UNIVERSITY OF NAIROBI

Faculty of Engineering

MAPPING CHANGE DETECTION OF INFORMAL SETTLEMENTS USING REMOTE SENSING AND GIS: CASE STUDY-KAWANGWARE, NAIROBI (2000-2020)

By

ANN NJERI MUTITU

F56/38249/2020

A project report submitted to the Department of Geospatial and Space Technology in partial fulfillment of the requirements for the award of the degree of:

Masters of Science in Geographic Information Systems

August, 2022

DECLARATION OF ORIGINALITY

Name of student: ANN NJERI MUTITU
Registration: F56/38249/2020
College: ARCHITECTURE AND ENGINEERING
Faculty/School/Institute: ENGINEERING
Department: GEOSPATIAL ENGINEERING
Course Name: PROJECT
Title of work**: MAPPING CHANGE DETECTION OF INFORMAL SETTLEMENTS USING REMOTE SENSING AND GIS: CASE STUDY- KAWANGWARE, NAIROBI (2000-2020)**

1 I understand what plagiarism is and I'm aware of the university policy in this regard

2 I declare that this research proposal is my original work and has not been submitted elsewhere for examination, award of a degree or publication. Where other works or my own work has been used, this has properly been acknowledged and referenced in accordance with the University of Nairobi's requirements

3 I have not sought or used the services of any professional agencies to produce this work

4 I have not allowed, and shall not allow anyone to copy my work with the intention of passing it off as his/her work

5 I understand that any false claim in respect of this work shall result in disciplinary action in accordance with University of Nairobi anti-plagiarism policy

Signature: Ann Date: 23rd August 2022

DECLARATION

I, **Ann Njeri Mutitu**, hereby declare that this project is my original work. To the best of my knowledge, the work presented here has not been presented for a project in any other university.

Name of student: ANN NJERI MUTITU Signature: Date: 23rd August 2022

This project has been submitted for examination with my approval as university supervisor.

Name of supervisor: PROF F.N. KARANJA Signature: Date: 23.08.2022

TURN IT IN REPORT SUMMARY

MAPPING CHANGE DETECTION OF INFORMAL SETTLEMENTS USING REMOTE SENSING AND GIS:CASE STUDY-KAWANGWARE,NAIROBI (2000-2020)

ORIGINALITY REPORT

14%	5%	10%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

STUDENT:

NAME: ANN NJERI MUTITU

SIGNATURE:

DATE: 23rd August 2022

SUPERVISOR:

NAME: PROF F.N. KARANJA

SIGNATURE:

DATE:23rd Augsut 2022

DEDICATION

To all informal settlements' dwellers and everyone who contributed in my academic voyage.

ACKNOWLEDGMENT

First, I am grateful to the Almighty God for His grace has been sufficient in my academic journey.

I wish to express my sincere thanks to University of Nairobi for giving me the opportunity and scholarship to enroll and complete my Masters in Geographic Information Systems. I extend appreciation note to the all the staff and of the Department of Geospatial Engineering for any support and guidance.

I wish to express my gratitude to my supervisor and role model Prof F.N Karanja for providing me with guidance and freedom necessary to complete this project.

I wish to extend a thank you note to my daughters (Wendo and Hera), parents, sisters and brothers and for their encouragement, support, prayers and understanding

ABSTRACT

The first stage in planning and developing cities in underdeveloped countries is to detect informal settlements and their changes. High spatial resolution satellite imageries are popular in the carrying out the changes in spatial extent of these settlements. However, these imageries are expensive to purchase, and as a result, it may not be affordable to all, particularly in countries with a large number of informal settlements. Using Kawangware, Nairobi as a case study, this project aimed to investigate the phenomenon of informal settlements in Kawangware and its development from 2000 to 2020 by means of remote sensing and GIS. Landsat, which is a Medium Resolution satellite imagery, was used to map this phenomenon. Random Forest classification method was the applied in this project after constructing the training set using some approaches. In the first approach, visual interpretation was done using Google Map imagery and composites in order to select training samples. Open Street Map (OSM) building blocks and street layers were used as the method for training during the second round of classification. The results showed a high-speed growth of the built-up class. Consequently, the informal settlements increased in area, especially on the account of the vegetation and bare ground classes. The results obtained showed that in 2000, the total area representing the informal settlement was 98.46ha and this increased to 99.08ha. This area increased to 143.94ha in 2020, an increase of 4.3% from 2000. Conclusions drawn were that the increase in informal settlements can be attributed to the proximity of Kawangware to formal settlements such as Lavington, Westlands and also Nairobi CBD where the dwellers work so that commuting easily and cheaply since most of them are casual workers. The classification accuracy was 93.63% for the year 2020,86.64% for the year 2010 and 88.71% for the year 2000.The proposed methodology presents the application of freely available remote sensing data to map change detection in an informal settlement of an extent not larger than a city. It was recommended that further studies on the OSM validation should be conducted to improve the reliability of this data source.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

ABBREVIATIONS

ANN-Artificial Neural Networks

KNN- k-Nearest Neighbors

KNBS- Kenya National Bureau of Statistics

MR- Medium Resolution

OSM- Open Street map

RF-Random Forest

UN-United Nations

VGI-Volunteered Geographic Information

VHR-Very High Resolution

Chapter 1 : **INTRODUCTION**

1.1 Background

In order to detect changes in the informal settlements, it is important to understand the meaning of the term informal settlements and its origin. The world is facing rapid growth of urban areas and this may be as a result of migration from rural to urban areas, which is also termed as urban sprawl. The world has about 55% of its entire population currently living in urban areas, which might increase to 60% by the year 2030. The increase will be experienced more in the developing counties, which by then will have the highest percentage of the population growth which is projected to reach 95% by the year 2030 (United Nations, 2019)

The demand for housing on available land is therefore increasing as a result of urbanization. This trend has resulted to land use/land cover change, land transactions that are informal, and consequently spatial expansion of these informal settlements (Selod & Durand, 2009). Informal settlements sometimes referred to as slums have emerged as illegal, irregularly constructed and built-up areas with narrow road networks with dead ends and building sub-divisions within the urban area borders that have little or no access to basic urban infrastructure and services.

Kenya’s Urban population will be almost at 50% of the total country’s population by the year 2050 .Kenya's Urban population has grown by 4.1% for the years 2015 to 2020 and 3.06-3.97% for the years leading to the year 2050 (UN-Habitat, 2016).Currently the urban population, estimated at about 15 million people, represents 31% of the total population which is about 47 million according to the 2019 census done by Kenya National Bureau of statistics (KNBS, 2019).

Nairobi was founded along the Kenyan Ugandan Railway formally known as Mombasa Kampala railway in 1901. Like almost all colonies, it grew and reached a population of about 118,000 in 1946. The spatial pattern which are seen today reflects the patterns that were done by the British colonists as opposed to the traditional African one (Obudho, 1984).Nairobi can be considered as one of the most modern cities of Africa due to how well areas are planned. Its population has grown quickly, like most cities of the continent, to reach 2 million people in less than 50 years due to immigration and natural increase.

Larger areas of the city where access to land was easier,have been occupied by low-income migrants. The immigrants had no option but to organize themselves in an informal way to cope with the lack of infrastructure. Everything became informal - employment, housing, transportation and industrial activities. In these areas of city, the population growth and informality has resulted to environmental degradation, poor housing, overcrowding, poor sanitation and limited access to water.

Informal settlements are no doubt a reality and their fast growth is a great challenge that authorities in charge of providing essential basic services are facing or will have to in the future. This reality has appeared when the global economy has is rapidly declining, leaving most of the developing countries with fewer resources available to cope with the population growth, unable to address basic infrastructures needs, like water, waste disposal, education, energy and health care facilities.

Remote sensing data and techniques have been used to monitor changes in urban areas. In recent years temporal satellite imagery has been of use to estimate spatial changes in urban environment (Wang et al., 2020).Remote sensing has also gained popularity in detecting and mapping changes in informal settlements which is the first step toward improving one of the world's most vulnerable groups

1.2 Problem statement

Just like most of Africa Countries, Kenya is experiencing rapid urbanization or urban growth, which has led to growth of informal settlements. Rapid growth of informal settlements is a leading factor for slow development in developing countries. Other factors such as Poverty, Unemployment, political instability, basic health care etc. has been seen to pose a great challenge especially to a sustainable development.

The Kenya's Urban population living in informal settlement was about 54.7% in the year 2009 and it was estimated that this increased to 56% by the year 2014 (UN-Habitat, 2016).People living here often face tenure insecurity ,live in sub-standard houses, lack adequate infrastructure and have to deal with environmental challenges.

Kenya Vision 2030 implementation program stated some reforms in the housing sector, which includes a program on informal settlement improvement. Mapping changes in informal settlement will serve as a major decision-making tool to achieve the above program. However, most studies on change detection have been done using Very high/High resolution imagery, which are expensive to acquire, and resources may not be available in developing countries. Open data sources when combined with machine algorithms can be used to detect changes in informal settlements.

1.3 Objectives

1.3.1 Main Objective

The overall objective of the study was to map changes in growth of informal settlement in Kawangware area between 2000-2020 using remote sensing data and techniques.

1.3.2 Specific Objectives

The specific objectives of the study were namely to: -

i Review change factors of informal settlements.

ii Identify suitable temporal satellite imagery for mapping changes in Kawangware informal settlement.

iii Map informal settlement for Kawangware for the years 2000-2020.

iv Analyze changes in the informal settlements for Kawangware for the years 2000-2020.

1.4 Research Questions

i How have informal settlements changed over time?

ii What satellite imagery is suitable for mapping Kawangware informal settlements?

iii How can remote sensing and GIS be used to map informal settlements of Kawangware for the years 2000-2020?

iv What is the rate of change of informal settlements in Kawangware area between the years 2000 to 2020?

1.5 Justification of the Study

The problem stated shows that there is an urgent need for planning and fast decision making in order for the spatial planners and government officials to maintain at least control of the city growth. This could be impossible without availability of updated information about various aspects in the urban areas, e.g. informal settlements data

Kenya vision 2030 has set up a program on informal settlements. The program's main aim is to improve the lives of at least 2.5 million people living and working in the slum. Mapping change detection in Kawangware informal settlements are will help the local administration and all stakeholders know the rate of change of informal settlements in the area and Kenya in general and therefore making proper decisions on how to improve the living conditions of the people.

1.5 Scope and Limitations of the Study

The study area was limited to the Kawangware in Nairobi County. It is located along Naivasha road as the main spine. This area is a typical informal settlements area with characteristics similar to any informal settlement. The project does not attempt to resolve the general housing problem in this area. It however attempts to map changes of the informal settlements over the study period.

1.6 Organization of the Report

The report is structured as follows. Chapter one contains background information about the study problem as well as objectives and research questions. Chapter two discusses available literature, making comparison and drawing conclusions. The methodology used in data collection and analysis is discussed in chapter three. Detailed data analysis, interpretation and presentations of findings are outlined in chapter four. Chapter five contains summary findings, conclusion and recommendations.

Chapter 2 : LITERATURE REVIEW

2.1 Urban Informality and Kenyan Context

Informal settlements mostly occur in underdeveloped and newly industrialized countries. Many definitions that are associated with informal settlements exist, squatters, shantytowns or slums are commonly used to refer to these types of settlements. The United Nations (UN Statistics, 2022) defines these settlements as:

“1. Groups of housing that have been built on land that those occupying have no legal claim to occupy;

2. Areas where the settlements are unplanned and not in agreement with the current planning regulations.”

Informal settlements could be because of rural-urban migration, which is known as inserting. Thus, a form of settlement is commonly found in peri-urban areas and it occurs mostly on vacant land (Dovey & King, 2011) .Setting which is another type of informal settlement consists of villages which are surrounded by built-up areas formed during the urbanization process. This means that people live on unregistered land. These villages lack much needed basic services and infrastructure and are considered self-governing.

One of the world’s largest slums is found in the Asia-Pacific region. Asia has the largest population and contains both types of informal settlements, with 80% of these settlements ‘dwellers found the eastern and Southern parts of the continent (Sweeting, 2017). Nairobi, which is the capital city, is home to a number of squatters. These include but not limited to Kibera Slums in Langata constituency, Kawangare slums in Dagoretti Riruta, Mukuru wa Njenga, Dandora slums etc. Kawangware settlement is considered a unique informal settlement in that its informal areas extend to areas with planned settlements in the city. The settlement is also characterized by rapid urban sprawl and unplanned built up areas.

2.2 Characteristics of Urban Informality

Informal settlements have buildings with no pre-thought pattern. Residential houses are mainly single-stores due to poverty of residents and considering the type of materials used in the building and uncommon construction procedure. Houses in these settlements have generally been formed without any pre-thought geometry and based on the financial ability of each family in occupying the land and under the influence of topography and natural bed.

Various reasons have been put forward to explain the emergence and the rapid growth of slums in third world countries. For instance, research has it that slums mostly occur in less valuable or marginal land such as steep slopes, river banks, dumping areas, along transportation networks, near industrial areas abandoned or unexploited plots, and market places, and in wetlands or low-lying areas (Blight et al., 1999).

The density and pattern of buildings are key factors to identifying informal settlements. A settlement can be characterized by both the building density unit (number of buildings per unit area) and building coverage (the ratio of the total area of built-up area to the administrative region area).

Roads in informal settlements are narrow and short and have very regular sections. The streets have poor accessibility, with a high proportion of dead ends. Road networks are often irregular with changeable widths, length and types of pavements. Buildings blocks have no clear textural features, are loosely distributed, and have branch shaped pattern. This trend poses a great challenge in effective provision of basic infrastructure and social services as well as accumulation of economic activities.

2.3 Remote Sensing Data Used for Detecting Informal Settlements

Informal settlements has been defined as having one of the following characteristics: a) lack of tenure security; b) deprivation or lack of basic infrastructure ;and c) housing that do not meet the current physical planning regulations .The first criterion refers to the aspect of land regulations while the other two consider the physical features which can be detected using remote sensing data (UN-Habitat, 2013).

Several efforts have been made to detect these kinds of settlements by using remote sensing technologies and data based on their physical features (Hofmann & Bekkarnayeva, 2017). However, most of these studies that have been done, VHR/HR imagery have been used. One of the disadvantages of these kind of imageries is that they are expensive to acquire and might not be affordable especially in least developed countries. Despite this high cost misclassification of formal and informal can still occur. This limitation has driven interest in the use open data sources to detect and map changes in informal settlements.

Open data can be classified into freely available statistical and census data that is mostly collected by the government or non-profit organizations and open spatial data which includes medium resolution (MR) satellite imageries and open street maps (OSM) that is remotely sensed using modern remote sensor and transmitted to ground stations for processing.

(Mahabir & Agouris, 2018) Mixed OSM and MR data statistical data and census data to identify suitable indicators such as street patterns, and level of vegetation to detect informal settlements in Kenya. The study has taken into account both physical and social-economic features and this requires high volumes of data (Verma & Jana, 2019) compared the performance of transfer learning model for detection of informal settlements on VHR and MR at the city level and the results revealed a higher performance on MR and finally suggested further studies on machine learning algorithms to improve the classification accuracy.

2.4 Methods for Developing the Training set for Classification

The method, which is commonly used, for developing the training set sample is called the in-situ data collection. This type of method is very time consuming as well as expensive for very large areas. There have been various options for training pixels but one of the common one is by using available data like informal settlement maps for this case.

There being a new era with increasing available data provides more options to replace the data sets that have been used previously as input data for training pixels. Developing the training sets based visual checks on higher spatial resolution imagery available on Bing (https://www.bing.com/maps) and Google earth (https://www.google.com/earth/) and is one of the uses of open data to train pixels which are to be used for classification. This however highly depends on expertise and is prone to human error.

Another approach, which is gaining popularity, is the used of Volunteered Geographic Information (VGI) for training pixels (Smith et al.,2007). One of the most popular VGI is the OSM. This is a collaborative project used to develop open spatial data worldwide and it provides layers in editable formats, such as, shapefiles, for analysis and manipulation. Depending on one's area of study different data layers such as building block, street layers; just to mention a few; might be available on the OSM platform, such as rivers, road networks, buildings, etc.

OSM accuracy has been questioned as it offers an editable data source however, case studies have shown that the accuracy of OSM is comparable to that of authoritative sources (Jokar Arsanjani et al., 2013). OSM data generally suffers from inaccuracy based on location depending on the data layers therefore it is advised that data validation should be done in order for this method to be relied up on.

2.5 Machine Learning Algorithms Pixel Based Classification

There exists different methods for the classification of satellite images. They vary from parametric (based on data distribution assumptions) which are like maximum likelihood to those which are non-parametric algorithms which are decision tree (DT) (Friedl & Brodley, 1997),artificial neural networks (ANN), k-Nearest Neighbors (KNN) as well as ensemble classifiers, such as Random Forest (RF) which are simply a combination of classifiers (Breiman, 1996).

The non-parametric algorithms are known to be accurate especially when dealing with large-scale mapping. This maybe be because assumption on data is not made and underlying functions (Rodriguez-Galiano & Chica-Olmo, 2012).Studies to compare the various machine learning algorithms have been done. It has been seen that RF is more robust since it has fewer user-defined parameters and requires very minimal supervision compared to the other algorithms. RF is also less sensitive to the training data and processes the data fast (Rodriguez-Galiano & Chica-Olmo, 2012).

In comparison to the other algorithms, ANN has sensitivity to hidden codes number, Kernel parameters usually have an effect on the accuracy in support vector machine and for KNN,the accuracy is determined by the ideal value of K that can be difficult for one to set (Noi & Jones, 2017). RF is seen to deliver more variety of classes than traditional classification methods. Differentiating similar land-use classes is very important in this case of informal settlement change detection. Distinguishing similar land-use/land cover classes, such as informal and formal built-up areas can be less manageable than the other, which are very noticeable like (e.g. built-up areas and green spaces or bare land and water bodies).

2.6 Random Forest Classification Algorithm

Random Forest classification is a model that is based on decision trees with bootstrap techniques and improved bagging. This classification model contains a large number of trees just a forest. These trees are grown from training pixels, which are randomly selected to do the classification. It uses a technique known as train-test-split. One sample is used to do the classification while the other sample is used to test the model or estimate the classification errors.

Each sample produces a tree and the number of trees is grown from a number of bootstraps. The parameters which should be defined in this classification are two: the ntree (these are the number of trees that are grown) and the mtry (variables to which are to be split at each node) and this is the square root of the input variables. After the model is built, each results of the bootstrap then votes for the most common class and the final output is a classification result. The more the number of variables means a more complex algorithm and better correlation of the trees (Breiman, 1996).

2.7 Land Use /Land Cover Change Detection Techniques

The first thing to consider when carrying out a land use -land cover change detection is registration of the satellite imagery so that the pixels overlaid are of the same location, there are many methods to carry out change detection. They are; multi-date composite imagery method; comparison of imagery; classified images comparison; combination of the classified images; radar classification.

2.7.1 Multi-date Composite Imagery Method

Combining satellite imageries of different periods into a composite helps to detect changes from variation in gray tone color or hue in the resultant composite. Image enhancement through overlay, rationing and vegetation index can be done to enhance areas containing changes suppress areas of no change. The resultant image is then classified. (Brazeau & Fung, 1989) described enhancement procedures as follows:

- Overlay involves combination of two images of various colors.
- Differencing is subtracting two images of two spectral bands pixel by pixel.
- Image rationing which involves diving bands to get the rationed imageries.
- Vegetation index, which is used to compare and contrast soil background and vegetation.

2.7.2 Image Comparison

This involves comparing satellite images of different dates. When the signals variate between these images then it means that, there is change in the land cover for that area (Murphy, 1989).

2.7.3 Classified Image Comparison

(Samahiji & Chaube, 1987) did a study on spatial expansion of urban areas in Algeria for the towns of Bilda an oasis of Lahghoutt using traditional maps and satellite images. He stated that the city of Bilda had expanded to about 10055 ha since the year 1962.A different study was done in the northern eastern parts of Cairo using data from multiple sources, Spot-HRV, KVR-1000 and Landsat-TM with resolution of 20 m, 30 m and 5 m respectively. They were then merged to provide a higher information content. Land use changes were mapped from 1945-1993 and it offered an indicator for urban growth.

2.7.4 Combination of Classified Images

This method was used by (Guan & Chen, 2021) in Thailand. The main objective was to develop land use-land cover change methodology. Two geocoded Landsat-TM Imagery, which were acquired during the dry and wet seasons, were classified. The resultant was an April image with 14 classes and September image containing of 11 classes. They were then combined to give a combination of 154 possible gray level image. The resultant imagery gave a land cover change with respect to the type of terrain.

2.8 Case Studies

2.8.1 Informal Settlements Change detection Using Multi-Temporal Aerial Photographs Case Study: Voi, SE Kenya

(Hursikainen & Petri, 2004) Carried out a change detection of informal settlements using Multi-Temporal Aerial photos in a town called Voi, coast region in Kenya for the years 1985, 1993, 2004. These photographs went through pre-processing where orthorectification, correcting for brightness variation was done, they were then mosaic using EncoMOSAIC and Erdas Imagine. The Mosaics were segmented and then classified to built-up and non-built-up areas. Change detection was done by comparing the three masks for the different years using Arcview software. From the study, the area had changed by the year 2004. A total of 1914 houses had been built from 1993-2004.

2.8.2 Using Open and Freely Available Data for Detection of the Pattern and Structure of Informal settlements

(Assarkhaniki & Sabri, 2021) Used open data and freely available data to detect pattern and structure of informal settlement in Jakarta, Indonesia. Landsat 8 (2020) imagery was classified to for the detection of these settlements. Training was carried out in two approaches, one approach used available survey data for 2015 and visual checks on high-resolution google map. In the second approach, OSM data was used to train the pixels.

The formal settlements were seen to be located in central and North Jakarta while the informal ones were seen to be located in the West and South of Jakarta. During second round of classification, the informal settlements were found mostly in the Eastern and Western area on Jakarta. Results validation was done using RF machine learning algorithm. Enhancement was noticed particularly in the built-up class with an increase in accuracy from 0.88 to 0.94 and an increase in precision from 0.58 to 0.79

2.8.3 Change Detection of Informal Settlements using Object-Based Techniques

(Hofmann & Bekkarnayeva, 2017) did a study on detection of informal settlements for an area in Cape Town, South Africa from 2000 to 2015 using object-based techniques. Very high satellites were used in this study. The imageries were sharpened using Hyper Spectral Color Space Resolution Merge with Nearest Neighbor resampling and then co-registered to the 2000 Ikonos reference scenes using about 15-20 GCPs per image and a third polynomial transformation with Root Mean Square less than 0.5.

The scenes were classified, objects segmented into base level and top level. The base level contained roofs that were red in color, bright objects, vegetation and dark objects and the top-level features described the base level properties and objects shapes and color were used. This is known as object-based structure and texture analysis. The change was detected by linking corresponding objects and then calculating the changes in the values of the object features. The author suggested that improving classification methods would help accurately classify slums and monitoring individual informal settlements in order to map the rate of change.

2.8.4 Detection of Urban Features from IKONOS Data Using an Object-Oriented Approach

(Hofmann, 2001) showed how high spatial resolution IKONOS data can be used to detect component pan sharpening after which image objects were generated using eCognition approach on an arbitrary number of scale levels and took into account shape and color homogeneity. After describing network's objects' semantic relationships in terms of neighborhood relationships, the classified objects were aggregated to the semantic groups.

Image segmentation was done which led to better outlining of objects like roads and houses. Classification was done based on spectral properties. Settlements areas were split into dense, medium, settlement areas with gardens, new and bright settlements. Informal settlements areas were detected satisfyingly and in cases were visual interpretation was hard, ground-truthing was recommended to give evidence of the land use.

Chapter 3 : **METHODOLOGY**

3.1 Study Area

Kawangware is located about 15 km west of Nairobi city. Kawangware area has a population of about 300,000 people (KNBS, 2019) with an estimated 80,000 of the population dwelling in informal settlements. The settlement is unique in that its informal areas extends to areas with planned settlement as well as peri-urban areas. The settlement consists of a number of wards, which spatially overlap with planned areas. The wards represented in the study area were Kabiro, Gatina, Riruta and Kawangware. Figure 3.1 shows the study area.

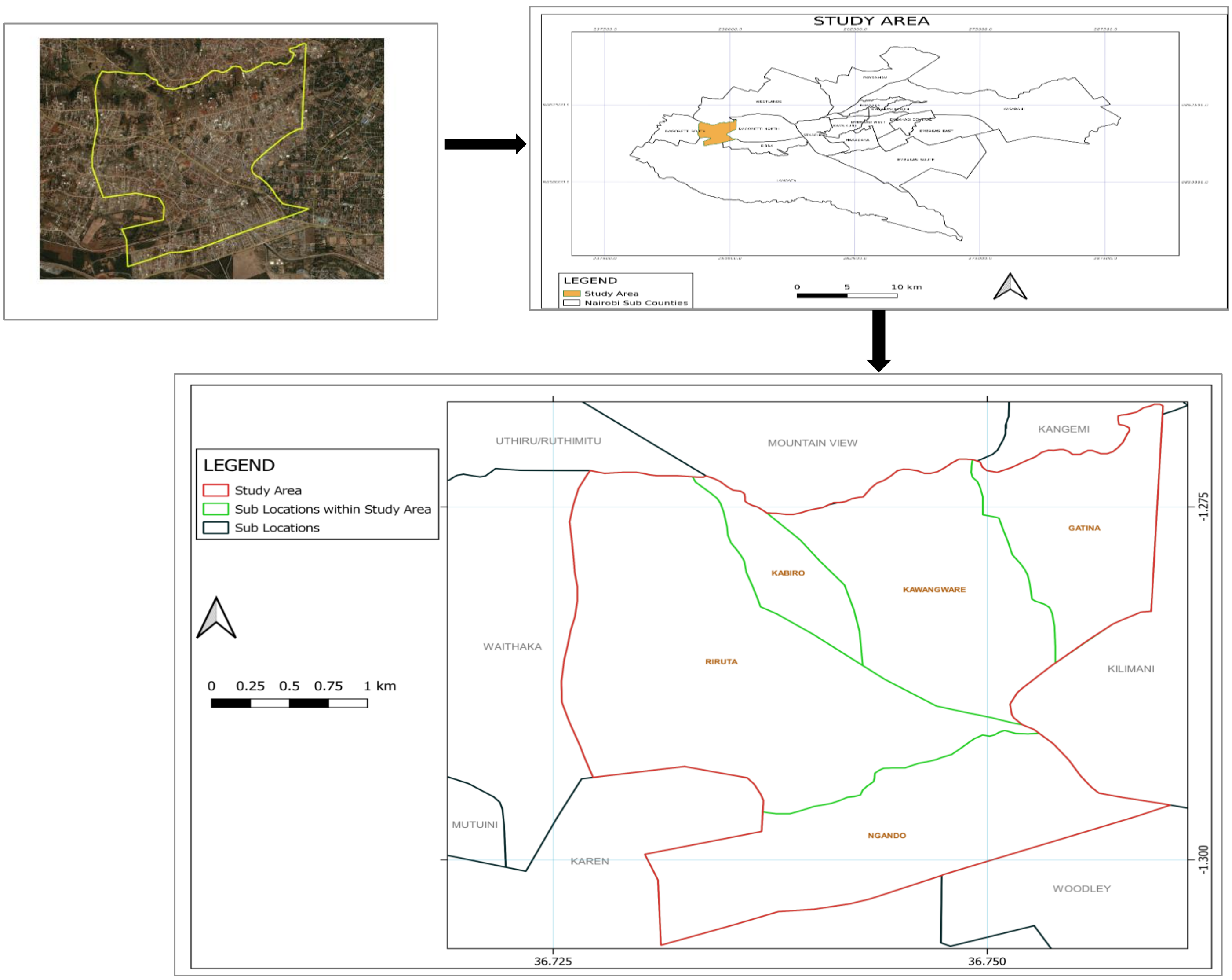

Figure 3.1:Study Area

3.2 Data and Methodology

3.2.1 Data

This project's objective was to map the changes in spatial extent of informal settlements in the Kawangware, Nairobi. It made use of the available remote sensing data and capabilities to determine the impact of this expansion. To achieve the Landsat imageries of the study area were acquired for the years 2000, 2010 and 2020 and were made to pass through classification. A google earth image of the study area was also used to help in the supervised classification based on the knowledge of the area. Buildings and road networks data for the study area were also sourced from open street map. Table 3.1 gives a description of the data used in this study.

Table 3.1:Data and Sources

Data	**Source**	**Description Year/Period**
Buildings, Road Networks	Open street Map	Vector format
Administrative Boundary	Kenya open data	Vector format
Landsat 7 (ETM+ sensor)	USGS Earth explorer	30 m spatial resolution, Image bands 1−7, Panchromatic band 15m, bands 10–11,100 m spatial resolution Path-167 Row -68year (2020)
Landsat 7 (ETM+ sensor)	USGS Earth explorer	Image bands 1 −7,30 m spatial resolution, Panchromatic band 15m, Path-167 Row -68 year (2000)
Landsat 5 (TM sensor)	USGS Earth explorer	Image bands 1 −7,30 m spatial, Band 120m resolution, Path-167 Row -68 Year (2010)
Google earth imagery		

3.2.2 Methodology

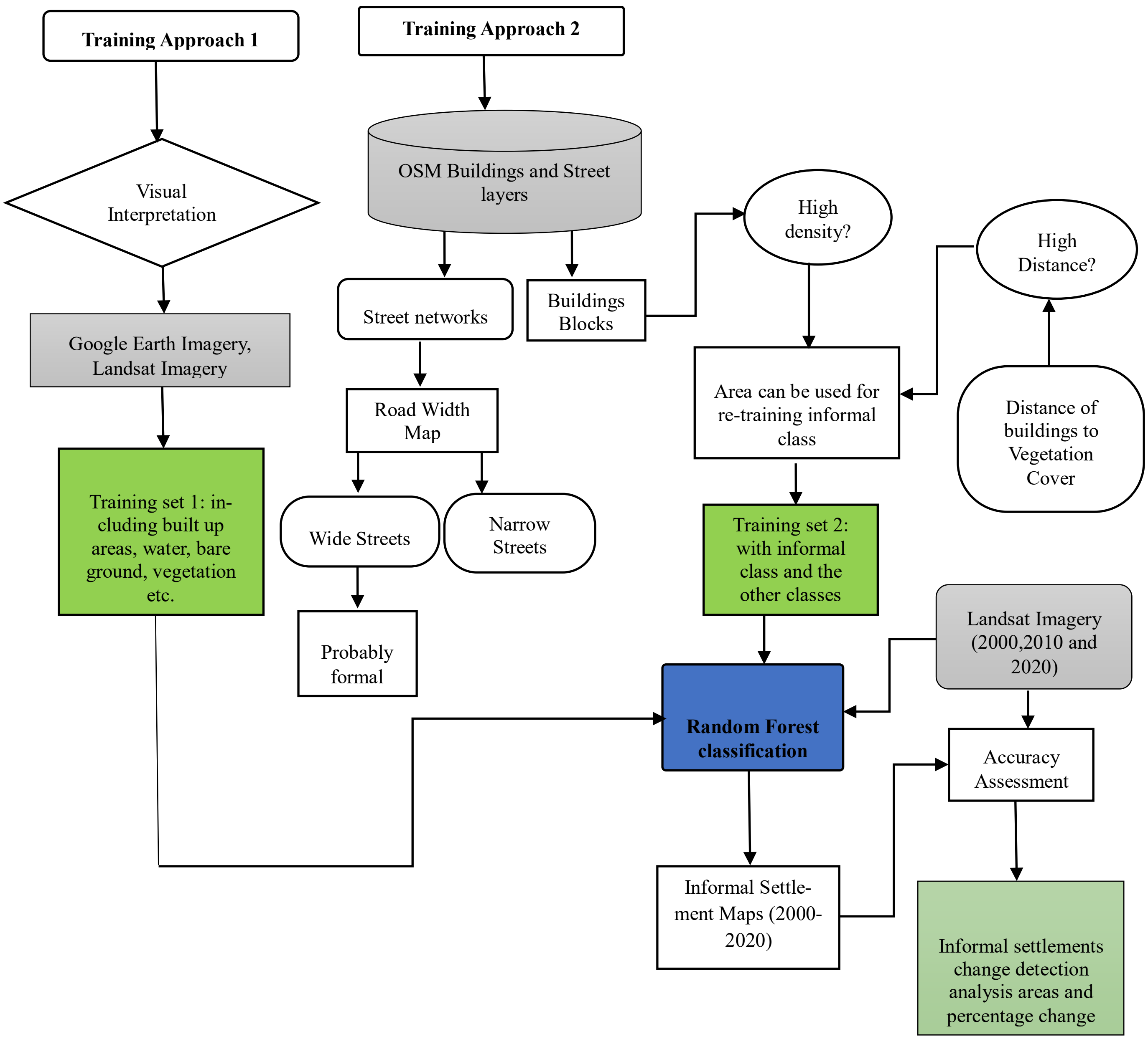

Figure 3.2: Flowchart of the Methodology

An intensive geographical study of the study area was done through literature review and this included basic description such as road networks, proximity to the CBD, and proximity to the formal settlements. In order to meet the objectives of the study Satellite imagery (Landsat 7 ETM+ sensor, Landsat 5 TM sensor) was acquired for the periods 2000, 2010 and 2020 respectively and free of clouds. OSM data on building blocks and street layers was also obtained.

Classification was done on Landsat images applying the Random Forest Algorithm. This algorithm requires produce highly accurate results and requires very few defined parameters (Duro et al., 2012). Training of sample points was done using two approaches. The first one involved the use of a google earth imagery. Selecting sample points for the built-up areas and other classes of land cover each containing the same number of points was done using visual interpretation google earth imagery. It should be noted that the higher the number of classes to be used, the deeper the depth of the decision trees which consequently has a higher accuracy (Breiman, 1996).

The second approach used OSM data to help retrain the built-up areas. Due to the aspect of uncertainty of the OSM data accuracy, attribute of data such as the roof types on building blocks can be difficult to tell. The use of OSM in this project was to extract the pattern of the buildings and the width of roads in the study area. The selected sample points were double-checked using Google Map Imagery. Table 3.2 shows a description of the resultant classes.

Table 3.2:Land Use Land Cover Classes

Classes	**Description**
Formal Settlements	Standard size buildings with vegetation in between and ample open space
Informal Settlements	Substandard settlements with very limited open space, few or no vegetation in between, and having narrow streets.
Bare Ground	Areas which are undeveloped and have no dominant vegetation cover
Tree Cover	Areas that are covered by trees
Roads	Street networks and have asphalt
Water	Areas that are covered by water, such as ponds and pools
Grassland	Areas that are covered by low vegetation such as grass and low bushes
Scattered vegetation	Areas covered partly by vegetation and partly by bare soil

3.2.2.1 Informal Settlements Detection

From literature review on informal settlements, detection was done using the characteristics that they are associated with. This includes lack of or very low vegetation, high building density per area, narrow streets, and concentration on major public transportation networks and water bodies such as rivers

Physical features of informal settlements may be attributed to the streets in the area. Therefore, since ownership of cars is not a very common occurrence among the dwellers of informal settlements, streets are narrow with dead ends and mainly suitable for motorcycle or pedestrians (Hidayati et al., 2020) showed that streets which have a width of about 2 meters or less are an indication of slums.

Although narrow streets are an indication of informal settlements' setup, these roads can also be found in formal settlements. Another indication has been used to detect these settlements, which is the urban fabric of informal settlements. The geometrical types of housing and alleyways can analyze this. This refers to the regular streets pattern as well as consistent placement of buildings (Suhartini & Jones, 2020). Figure 3.2 illustrates the flowchart of the methodology.

Chapter 4 : RESULTS AND DISCUSSION

4.1 Land use/ land cover Classification map

First round of classification containing built-up areas in general and other classes. Figure 4.1 shows a map of land use /land cover of the study area containing 7 classes for the year 2020. The built-up area is seen to cover most of the study area compared to other classes. At this stage the informal and informal settlements classes have not been trained.

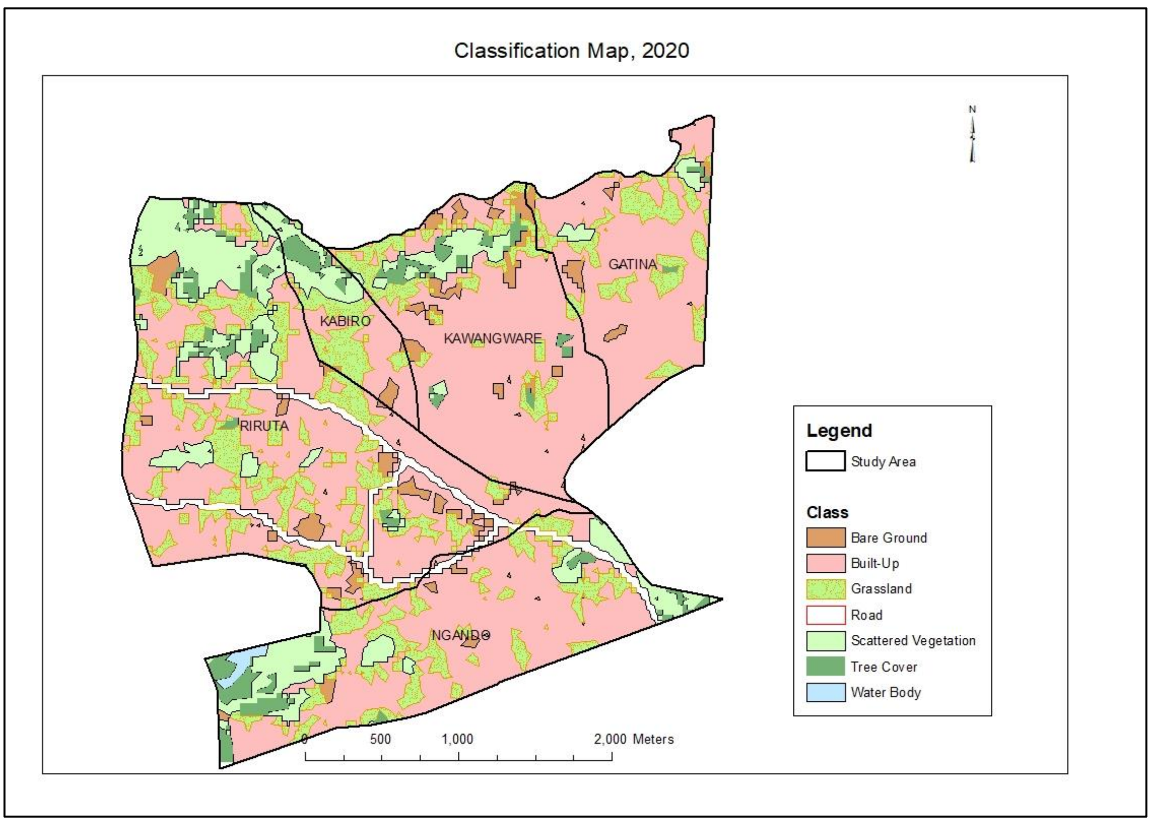

Figure 4.1:Land use/Land cover Map 2020

4.2 Informal Settlements Maps based on Road Width, Vegetation cover and Buildings Density

Figure 4.2 shows areas covered by informal settlements based on the width of the street networks passing through them. It was noted that Kawangware and Gatina wards have narrow roads within them and this could be an indication of existence of informal settlements. This data on road helped train the informal settlements class.

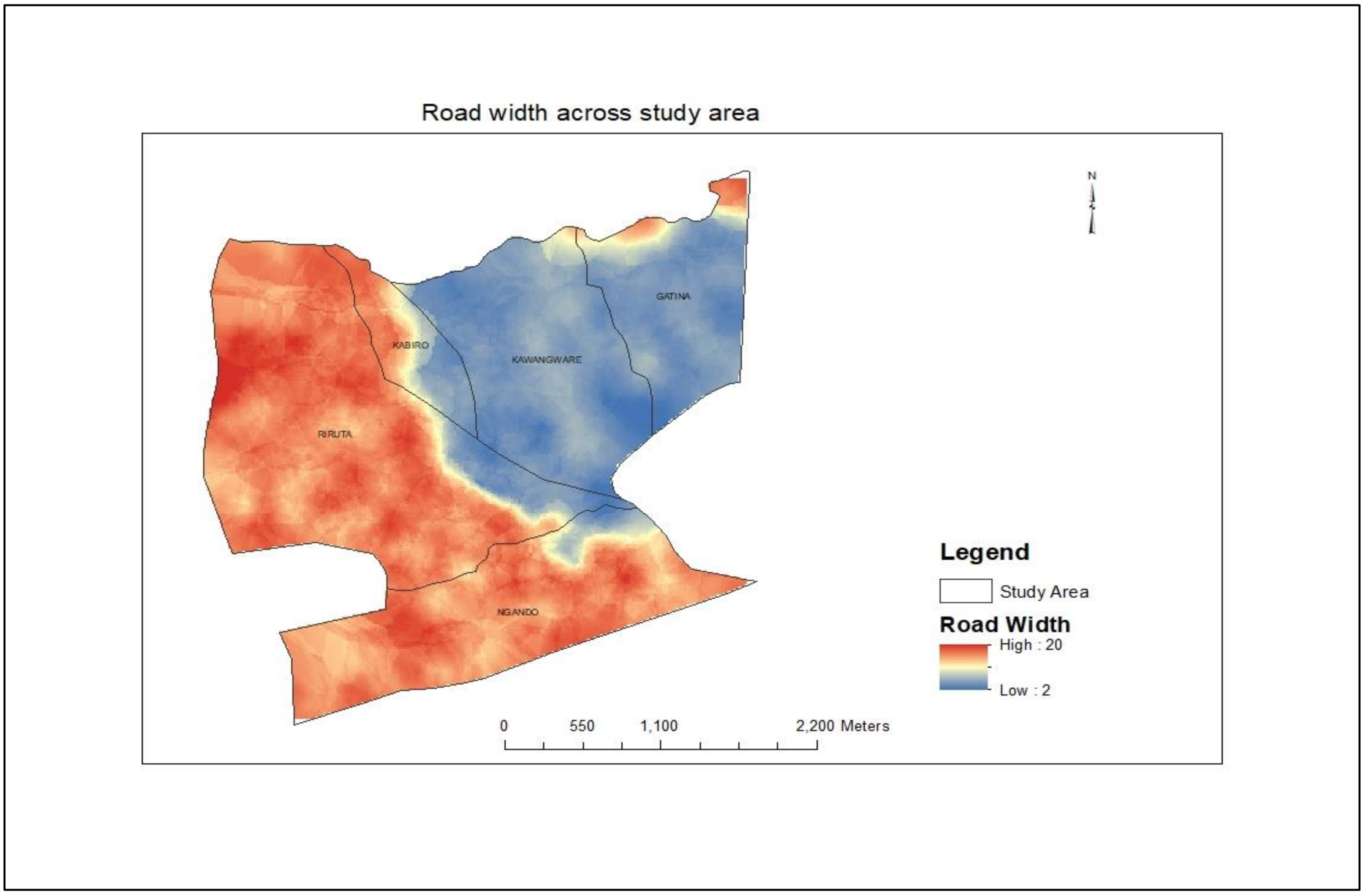

Figure 4.2:Road Width Map

Informal settlements are characterized by unplanned and overcrowded buildings, which have no regular pattern. Figure 4.3 is a map showing the density of buildings in the study area. Kawangware ward has notably high density of buildings, which indicates a possibility of informal settlements in the area.

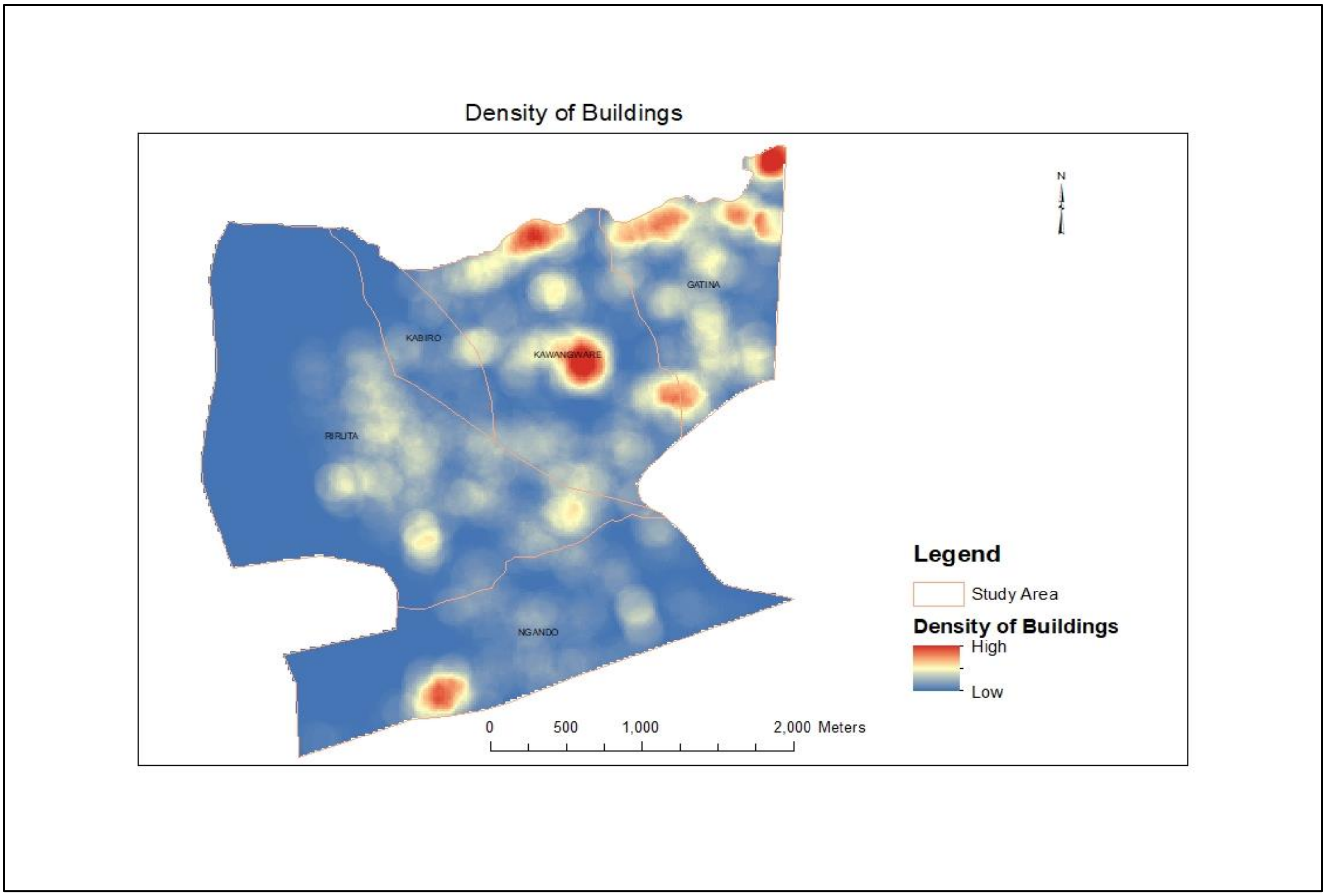

Figure 4.3: Buildings Density Map

Due to overcrowding in squatters, people tend to clear nearly all the vegetation cover to pave way for the houses. Vegetation class can therefore be used to help retrain the informal settlements. Figure 4.4 shows the distance that a settlement is to the nearest vegetation. This vegetation included scattered one and tree cover. The red area shows the area has little of no vegetation since the distance is high.

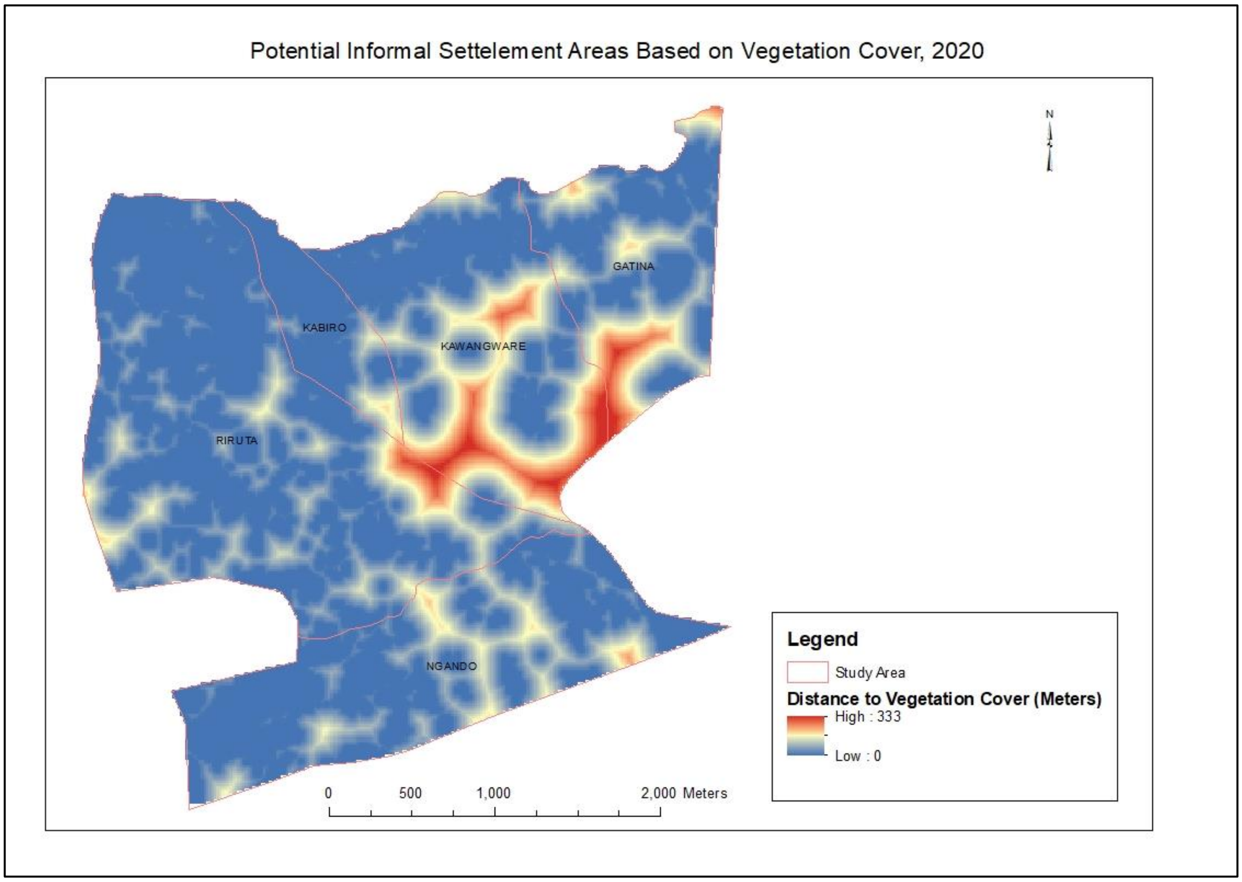

Figure 4.4:Distance of Built-up areas to Vegetation Cover

The three informal settlements maps were combined to give an overall informal settlement probability map probability map which is shown in Figure 4.5. The area in red show a very high probability than the area in blue. It was noted that Kawangware, Gatina and parts of Kabiro and Riruta wards show a very high probability of having informal settlements while in Ngando ward there is a high probability of it few of these settlements.

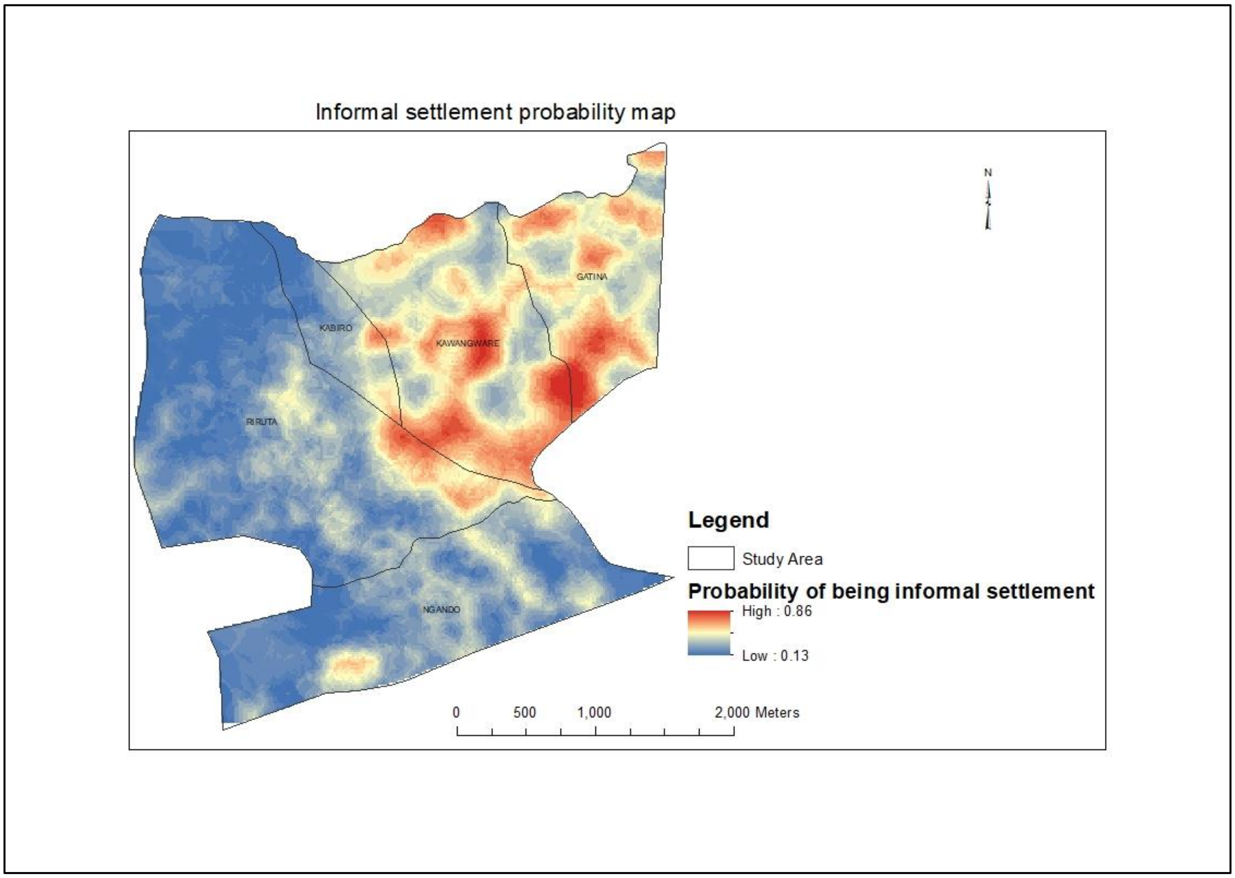

Figure 4.5: Informal Settlements Probability map (2020)

4.3 Kawangware Informal Settlements Spatial Extents for the year 2000-2020

The map shown in figure 4.6 represents the spatial extents informal settlements for year 2000. The areas in yellow show the area covered by these settlements. Most of these settlements were found in Gatina, Kawangware and Kabiro wards. Ngando and Riruta wards were largely covered by vegetation and bare land classes.

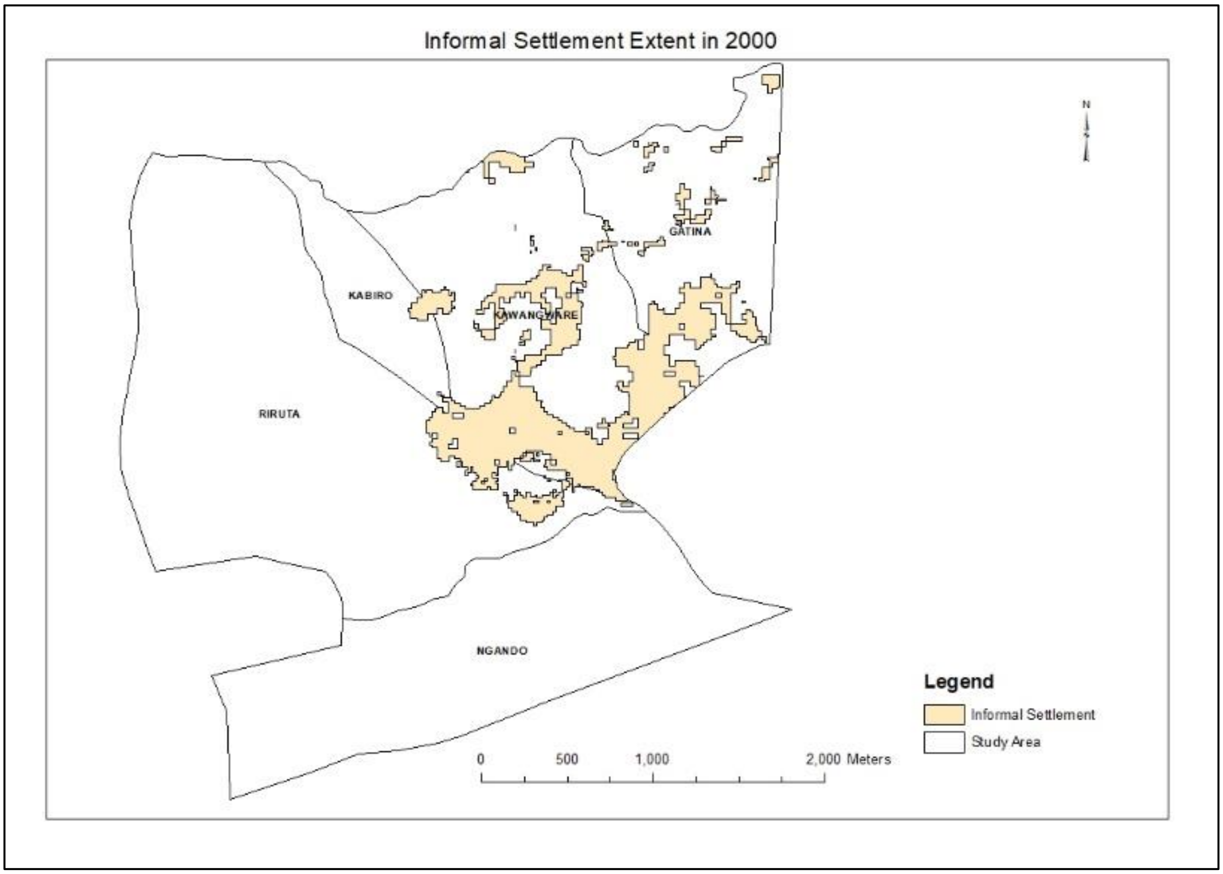

Figure 4.6: Informal Settlements Extent (2000)

The map shown in figure 4.7 represents the spatial extents informal settlements for the year 2010. The areas in yellow show the area covered by these settlements. Most of these settlements were found in Gatina, Kawangware and Kabiro wards. Ngando and Riruta wards were largely covered by vegetation and bare land classes.

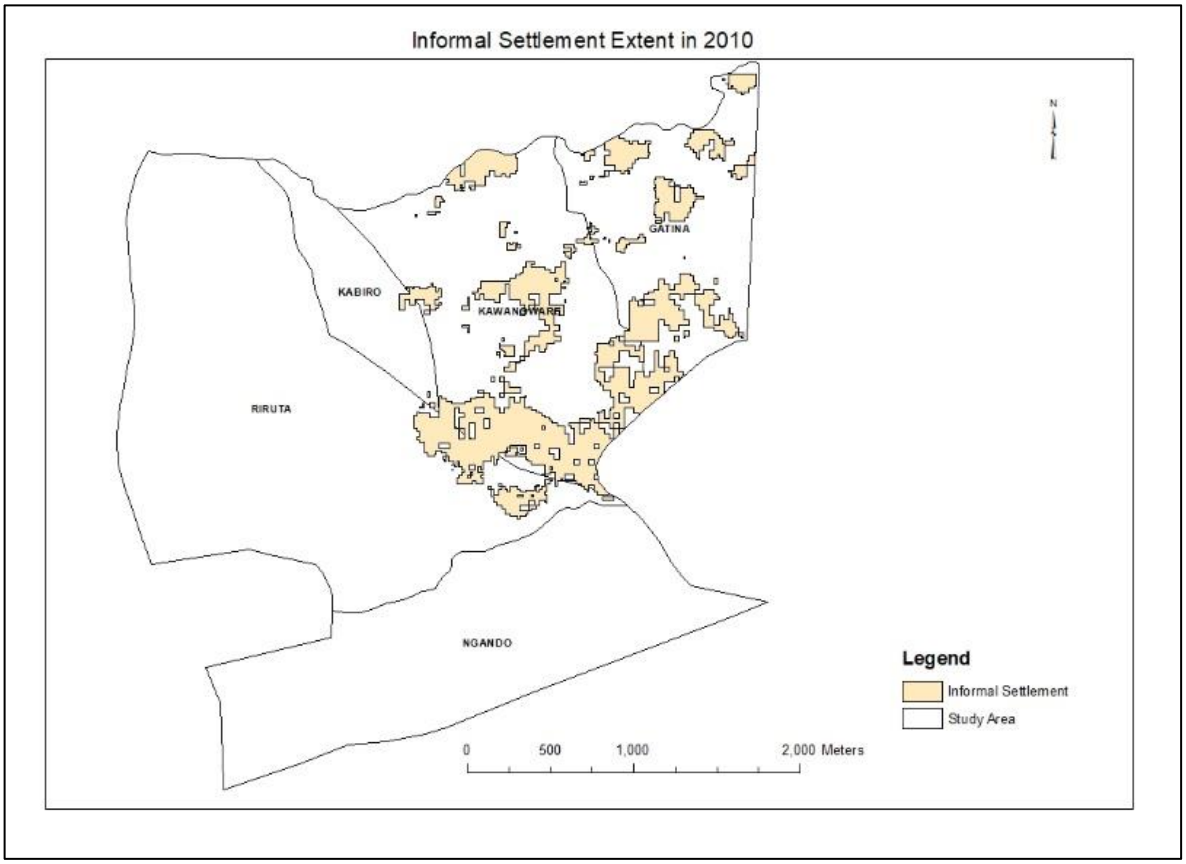

Figure 4.7: Informal Settlements Extent (2010)

The map shown in figure 4.7 represents the spatial extents informal settlements for the year 2020. The areas in yellow show the area covered by these settlements. Most of these settlements were found mostly in Gatina, Kawangware and Kabiro wards. Ngando and Riruta wards were largely covered by vegetation and bare land classes. It can be seen that these settlements tend to increase from 2000 to 2020 by just looking at the extent they cover in each of the map shown.

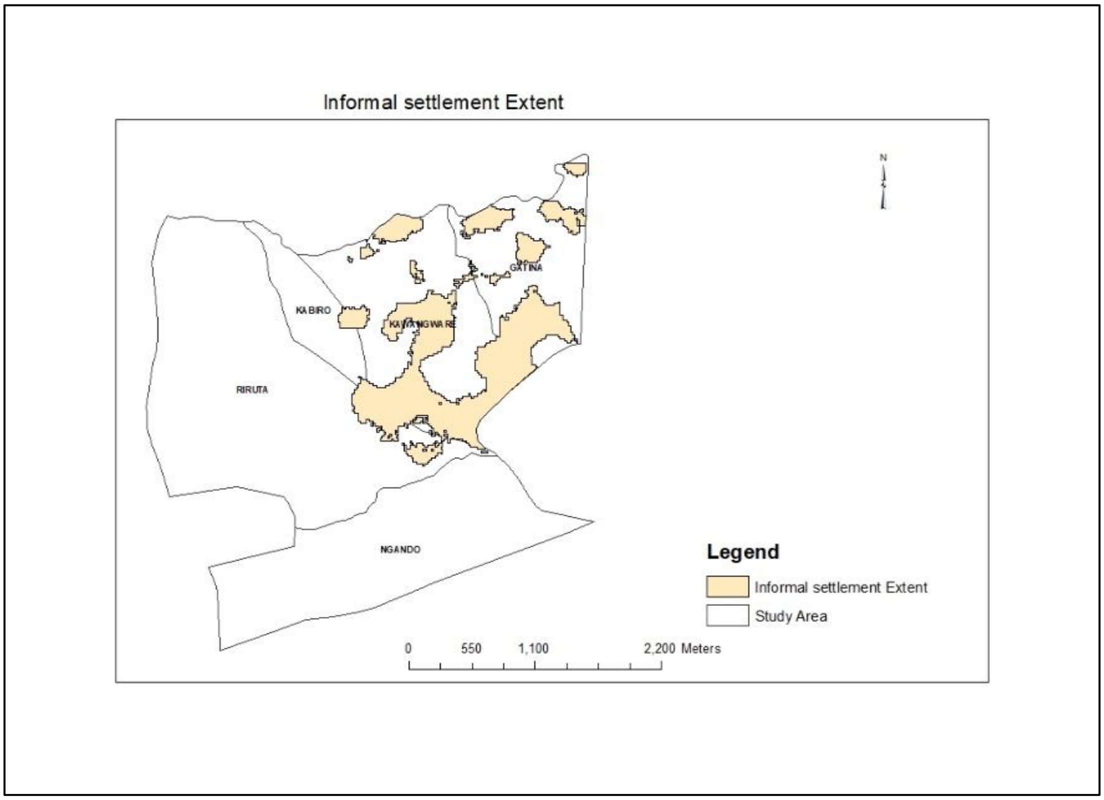

Figure 4.8: Informal Settlements Extent (2020)

4.4 Change Detection Tabular Analysis

Table 4.1 shows area in ha covered by the classes from the year 2000-2020.In 2000 Bare land covered 158.55 ha of the study area and this decreased to 76.98 ha in 2010 and eventually 37.20 ha in 2020. Informal Built-up area increased from 352.08 ha in 2000 to 5598.77 ha in 2020 while vegetation cover decreased from 288.14 ha in 2000to 223.71 ha in 2020.

Table 4.1: Classification Output

Classes/Year	2000 (Ha)	2010 (Ha)	2020 (Ha)
Bare land	158.55	76.98	37.20
Informal Built-up	352.08	356.94	598.77
Grassland	259.63	86.87	198.72
Vegetation	288.14	537.61	223.71

Table 4.2 combined the grassland land class with the vegetation class. There was decrease in the bare land and the vegetation cover from the year 2000 to 2020 indicating that the conversion of land covered by vegetation and bare land was extensive during informal settlement expansion. The area covered by vegetation in Kawangware declined from 547.77 ha in 2000 to 422.43ha in 2020, suggesting an increase in encroachment of farmland. Bare land area, which was 158.55 in 2000, decreased to 37.20 ha in 2020. Informal area increased from 352.08 ha in 2000 to 356.94 ha in 2010 and eventually to 598.77 ha. in 2020.

Table 4.2: Refined Classes- Combining Vegetation & Grassland

Classes/Year	2000 (Ha)	2010 (Ha)	2020 (Ha)
Bare land	158.55	76.98	37.20
Informal Built-up	352.08	356.94	598.77
Vegetation & Grassland	547.77	624.48	422.43
Total	1058.4	1058.4	1058.4

Table 4.3 shows the changes in the classes in the period of 20 years being studied. Informal built up area increased by4.86 ha from 2000 to 2010, 241.83ha from 2010 to 2020 and 246.69 ha from 2000 to 2020.There was a decrease in the areas covered by bare land by 81.57 ha from 2000 to 2010, 39.78 ha from 2010 to 2020 and 121.35 ha from 2000 to 2020.

Table 4.3: Changes in Land use/Land cover from 2000-2020

Classes/Year	2000-2010 (Ha)	2010 -2020 (Ha)	2000-2020 (Ha)
Bare land	-81.57	-39.78	-121.35
Informal Built Up	+4.86	+241.83	+246.69
Vegetation	+76.71	-202.05	-125.34
Change Error	0	0	0

Figure 4.10 shows the bare ground, vegetation cover and informal built-up areas graphically over the three epochs. From the graphs, it is clear that the bare land and vegetation cover decreased over the years. The informal built up area on the other had increase meaning that the area covered by other classes was converted to build up areas. This change was seen to have been experienced more between the year 2010 and 2020 than 2000 and 2010

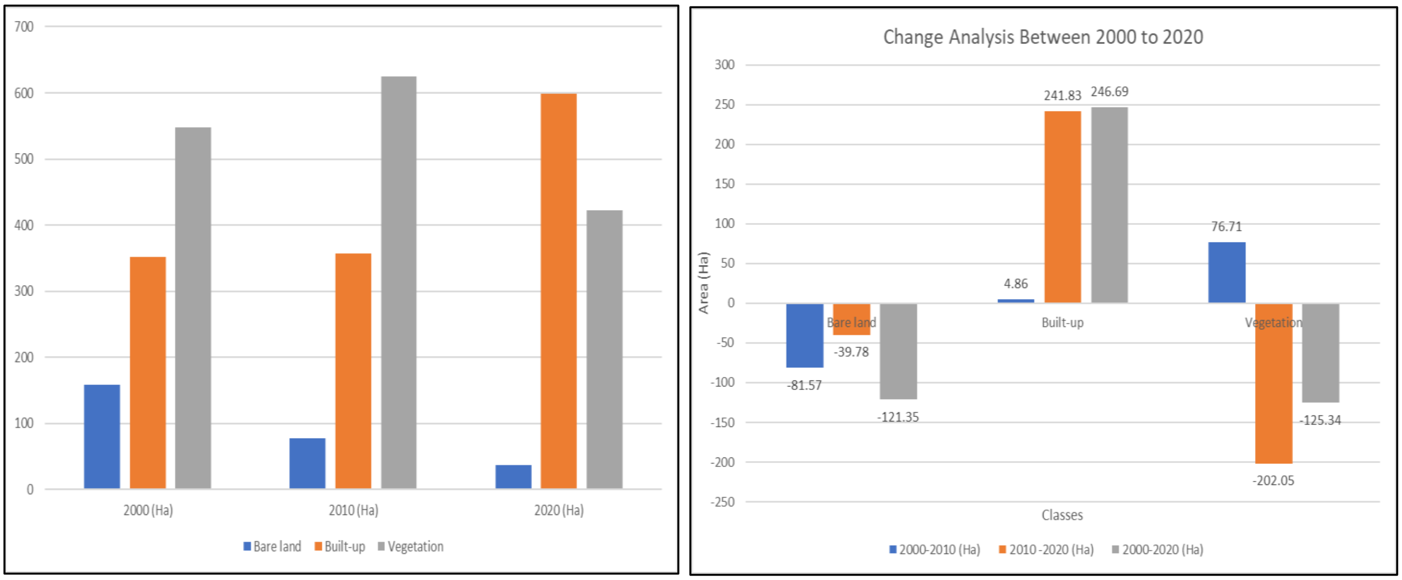

Figure 4.9: Graphical representation of Changes in Informal built up classes and other classes

The percentage increase in the informal built up area were 0.1% from year 2000-2010,4.2% from year 2010 to 2020 and 4.3% from year 2000 to 2020.This showed that from year 2010 to 2020 more of these settlements emerged as compared to the previous decade. These results showed that there has been an increase in the informal settlements in Kawangware area from the year 2000 to 2020.Table 4.6 shows the increase in the informal built up area

Table 4.4: Informal settlements Areas in percentages for the years 2000,2010 and 2020

Year	**Area (Ha)**	**Percentage (%)**	**Percentage Increase**
2000	98.46	9.3	0
2010	99.08	9.4	0.1
2020	143.94	13.6	4.3

The bar chart in Figure 4.10 illustrates informal settlements area in (Ha) for the years 2000, 2010 and 2020. The Y-axis represents the area covered while the X-axis represents the year. It can be seen that the change from year 2000 to 2010 was from 98.46 ha to 99.08 ha while that from 2010 to 2020 was from 99.08 ha to 143.94 ha. This is a clear upward trend in the increase of informal settlements in the study area.

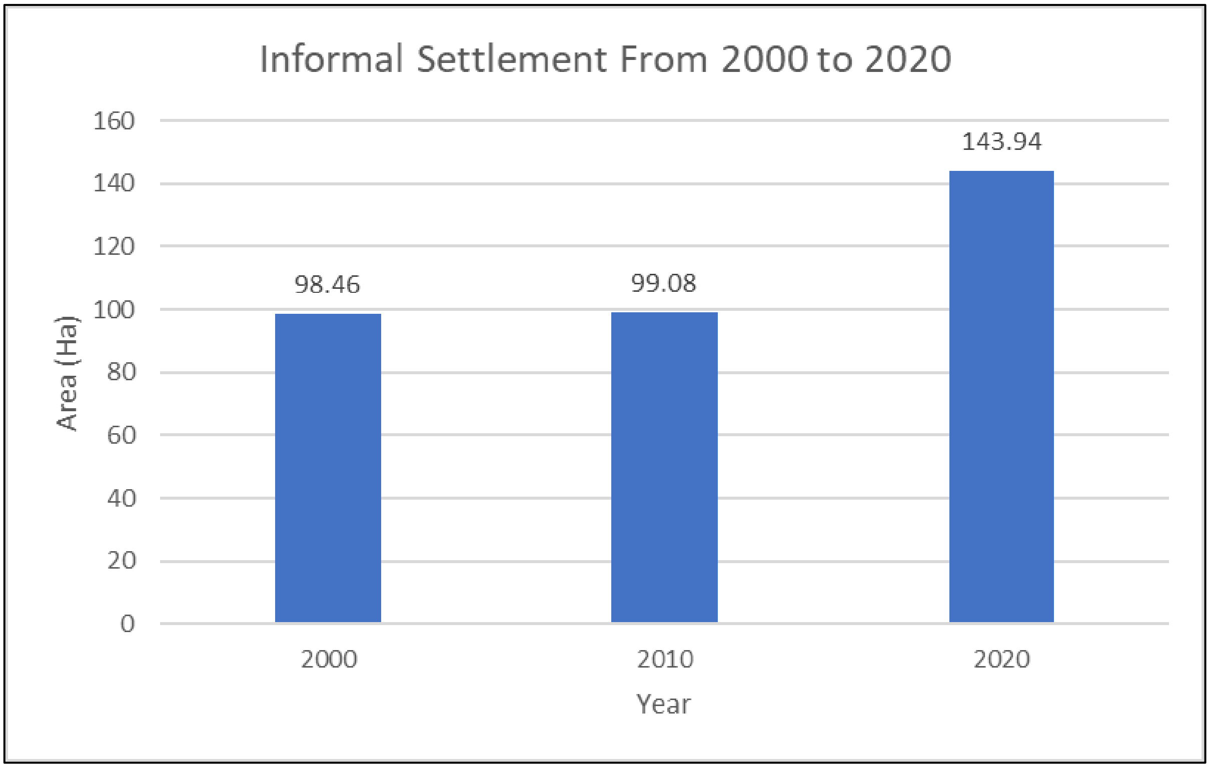

Figure 4.10: Informal settlements Area in (Ha)for the years 2000,2010 and 2020

4.5 Accuracy Assessment

Cells on the diagonal are the numbers of pixels that were correctly identified. Overall accuracy is obtained by dividing this number of pixels with the total number of pixels classified. In order to identify class accuracy, non-diagonal cells in the matrix are used. The non-diagonal cells are used to identify the class accuracy. The cells have the classification errors where the reference imagery and the classified image do not match. These are omission and commission errors.

The commission errors are those that occurs when the classification method assigns pixels to a class to which they do not belong. The number of these pixels are found in the column of the class below or above the main diagonal. The absolute class commission error is the sum of these pixels. The relative class commission error is obtained by dividing the sum by the overall number of class pixels. The producer's accuracy refers to the number of correctly identified cells divided by the total number of pixels in the reference image.

Producers accuracy = 1-commission error

The errors of omission on the other hand is when pixels are mistakenly included into other classes where they do not belong. They are found on the rows cells, which are to right and left from the main diagonal sum is the absolute while when the sum is divided by the total number of pixels in the classified image, it becomes the relative error of omission.

The significance of error of omission is describe by User's accuracy. It is the total number of correctly identified pixels divided by the total number of pixels in the classified image.

Users accuracy = 1- omission error

<u>Equation 4-1: Kappa Coefficient Equation</u>

$$\boxed{Cohen's\ kappa = \frac{N\sum_{i=1}^{m} CM_{ii} - \sum_{i=1}^{m} Ci_{corr} Ci_{pred}}{N^2 - \sum_{i=1}^{m} Ci_{corr} Ci_{pred}}}$$

Source: (Tallón-Ballesteros & Riquelme, 2014)

Kappa coefficient is a measure of how the final classification results compare to values, which are assigned by chance. It usually takes value 0 to 1. If this value is 0, it means that the reference image and the classified image have no relationship. If it is 1, then the ground truth image or the reference image and the classified image are one and the same image. The higher the kappa coefficient is, the more accurate and reliable the classification algorithm is.

4.5.1 Confusion Matrix Tables for the Years 2000 ,2010 and 2020

Table 4.5: Confusion Matrix Table (2000)

REFERENCE DATA											
		Tree Cover	**Built-up**	**Grassland**	**Scattered Vegetation**	**Water Body**	**Bare Ground**	**Total**	**underestimated Pixels**	**Omission error (%)**	**User's Accuracy (%)**
CLASSIFIED DATA	**Tree Cover**	**57**	0	3	2	0	0	**62**	5	8.06	91.94
	Built-up	0	**31**	0	0	1	2	**34**	3	8.82	91.18
	Grassland	2	0	**26**	2	0	2	**32**	6	18.75	81.25
	Scattered Vegetation	5	0	2	**67**	0	0	**74**	7	9.46	90.54
	Water Body	1	0	0	0	**15**	1	**17**	2	11.76	88.24
	Bare Ground	0	3	0	0	2	**24**	**29**	5	17.24	82.76
	Total	**65**	**34**	**31**	**71**	**18**	**29**	**248**			
	overestimated Pixels	8	3	5	4	3	5				
	Commission error (%)	12.31	8.82	16.13	5.63	16.67	17.24				
	Producer's Accuracy (%)	87.69	91.18	83.87	94.37	83.33	82.76				

Overall Accuracy = (57+31+26+67+15+24) * 100/248 = 88.71 %. The overall classification accuracy for the 2010 image is equal to 88.71%. This means that about 89% of the pixels are correctly assigned and 11% of pixels have errors. This is quite high accuracy. Also, the producer's and user's accuracy for each class have been shown in the matrix. The size of omission errors is more common for grassland on the contrary, for bare ground, commission errors prevail. Kappa Coefficient is 0.87. This value means that the reference image and the classified are almost in perfect agreement.

Table 4.6: Confusion Matrix Table (2010)

REFERENCE DATA											
	Tree Cover	**Built-up**	**Grassland**	**Scattered Vegetation**	**Water Body**	**Bare Ground**	**Total**	**underestimated Pixels**	**Omission error (%)**	**User's Accuracy (%)**
CLASSIFIED DATA											
	Tree Cover	**64**	0	5	4	1	1	**75**	11	14.67	85.33
	Built-up	1	**41**	0	0	0	3	**45**	4	8.89	91.11
	Grassland	2	0	**34**	5	0	1	**42**	8	19.05	80.95
	Scattered Vegetation	4	0	3	**71**	2	0	**80**	9	11.25	88.75
	Water Body	1	2	0	1	**18**	0	**22**	4	18.18	81.82
	Bare Ground	0	3	0	0	0	**25**	**28**	3	10.71	89.29
	Total	**72**	**46**	**42**	**81**	**21**	**30**	**292**			
	overestimated Pixels	8	5	8	10	3	5				
	Commission error (%)	11.11	10.87	19.05	12.35	14.29	16.67				
	Producer's Accuracy (%)	88.89	89.13	80.95	87.65	85.71	83.33				

Overall Accuracy = (64+41+34+71+18+25) * 100/292 = 86.64 %. The overall classification accuracy for the 2010 image is equal to 86.64%. This means that about 87% of the pixels are correctly assigned and 13% of pixels have errors. This is quite high accuracy. In addition, the producer and user's accuracy for each class have been shown in the matrix. The size of omission errors is more common for grassland on the contrary, for bare ground, commission errors prevail. Kappa Coefficient is 0.83. This value means that the reference image and the classified are almost in perfect agreement.

Table 4.7: Confusion Matrix Table (2020)

REFERENCE DATA											
CLASSIFIED DATA		**Tree Cover**	**Built-up**	**Grassland**	**Scattered Vegetation**	**Water Body**	**Bare Ground**	**Total**	**underestimated Pixels**	**Omission error (%)**	**User's Accuracy (%)**
	Tree Cover	**69**	0	0	5	0	1	**75**	6	8	92
	Built-up	0	**34**	0	0	0	0	**34**	0	0	100
	Grassland	0	0	**29**	0	0	0	**29**	0	0	100
	Scattered Vegetation	3	0	0	**75**	0	4	**82**	7	8.54	91.46
	Water Body	1	0	0	0	**13**	0	**14**	1	7.14	92.86
	Bare Ground	1	2	0	0	0	**30**	**33**	3	9.09	90.91
	Total	**74**	**36**	**29**	**80**	**13**	**35**	**267**			
	overestimated Pixels	5	2	0	5	0	5				
	Commission error (%)	6.76	5.56	0	6.25	0	14.29				
	Producer's Accuracy (%)	93.24	94.44	100	93.75	100	85.71				

Overall Accuracy = (69+34+29+75+13+30) * 100/267 = 93.63%. The overall classification accuracy for the 2010 image is equal to 93.63%. This means that about 94% of the pixels are correctly assigned and 6% of pixels have errors. This is quite high accuracy. In addition, the producer and user's accuracy for each class have been shown in the matrix. The size of omission errors and commission is more common for bare ground on the contrary, for bare ground, commission errors prevail. Kappa Coefficient is 0.83. This value means that the reference image and the classified are almost in perfect agreement.

4.6 Discussion of the Results

Mapping changes from 2000 to 2020, a time-span of almost 20 years, revealed how the Kawangare informal settlements have grown and changed. It is also important to understand why, these changes have taken place. Secondary data sources should be taken into consideration when trying to answer to those questions.

The results obtained suggests that although Random Forest classification is effective in differentiating classes which have similarities such as informal and formal settlements misclassification can occur, however the method applied in this study enhanced accuracy in differentiating the two classes. This method however is limited to areas where these two classes encompass distinctive structure, pattern and physical features.

The results obtained showed that there has been an increase in informal settlements spatial extent in the study area but do not provide any direct causal informal on the cause of this change. However, from the reviewed literature some of the factors can be thought to have contributed to the above-mentioned change. These factors are like proximity to formal settlements and Nairobi CBD.

From the year 2010 to 2020, it was noted that the increase in informal settlements was more than the previous decade. If this trend continues, the increase in the spatial expansion of the study area could even be more. The cause of this increase could however not be known from the results obtained. The overall accuracy for the three years were 88.71%, 86.64% and 93.63 % for year 200, 2010 and 2020 respectively. The different levels of accuracy can be attributed to the user's accuracy differences and the data that was used.

Chapter 5 : **CONCLUSIONS AND RECOMMENDATIONS**

5.1 Conclusions

The main objective was to map changes in informal settlements of Kawangware using GIS and remotely sensed imagery from year 2000 to 2020. Based on the results which were obtained and analysis, which was done it, can be concluded that the objectives of the project were achieved.

Urban sprawl, which is usually because of conversion of farmland to residential areas due to rural-urban migration, has been experience in the study area over the period of the study. If no measures are taken to curb this, the trend will continue even in the coming decades.

Proximity to formal settlements like areas in Westlands and Lavington and Nairobi CBD has been one of the major causes of increase in the spatial extent of Kawangware informal settlements. This is due to the availability of labor but of very low income. People living in informal settlements obviously earn very little income and thus tend to settle near to their places of work to cut on transportation cost.

Although high resolution remotely sensed data provide an opportunity to map changes in informal settlements, the issue of it being a cost-prohibitive data source is still a drawback. In this project, the application of random forest algorithm for mapping changes of informal settlement on freely available Landsat has been tested. The classification output and the accuracy assessment revealed that this data combined with RF is suitable for mapping changes in informal settlements of an extent not larger than a city due to the visual checks involved.

Freely available data, which can be used in detecting changes in informal settlements, has its limitations. One limitation is the validation issue of this data source. OSM data might be less reliable compared to the data provided by local or international authorities; in this case; data like informal settlements maps created by agencies. Another limitation is the fact it can only be used where informal settlements are structurally different from formal settlements. The opportunities that Open Street Map provides has made it a competitive source of data.

5.2 Recommendations

Volunteered Geographic Information is often not reliable due to various factors such as the level of skills of the contributor, the method used to collect the data provided on the platform. Future research should be conducted for the validation of OSM data source to enhance its reliability.

Where the cost of data to be used is not an issue, high-resolution imagery can replace the freely available Landsat imagery. This will give a better classification output due to how distinctively features appear of high-resolution imageries.

The proposed methodology supports the use of remotely sensed data for mapping changes of informal settlements. However, when considering the visual checks that must be done, the application of this method is more reliable for areas not larger than a city where formal, informal settlements are structurally diverse, and their roof materials are different.

Improvement of classification method could be done by adding 'auxiliary or secondary data to distinguish between informal and formal settlements should be done. This information could be collected by interviews, questionnaires and they could help answer questions on household income, property value, age of the household and ownership. Improving classification methods will help accurately classify slums and monitoring individual informal settlements in order to map the rate of change.

5.3 Areas for Further Research

While the above stated problem is mostly urban related, a slum information system and more broadly an urban information system can be a major data source that could also keep track of the changes of these settlements.

References

Assarkhaniki, Z., & Sabri, S. (2021). Using open data to detect the structure and pattern of informal settlements: an outset to support inclusive SDGs' achievement. *Big Earth Data*, 497-526. Retrieved from 10.1080/20964471.2021.1948178

Blight, G., Fourie, A., Shamrock, J., & Mbande, C. (1999). The effect of waste composition on leachate and gas quality: A study in South Africa. *Waste Management & Research: The Journal for a Sustainable Circular Economy*, 124-140. Retrieved from 10.1177/0734242x9901700208

Brazeau, G. A., & Fung, H. L. (1989). Interferences with assay of creatine kinase activity in vitro. *Biochemical Journal*, 619-621.

Breiman, l. (1996). Bagging predictors. *Machine Learning*, 123-140.

Dovey, K., & King, R. (2011). Forms of informality: Morphology and visibility of informal settlements. *Built Environment*, 11-29.

Duro, C., Franklin, E. S., & Dubé, M. G. (2012). Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. *International Journal of Remote Sensing*, 4506-4526.

Friedl, M., & Brodley, C. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 399-409.

Guan, H., & Chen, Y. (2021). Optimal bands combination selection for extracting garlic planting area with multi-temporal Sentinel-2 imagery. *Sensors*, 5556.

Hidayati, P., Fikri, A., & Hidayat, R. (2020). Aktivitas Dakwah Komunitas Muslim bikers Indonesia (KOMBI) Pekanbaru. *Idarotuna*.

Hofmann, P. (2001). Detecting urban features from IKONOS data using an object-oriented approach.

Hofmann, P., & Bekkarnayeva, G. (2017). Object-based change detection of informal settlements. *2017 Joint Urban Remote Sensing Event (JURSE)*.

Hofmann, P., & Bekkarnayeva, G. (2017). Object-Based Change Detection of Informal Settlements. *2017 Joint Urban Remote Sensing Event (JURSE)*.

Hursikainen, P., & Petri, K. E. (2004). Change detection of informal settlements using multi-temporal aerial photographs: the case of Voi, SE Kenya.

Jokar Arsanjani, J., Helbich, M., Bakillah, M., & Loos, L. (2015). The emergence and evolution of OpenStreetMap: A cellular automata approach. *International Journal of Digital Earth*, 76-90. Retrieved from 10.1080/17538947.2013.847125

KNBS. (2019). 2019 Kenya Population and Housing Census. *Distribution of Population by Administrative Units, II*.

Mahabir, R., & Agouris, P. (2018). Detecting and mapping slums using open data: A case study in Kenya. *International Journal of Digital Earth*, 683-707.

Murphy, M. (1989). Blanco White: Self-banished Spaniard. In M. Murphy, *Blanco White: Self-banished Spaniard* (p. 320).

Noi, P., & Jones, P. (2017). Comparison of Random Forest,k Nearest Naeighbor and Support Vector Machine Classifiers for Land Cover using Sentinel 2 Imagery Sensors. 18.

Obudho, R. (1984). National urban and regional planning policy in Kenya. *Third World Planning Review*, 636.

Rodriguez-Galiano, v., & Chica-Olmo, m. (2012). Random forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, 93-107.

Samahiji, J., & Chaube, A. (1987). Evolution of Bahrain Field in relation to western Arabian Gulf basin. *All Days*.

Selod, H., & Durand, A. (2009). The formalization of urban land tenure in developing countries. *Urban Land Markets*, 101-132.

Shirgaokar, M., & Rumbach, A. (2017). Investigating housing tenures beyond homeownership: A study of informal settlements in Kolkata, India. *Housing Studies*, 117-139.

Smith, Michael, J., Goodchild, & Michael, F. (2007). *Geospatial analysis: A comprehensive guide to principles, techniques and software tools.* Troubador Publishing.

Suhartini, N., & Jones, P. (2020). Better understanding self-organizing cities: A typology of order and rules in informal settlements. *Journal of Regional and City Planning*, 237-263.

Sweeting, P. (2017). Financial enterprise risk management.

Tallón-Ballesteros, J. A., & Riquelme, C. J. (2014). Data mining methods applied to a digital forensics task for supervised machine learning. *Studies in Computational Intelligence*, 413-428.

UN Statistics. (2022, April 22). *United Nations statistics division - Environment statistics.* Retrieved from https://unstats.un.org/unsd/environmentgl/gesform.asp?getitem=467

UN-Habitat. (2013). State of the World's Cities. *Prosperity of Cities*.

UN-habitat. (2014). UN-habitat urban indicators. *Encyclopedia of Quality of Life and Well-Being Research*, 6774.

UN-Habitat. (2016). *Urbanization and Development.* UN-Habitat. Retrieved from https://unhabitat.org/world-cities-report-2016

United Nations. (2019). Urban and rural population growth and world urbanization prospects. *World Urbanization Prospects: The 2018 Revision*, 9-31.

Verma, D., & Jana, A. (2019). Transfer learning approach to map urban slums using high and medium resolution satellite imagery. *Habitat International*, 101-981.

Wang, M., Ding L, Wu , z., & Wang , F. (2020). PSDet: Efficient and universal parking slot detection. *2020 IEEE Intelligent Vehicles Symposium (IV)*.

Appendix

```
# -*- coding: utf-8 -*-
"""
@author: Ann Mutitu
"""
#-------------------Libraries----------------------------
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import cohen_kappa_score
from sklearn import metrics
import rasterio as ro
import pandas as pd

#------------------------Data Extraction----------------------------

init={'l5':'Clipped/LT05/','l7':'Clipped/LT07/','l7':'Clipped/LT07/'}

l5_images= {'Band1':'LT05_B1_Clipped.tif','Band2':'LT05_B2_Clipped.tif',
 'Band3':'LT05_B3_Clipped.tif','Band4':'LT05_B4_Clipped.tif',
 'Band5':'LT05_B5_Clipped.tif','Band6':'LT05_B6_Clipped.tif',
 'Band7':'LT05_B7_Clipped.tif'}

l7_images= {'Band1':'LT07_B1_Clipped.tif','Band2':'LT07_B2_Clipped.tif',
 'Band3':'LT07_B3_Clipped.tif','Band4':'LT07_B4_Clipped.tif',
 'Band5':'LT07_B5_Clipped.tif','Band6_1':'LT07_B6_VCID_1_Clipped.tif',
 'Band6_2':'LT07_B6_VCID_2_Clipped.tif','Band7':'LT07_B7_Clipped.tif'}
```

```
def extract bands(t,l,initial,x,y):
 bands=list(l.keys())
 data={}
 for b in bands:
 image=ro.open(initial+l[b])
 data[b]=image.read(1).flatten()
 data['X_coord']=x
 data['Y_coord']=y
 df=pd.DataFrame(data)
 df.to_csv(t+'.csv',index=None)
 return('Done')
def extract_coords(image):
 x_coord=[]
 y_coord=[]
 bands=image.read()
 shape=bands.shape
 for r in range(shape[1]):
 for c in range(shape[2]):
 coord=ro.transform.xy(image.transform,r,c)
 x_coord.append(coord[0])
 y_coord.append(coord[1])
 return(x_coord,y_coord)

x,y=extract_coords(ro.open('Clipped/LT05/LT05_B1_Clipped.tif'))

landsat5_2000=extract_bands('Landsat5_2000',l5_images,init['l5'],x,y)
landsat7_2010=extract_bands('Landsat7_2010',l7_images,init['l7'],x,y)
```

```
landsat7_2020=extract_bands('Landsat7_2020',l7_images,init['l7'],x,y)

#----------------Random Forest Classification-------------------------

image_data=pd.read_csv('Extracted Data/Landsat7_2020.csv')
unclassified_image=image_data[['Band1','Band2','Band3','Band4',
 'Band5','Band6','Band7','Band9',
 'Band10','Band11']]

training_data=pd.read_csv('Training/training_points.csv')

X=training_data[['Band1','Band2','Band3','Band4','Band5','Band6','Band7',
 'Band9','Band10','Band11']]

y=training_data['class']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25)
clf=RandomForestClassifier(n_estimators=100)

clf.fit(X_train,y_train) #Train the model
y_pred=clf.predict(X_test) #Classify know classes for accuracy Assessment

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
classification=clf.predict(unclassified_image) #Classify the image
image_data['Class']=classification
image_data.to_csv('Classified_Data/Landsat 8 Classified.csv',index=None) #Save

#----------------Accuracy Assessment-------------------------
```

```
training_data=pd.read_csv('Training/training_points.csv')
X=training_data[['Band1','Band2','Band3','Band4','Band5','Band6',
'Band7','Band9','Band10','Band11']]
y=training_data['class']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5)
clf=RandomForestClassifier(n_estimators=100)
clf.fit(X_train,y_train)
y_pred=clf.predict(X_test)

d={'Actual Class':y_test,'Predicted Class':y_pred}

df=pd.DataFrame(d)

pred=df.groupby('Predicted Class').size()
actaul=df.groupby('Actual Class').size()

veg=df[df['Actual Class']=='Vegetation']
built=df[df['Actual Class']=='Built-up']
grass=df[df['Actual Class']=='Grassland']
scatttered=df[df['Actual Class']=='Scattered Vegetation']
water=df[df['Actual Class']=='Water Body']
bare=df[df['Actual Class']=='Bare Ground']

#Confusion Matrix and Kappa Coefficient

m=confusion_matrix(y_test, y_pred, labels=["Vegetation", "Built-up",
"Grassland","Scattered Vegetation",
"Water Body","Bare Ground"])
```

```
df1=pd.DataFrame(m,columns=["Vegetation", "Built-up", "Grassland",
"Scattered Vegetation","Water Body","Bare Ground"],
 index=["Vegetation", "Built-up", "Grassland",
 "Scattered Vegetation","Water Body","Bare Ground"]

k=cohen_kappa_score(y_test, y_pred, labels=["Vegetation", "Built-up",
"Grassland","Scattered Vegetation",
"Water Body","Bare Ground"])

#-----------------------------------END--------------------------------------------------------
```