita Sharma, and Manish Gupta. Theoretical study of decision tree algorithms to identify pivotal factors for performance improvement: A review. *International Journal of Computer Applications*, 141:19–25, 05 2016.

- [17] Moshhood Hambali, Tinuke Oladele, Adewole S., Arun Kumar, and Wei Gao. Feature selection and computational optimization in high-dimensional microarray cancer datasets via infogain-modified bat algorithm. *Multimedia Tools and Applications*, 81, 08 2022.
- [18] S R Jiao, J Song, and B Liu. A review of decision tree classification algorithms for continuous variables. *Journal of Physics: Conference Series*, 1651(1):012083, nov 2020.
- [19] Lazhar Khelifi and Max Mignotte. Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis. *IEEE Access*, 8:126385–126400, 2020.
- [20] Jan Knorn, Andreas Rabe, Volker C Radeloff, Tobias Kuemmerle, Jacek Kozak, and Patrick Hostert. Land cover mapping of large areas using chain classification of neighboring landsat satellite images. *Remote sensing of environment*, 113(5):957–964, 2009.
- [21] John N. Kwiringira, Ronald Kabumbuli, Henry Zakumumpa, and et al. Re-conceptualizing sustainable urban sanitation in uganda: Why the roots of 'slumification' must be dealt with. *BMC Public Health*, 21(992), 2021.
- [22] Alfredo MarÃn, Luisa I. MartÃnez-Merino, Justo Puerto, and Antonio M. RodrÃguez-ChÃa. The soft-margin support vector machine with ordered weighted average. *Knowledge-Based Systems*, 237:107705, 2022.
- [23] Farid Melgani and Lorenzo Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8):1778–1790, 2004.
- [24] Olli Nevalainen, Eija Honkavaara, Sakari Tuominen, Niko Viljanen, Teemu Hakala, Xiaowei Yu, Juha Hyppä, Heikki Saari, Ilkka Pöölönen, Nilton N Imai, et al. Individual tree detection and classification with uav-based photogrammetric point clouds and hyperspectral imaging. *Remote sensing*, 9(3):185, 2017.
- [25] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.
- [26] Bruno Emmanuel Ondo Nkoa and Jacques Simon Song. Urbanization and inequalities in africa: A study based on disaggregated indices. *Journal of Regional and Urban Economics*, 2019(3):447–484, 2019.
- [27] Y Ouma, B Nkwa, D Moalafhi, P Odirile, B Parida, G Anderson, and J Qi. Comparison of machine learning classifiers for multitemporal and multisensor mapping of urban lulc features. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:681–689, 2022.
- [28] Yuguo Qian, Weiqi Zhou, Jingli Yan, Weifeng Li, and Lijian Han. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing*, 7(1):153–168, 2014.
- [29] Oliver Sefrin, Felix M. Riese, and Sina Keller. Deep learning for land cover change detection. *Remote Sensing*, 13(1), 2021.
- [30] Ayesha Shafique, Guo Cao, Zia Khan, Muhammad Asad, and Muhammad Aslam. Deep learning-based change detection in remote sensing images: A review. *Remote Sensing*, 14(4), 2022.

- [31] Ayush Singhal and Gaurav Shukla. Decision trees: Classification and regression trees (cart). February 2023.
- [32] Yu-Yang Song and Yan Lu. Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*, 27(2):130–135, 2015.
- [33] Richard Stren. Urban service delivery in africa and the role of international assistance. *Development Policy Review*, 32:s19–s37, 07 2014.
- [34] United Nations African Institute for Economic Development and Planning (IDEP) and United Nations Economic Commission for Africa (CEA). *Urbanization and Inclusive Economic Growth in Africa*. IDEP, Rue du 18 juin, BP 3186 CP 18524, Dakar - SENEGAL.
- [35] Boipuso Nkxae Phillipon Odirile Ditiro Moalafhi Yashon O. Ouma, Amantle Keitsile and Jiaguo Qi. Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach. *European Journal of Remote Sensing*, 56(1):2173659, 2023.
- [36] Boipuso Nkxae Phillipon Odirile Ditiro Moalafhi Yashon O. Ouma, Amantle Keitsile and Jiaguo Qi. Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach. *European Journal of Remote Sensing*, 56(1):2173659, 2023.
- [37] Xinying Yu, Shie-Yui Liong, and Vladan Babovic. Ec-svm approach for real-time hydrologic forecasting. *Journal of Hydroinformatics*, 6(3):209–223, 2004.
- [38] Xin Zhang, Liangxiu Han, Lianghao Han, and Liang Zhu. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? *Remote Sensing*, 12(3), 2020.
- [39] Shengyu Zhao, Kaiwen Tu, Shutong Ye, Hao Tang, Yaocong Hu, and Chao Xie. Land use and land cover classification meets deep learning: A review. *Sensors*, 23(21), 2023.
- [40] Zhuo Zheng, Ailong Ma, Liangpei Zhang, and Yanfei Zhong. Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15173–15182, 2021.

Regine_Essay_afterCorrection.pdf

ORIGINALITY REPORT



PRIMARY SOURCES

- | | | |
|----------|---|-----------|
| 1 | researchhub.buan.ac.bw
Internet Source | 2% |
| 2 | Lazhar Khelifi, Max Mignotte. "Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review and Meta-Analysis", IEEE Access, 2020
Publication | 1% |
| 3 | Yashon O. Ouma, Amantle Keitsile, Boipuso Nkwaе, Phillimon Odirile, Ditiro Moalafhi, Jiaguo Qi. "Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach", European Journal of Remote Sensing, 2023
Publication | 1% |
| 4 | www.ncbi.nlm.nih.gov
Internet Source | 1% |
| 5 | Submitted to The African Institute for Mathematical Sciences
Student Paper | 1% |
-

Exclude quotes	On	Exclude assignment template	On
Exclude bibliography	On	Exclude matches	< 1%

Regine_Essay_afterCorrection.pdf

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51
