Analysis of High-Resolution Land Cover Classification Methods and Evaluation of Vegetation's Role in Temperature Mitigation: A Case Study in Ouagadougou

Author

Farnoosh Noroozi

(289884)

Supervisor

Prof. Dr. E. Belcore

Co supervisor

Prof. Dr. P. Dabove

The Department of ENVIRONMENT, LAND AND INFRASTRUCTURE ENGINEERING (DIATI)

The POLYTECHNIC UNIVERSITY OF TURIN

October 2024

Abstract

Monitoring and analyzing land cover in urban areas like Ouagadougou, Burkina Faso, is essential for effective urban planning and environmental management. This study aimed to generate a high-resolution land cover map of Ouagadougou using PlanetScope satellite imagery from May 10, 2023. The classification system focused on distinguishing key land cover types, including vegetation, water, bare soil, and built-up areas. NDVI (Normalized Difference Vegetation Index) was computed using RGB and NIR bands to assist in the classification process.

In this study, several methods were compared to classify Ouagadougou's land cover. First, NDVI thresholds were defined in QGIS to classify the land cover types. These thresholds were then applied in eCognition through Object-Based Image Analysis (OBIA) as a second method, focusing on rule-based classification. To refine the classification, GLCM (Gray Level Co-occurrence Matrix) texture analysis was incorporated as an additional condition. As a final method, supervised classification using machine learning (Random Forest) was performed with manually defined training samples. This method, also using OBIA, provided a comparative approach. Upon comparing the accuracy of these methods, the supervised classification achieved the highest overall accuracy at 75%.

In addition, we analyzed the relationship between vegetation coverage and surface temperature in the area. By correlating vegetation percentage with surface temperature, we observed an inverse relationship, where areas with lower vegetation coverage exhibited higher surface temperatures. This finding highlights the role of green spaces in moderating urban temperatures.

These results demonstrate the effectiveness of combining multiple techniques for urban land cover mapping, providing a methodological framework adaptable for similar studies in other regions.

Keywords: Land Cover Classification, Urban Planning, Object-Based Image Analysis (OBIA), Supervised Classification, Random Tree Classifier, Normalized Difference Vegetation Index (NDVI), Texture Analysis, Gray Level Co-occurrence Matrix (GLCM), Remote Sensing, High-Resolution Satellite Imagery, Surface Temperature, Urban Heat Islands, Thermal Regulation, Land Surface Temperature (LST).

Acknowledgements

I would like to begin by expressing my sincere gratitude to my supervisor, Professor Elena Belcore, for her constant guidance, availability, and for teaching me so much throughout this journey. I am also thankful to Professor Paolo Dabove, who inspired my enthusiasm for this subject and introduced me to valuable opportunities for this project.

I owe heartfelt thanks to my family, especially my parents, for their endless love, encouragement, and unwavering support. To my older brother, thank you for always supporting me, even from afar, and for being my motivation.To my younger brother, thank you for offering me thoughtful advice when I needed it and given me the confidence to move forward.

To my dear friends Ameneh Darban and Sepehr Razavian, I am especially grateful for always being by my side and never hesitating to help. I also thank my other friends who have shared many wonderful moments with me, creating memories that I will cherish forever.

A special thanks to the city of Torino and its people, as well as my Italian friends, who have shown me warmth and generosity. Their friendliness and kindness made me feel close to them, and because of them, I have truly come to love Italy.

Contents

List of Figures:

List of Tables:

1. Introduction

1.1. Background

Accurate land cover classification is essential for understanding environmental dynamics and supporting sustainable development, particularly in rapidly urbanizing areas. Land cover refers to the physical materials present on the Earth's surface, including vegetation, urban infrastructure, water bodies, and bare soil. The classification of these features is crucial for urban planning, natural resource management, biodiversity conservation, and climate change monitoring.

Advancements in remote sensing technology have significantly enhanced our ability to capture high-resolution imagery and derive meaningful insights into land cover dynamics. Remote sensing serves as a vital tool in environmental monitoring, offering detailed and up-to-date information on land cover patterns. This is particularly important in urban environments, where land cover classification supports effective urban planning and environmental management.

1.2. Objectives of the Study

This project aims to improve how land cover is classified in Ouagadougou, the capital city of Burkina Faso, which is expanding quickly. The research uses high-resolution satellite images from May 10, 2023, and combines several analysis techniques to identify different types of land cover within the city’s urban environment.

The study starts by using the Normalized Difference Vegetation Index (NDVI) to create initial maps of vegetation cover. These maps are then processed using Object-Based Image Analysis (OBIA) in the eCognition software, with added detail from Gray Level Co-occurrence Matrix (GLCM) texture analysis, to enhance classification accuracy. Additionally, a supervised classification test utilizing the random forest algorithm is conducted to evaluate its effectiveness of machine learning against other OBIA-based methods.

As a final step, the study correlates vegetation percentage with surface temperature across the area to explore the relationship between Vegetation Cover and urban Heat Distribution. This analysis

provides insights into the impact of vegetation on urban temperature, which can inform future urban planning and environmental management strategies.

A key part of this research is to check the accuracy of each method by comparing it to actual ground data. This will help determine the most reliable approach for mapping different land covers in Ouagadougou's complex urban area. The analysis is carried out using QGIS and eCognition software, giving insights into how well each method performs in urban settings.

1.3. Research Questions

In this study, we aim to address several key questions that will guide the analysis and classification of Ouagadougou's land cover. The following questions will help assess the effectiveness of different classification techniques and their impact on the overall accuracy:

- Which classification method proves most effective at distinguishing between different types of urban land cover?
- To what extent does incorporating NDVI thresholding improve the accuracy of land cover classification in urban environments?
- Does NDVI thresholding provide a reliable preliminary layer for further analysis in OBIA workflows?
- How does the integration of Gray Level Co-occurrence Matrix (GLCM) texture analysis influence the precision of land cover classifications obtained from OBIA?
- How does the use of a machine learning approach, specifically the random forest algorithm, compare with traditional supervised classification methods in terms of classifying urban land cover?
- What are the benefits and limitations of employing machine learning algorithms like random forest in urban land cover analysis?
- What are the synergistic effects, if any, of combining NDVI thresholding, GLCM texture analysis, and machine learning techniques on the overall accuracy of land cover classification?

- What is the relationship between vegetation cover percentage and surface temperature in the study area, and how does this correlation inform our understanding of urban heat distribution?

1.4. Structure of the Thesis

This thesis is organized into several key sections. The **Introduction** gives an overview of Ouagadougou's urbanization trends, explains the importance of land cover classification, and outlines the research objectives. The **Literature Review** looks at past studies on remote sensing, land cover classification methods, urban heat islands, and the effects of climate change. The **Methodology** chapter describes how the data was collected, processed, and analyzed, including the use of OBIA, supervised classification, GLCM texture analysis, and NDVI thresholding to classify land cover types. It also explains the steps taken to calculate vegetation percentage and surface temperature for each grid in the study area.

The **Results** section shows the findings from the classification process, describing the land cover types in Ouagadougou. It also looks at the relationship between vegetation percentage and surface temperature, showing how areas with less vegetation tend to have higher temperatures.

The **Discussion** explains these results in the context of urban development and environmental impacts, focusing on how vegetation can help reduce surface temperatures in urban areas. The **Conclusion** summarizes the key findings, explains their importance for urban planning and environmental management, and suggests ideas for future research. Finally, the **References** section lists all the sources used in the thesis, and the **Appendices** include extra technical details and supporting information.

This structured approach ensures a comprehensive examination of the study area and the methodologies used, contributing valuable insights into the environmental challenges and opportunities in rapidly urbanizing regions like Ouagadougou.

2. Literature Review

2.1. Land Use and Land Cover (LULC)

Land cover refers to the physical and visible features on the Earth's surface, such as forests, grasslands, water bodies, urban infrastructure, and bare soil. It describes what is physically present in a given area, whether natural or human-made, and is commonly identified through satellite imagery and remote sensing technologies. In contrast, land use refers to how these physical areas are utilized by humans for various purposes, such as agriculture, urban development, industry, or recreation. While land cover can be easily observed through visual data, land use involves a deeper understanding of human activities and their impact on the landscape[1].

The distinction between land cover and land use is important for environmental planning and resource management. Land cover provides insight into the natural environment, whereas land use reveals the influence of human activities. Understanding both helps in effective planning for urban development, agriculture, and environmental conservation, making them critical components in geographical and environmental studies.

2.2. Geographic Information System (GIS)

A Geographic Information System (GIS) is a computer-based tool that analyzes and visualizes spatial and geographic data to understand complex relationships between data layers, enabling the mapping and study of changes on Earth. GIS facilitates the integration of geospatial data from sources like satellite imagery, GPS data, and descriptive attributes tied to specific locations. By overlaying and analyzing this data, GIS helps uncover spatial patterns, trends, and relationships, making it essential for tasks such as land use and land cover (LULC) analysis, resource management, and public health[2].

GIS is composed of several key components that work together for effective spatial data analysis. Hardware includes the physical devices like computers and GPS units used for data collection and processing. Software, such as ArcGIS and QGIS, offers tools for spatial analysis, map production, and data management.

GIS handles two main types of spatial data: Raster and Vector. Raster data consists of grid-based pixels and is typically used for continuous data such as satellite imagery, digital elevation models, and vegetation indices. This data is particularly useful for representing phenomena like temperature, land cover, and elevation. On the other hand, Vector data represents geographic features through points, lines, and polygons, making it ideal for mapping discrete elements like locations (e.g., cities), linear features (e.g., roads), and areas (e.g., land parcels). Both Raster and Vector data are fundamental in capturing and analyzing spatial information in GIS applications, each serving distinct purposes depending on the nature of the analysis.

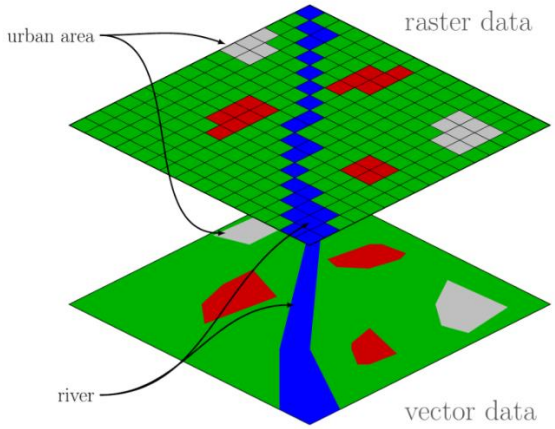

Figure 1. Vector and Raster data representation[3]

GIS's core functionalities include data capture and integration from various sources, efficient data storage and management, spatial analysis operations like buffer and overlay analysis, and visualization techniques like thematic maps and 3D models to communicate geographic information effectively. It is widely applied across fields, including urban planning for infrastructure development, environmental monitoring to track changes and manage resources, disaster management for risk mapping and recovery, and public health for disease tracking and healthcare planning.

Moreover, GIS is often integrated with remote sensing to enhance analysis of Earth's surface changes. Remote sensing provides critical data like vegetation indices and land cover maps, which GIS combines with socio-economic information for complex analyses, such as assessing the impact of

urban expansion on biodiversity. The integration of technologies like GPS and IoT further enhances GIS's capabilities in dynamic spatial decision-making across diverse sectors.

2.3. Remote Sensing

Remote sensing is an innovative technology that allows the collection of information about the Earth's surface from a distance, without any physical interaction. By using sensors on satellites or aircraft, electromagnetic radiation—either reflected or emitted by objects on the ground—is detected and recorded. This technique is highly versatile and plays a key role in many fields, such as environmental monitoring, land use and land cover mapping, agriculture, and urban planning. The remotely captured data offers valuable insights into surface conditions, contributing to better decision-making and resource management.

The data gathered from these platforms, whether from satellites or drones, can be processed and analyzed to reveal meaningful patterns and trends in the landscape. This makes remote sensing particularly useful in understanding geographical changes, monitoring vegetation health, assessing urban expansion, and tracking natural phenomena. By applying these insights, governments, researchers, and planners can take informed actions to protect the environment, improve urban development, and manage resources efficiently[4].

2.4. Observing with the Electromagnetic Spectrum

Electromagnetic energy, generated by the movement of charged particles, travels in wave form through both the atmosphere and the vacuum of space. These waves vary in wavelength (the distance between consecutive wave crests) and frequency, with shorter wavelengths corresponding to higher frequencies. Waves like radio, microwave, and infrared have longer wavelengths, while ultraviolet, x-rays, and gamma rays possess much shorter wavelengths. Visible light, which lies in the middle of this spectrum, is the only type of electromagnetic energy that the human eye can perceive. All other forms of electromagnetic energy require specialized instruments for detection[5].

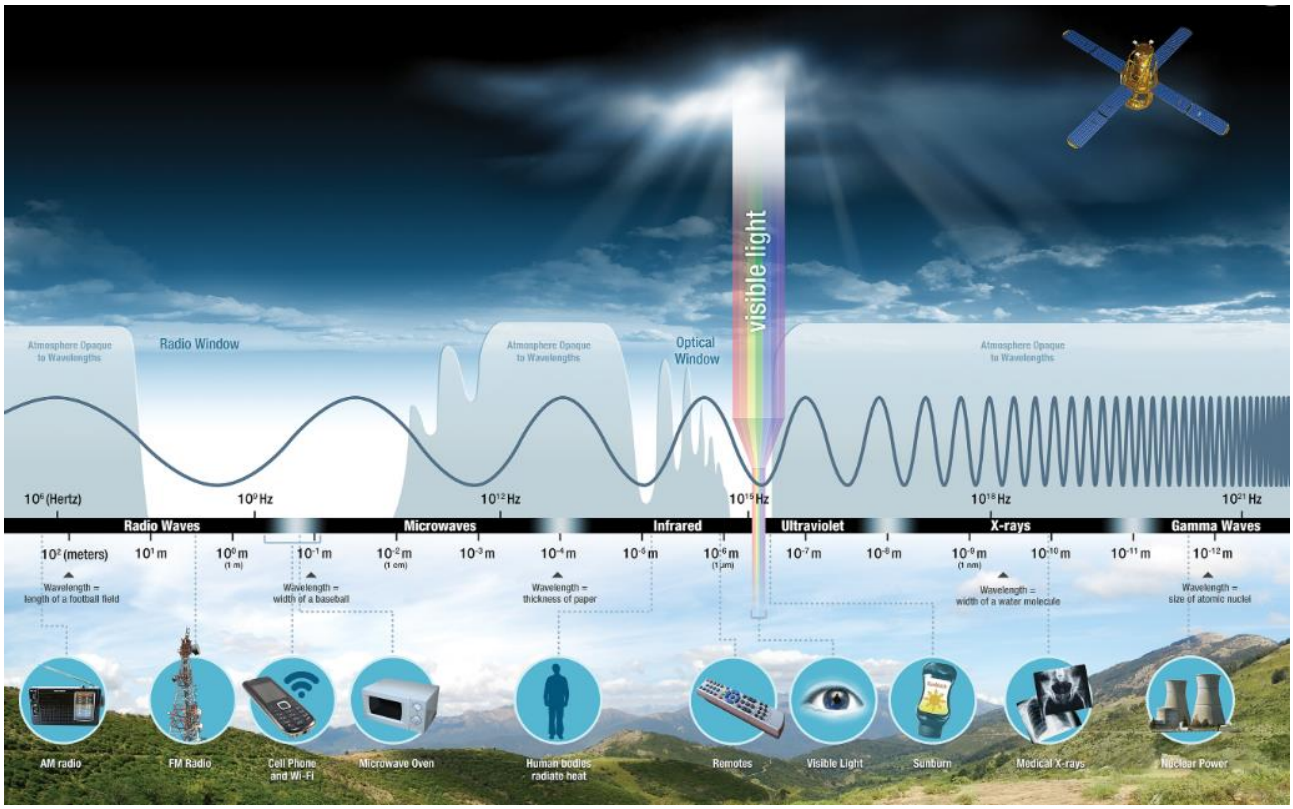

Figure 2. Diagram of the Electromagnetic Spectrum[6]

Some waves are either absorbed or reflected by atmospheric components such as water vapor and carbon dioxide, while others pass through the atmosphere without obstruction; visible light falls into the latter category, allowing it to be transmitted through the atmosphere. Microwave energy also has the ability to penetrate clouds, making it valuable for weather and communication satellites.

The Sun is the primary source of the energy observed by satellites. The amount of sunlight reflected by a surface is influenced by its texture and albedo (a measure of how well a surface reflects light instead of absorbing it). For instance, snow has a very high albedo, reflecting up to 90% of incoming solar radiation, whereas the ocean reflects only about 6%, absorbing the remainder. Absorbed energy is often re-emitted at longer wavelengths; for example, the energy absorbed by the ocean is re-emitted as infrared radiation.

Every object on Earth reflects, absorbs, or transmits energy differently, creating unique spectral signatures that can be identified by remote sensing instruments. These spectral signatures, including those captured by RGB and NIR bands, allow researchers to distinguish between different Earth features, rock types, and vegetation states. The ability to differentiate materials is dependent on an

instrument’s spectral resolution, which is determined by the number of spectral bands it can detect. This makes RGB and NIR data essential tools in environmental monitoring, land use studies, and climate change research.

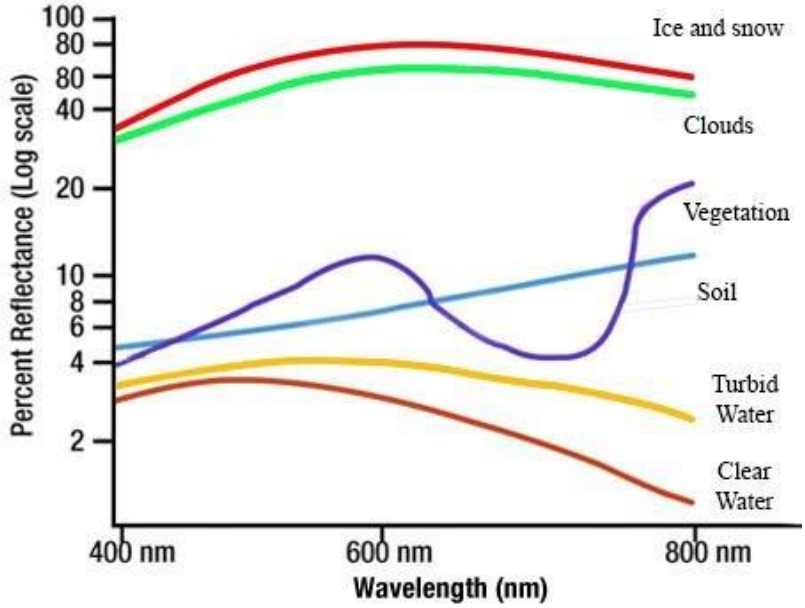

Figure 3. Spectral signatures of different Earth features within the visible light spectrum[7].

2.5. Spectral Bands Classification

Satellite imagery refers to systematically captured photographs of specific areas of the Earth's surface. These images are taken by satellite sensors using various bands of the electromagnetic spectrum. These images possess distinct characteristics based on the spectral bands used. Spectral bands are groups of wavelengths. For example, ultraviolet, visible, near-infrared, thermal infrared, and microwave are spectral bands. We categorize each spectral region based on its frequency (v) or wavelength.

Panchromatic Imaging Systems utilize a single channel sensor sensitive to a broad wavelength range, often within the visible spectrum, producing black-and-white images resembling photographs. These systems capture brightness but lose spectral (color) information.

Multispectral Imaging Systems feature multiple spectral bands, each capturing specific wavelength ranges in different bands of electromagnetic spectrum e.g. infrared, visible, Ultraviolet etc. These systems produce multilayer images containing both brightness and spectral information, making them versatile for applications like land use analysis. Multi-spectral imagery uses 3-10 bands for capturing a single imagery.

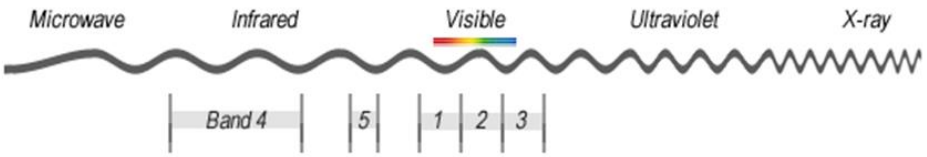

Figure 4. Multispectral Imaging Systems[8]

Hyperspectral Imaging Systems, or imaging spectrometers, capture images in hundreds of contiguous spectral bands, providing detailed spectral information that allows precise characterization and identification of targets. These systems are invaluable in fields like precision agriculture and coastal management.

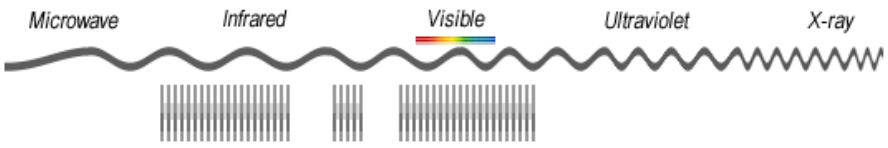

Figure 5. Hyperspectral Imaging Systems

The main difference between multispectral and hyperspectral is the number of bands and how narrow the bands are. Hyperspectral images have hundreds of narrow bands, multispectral images consist of 3-10 wider bands.

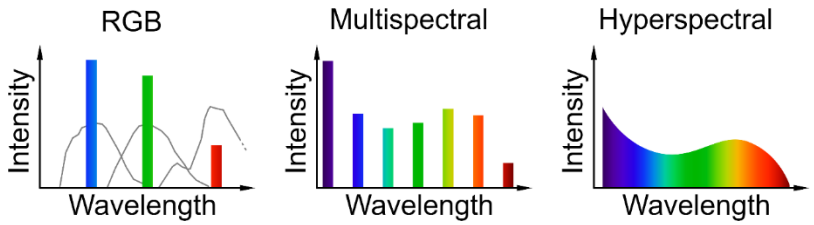

Figure 6. Definition of different imaging approaches based on spectral resolution[9].

2.6. Types of Resolution

Remote sensing systems are characterized by their spatial, spectral, temporal, and radiometric resolutions. These resolutions determine the level of detail, the range of wavelengths, the frequency of data acquisition, and the sensitivity of the sensor, respectively. High-resolution imagery provides detailed information, while lower resolution data covers larger areas.

Radiometric resolution is the amount of information in each pixel, that is, the number of bits representing the energy recorded. Thus, the higher the radiometric resolution, the more values are available to store information, providing better discrimination between even the slightest differences in energy. For example, when assessing water quality, radiometric resolution is necessary to distinguish between subtle differences in ocean color.

Spatial resolution is the area represented by a pixel. It describes how much detail in an image is visible to the human eye. Spatial resolution is the ability of the sensor to differentiate between various objects and features. The clarity of features on earth's surface depends on the size of the pixel and the number of pixels in a given imagery. A single pixel can only represent one color. Therefore, if a pixel in an image corresponds to a large area of land, it may obscure smaller details within that area[10].

Figure 7. Spatial Resolution[11].

Spectral resolution is the ability of a sensor to detect finer wavelengths, that is, having more and narrower bands. The more wavelengths a sensor can capture, the more detailed the land-use and land-cover information it provides. Many sensors are considered to be multispectral, meaning they have 3-10 bands. Some sensors have hundreds to even thousands of bands and are considered to be hyperspectral. The narrower the range of wavelengths for a given band, the finer the spectral

resolution. For instance, using only the visible spectrum limits us to features detectable by the human eye, such as iron ore presence. However, utilizing additional spectra like gamma rays enables the detection of elements like potassium, uranium, and thorium. Similarly, the infrared spectrum is essential for analyzing heat signatures. Essentially, the more spectral bands a satellite sensor captures, the higher its spectral resolution, leading to more detailed imagery[12].

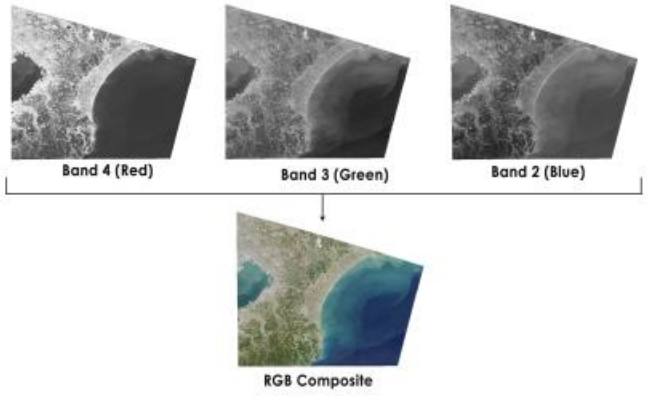

Figure 8. Spectral Resolution[13]

Temporal resolution refers to the frequency with which a satellite captures images of a specific area over a given time period. It is determined by how often a satellite revisits and photographs the same location. The amount of temporal data available depends on the satellite's orbit and its speed around Earth. Essentially, higher temporal resolution means more frequent updates of imagery for a particular region[14].

Figure 9. Temporal Resolution

2.7. Type of Sensors

Instruments or sensors on satellites and aircraft either utilize the Sun as a source of light or generate their own illumination, measuring the energy that is reflected back. Sensors that rely on sunlight are known as passive sensors, while those that emit their own energy are referred to as active sensors.

Passive sensors encompass various radiometers, which quantitatively measure the intensity of electromagnetic radiation in specific bands, and spectrometers, which detect, measure, and analyze the spectral content of reflected electromagnetic radiation. These sensors primarily operate within the visible, infrared, thermal infrared, and microwave regions of the electromagnetic spectrum. They are widely used to measure attributes such as land and sea surface temperature, vegetation characteristics, cloud and aerosol properties, and other physical features. However, most passive sensors are unable to penetrate dense cloud cover, limiting their effectiveness in regions like the tropics, where such conditions are frequent. For example, Landsat and Sentinel are passive sensors. They capture images by sensing reflected sunlight in the electromagnetic spectrum[15].

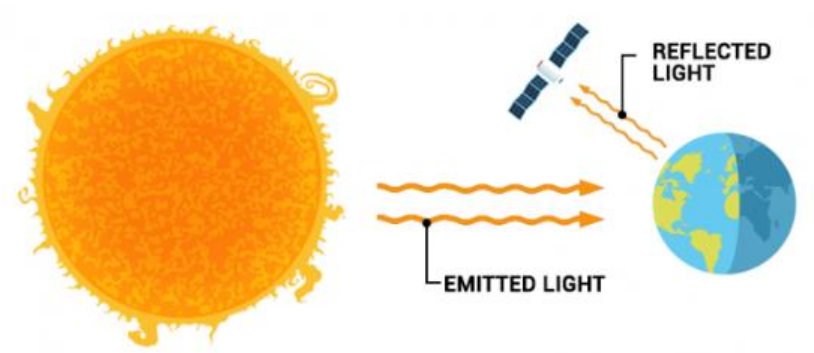

Figure 10.passive sensors (measuring reflected energy emitted from the sun)[16]

Active sensors, on the other hand, include radar sensors, altimeters, and scatterometers. These typically operate in the microwave band of the electromagnetic spectrum, enabling them to penetrate the atmosphere under most conditions. Such sensors are particularly useful for assessing vertical profiles of aerosols, forest structures, precipitation and winds, sea surface topography, and ice, among other features[17].

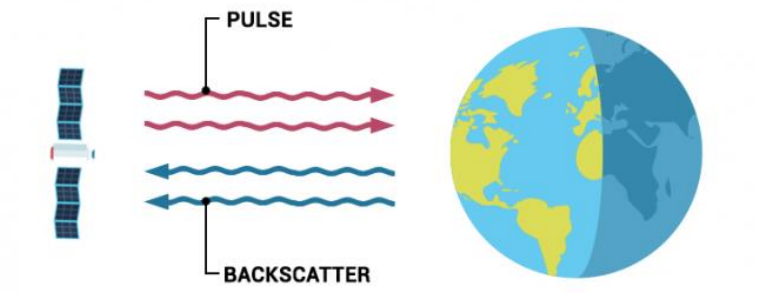

Figure 11. Active sensors (illuminating its target and measuring its backscatter)[18]

2.8. Overview of Sensors and Platforms

Satellites play a crucial role in various scientific and technological applications, classified primarily into four main types based on their functions: communication satellites, Earth observation satellites, navigation satellites, and astronomical satellites. Each type serves distinct purposes, from enabling global communications and broadcasting to providing weather forecasts, navigational data, and deep space exploration.

Different types of satellites are designed and built to fulfill these specific functions, with varying sizes, orbits, frequencies, and technology. The diversity of satellite types allows for better coverage, higher efficiency, and more accurate results in their respective domains.

Among these, Earth observation satellites are particularly vital for monitoring and studying our planet. They are designed to acquire data about the earth's surface and atmosphere. including vegetation cover, buildings, water surfaces, air temperatures, ground elevations, and many other characteristics. Earth observation satellites can be further categorized based on their specific functions, such as weather satellites for monitoring atmospheric conditions and remote sensing satellites for mapping and environmental analysis.Specific examples include weather satellites in geostationary orbits for consistent cloud pattern tracking, and remote sensing satellites like Sentinel, Landsat, and PlanetScope, which provide valuable data for various environmental and geographic applications[19].

The type of sensor on Earth observation satellites determines the kind of data they collect. These sensors are generally classified into two categories: optical sensors and radar (or microwave) sensors.

2.8.1. Optical Platforms

Optical remote sensing uses visible, near-infrared, and short-wave infrared sensors to capture images of the Earth's surface by detecting solar radiation reflected from objects on the ground, a process known as passive remote sensing. These platforms are crucial for applications requiring detailed imagery, such as land cover classification, vegetation health monitoring, and environmental change detection. Different materials reflect and absorb light uniquely at various wavelengths, allowing for the identification and differentiation of objects based on their spectral reflectance signatures. Depending on their spectral and spatial resolutions, optical platforms like Sentinel-2, Landsat, and PlanetScope offer versatile data suitable for a wide range of analyses, from large-scale environmental monitoring to high-resolution agricultural assessments.

Sentinel-2 (ESA): Sentinel-2, operated by the European Space Agency (ESA), is equipped with multispectral sensors that capture data across 13 spectral bands, making it highly versatile for environmental monitoring. These bands include the visible (RGB), near-infrared (NIR), and shortwave infrared (SWIR) regions, allowing for detailed analysis of various land surface characteristics.

The visible bands (2, 3, 4) cover the blue, green, and red wavelengths, which are commonly used in creating true-color images. The near-infrared band (Band 8) is particularly useful for vegetation analysis, as healthy plants reflect more NIR light, making it effective for calculating indices like NDVI. The shortwave infrared bands (11, 12) help assess moisture content in soil and vegetation, and they are valuable for detecting water stress, drought conditions, and fires. Sentinel-2's high spatial resolution, with bands ranging from 10 to 60 meters, makes it suitable for detailed land cover classification, monitoring agricultural productivity, and managing natural disasters such as floods and fires.

This combination of bands, especially the NIR, allows Sentinel-2 to be a powerful tool for vegetation analysis, urban monitoring, and natural resource management. It plays a critical role in understanding changes in land use and environmental conditions[20].

Figure 12. Sentinel-2 Satellite

Landsat (NASA/USGS): Landsat satellites have been providing valuable Earth observation data since the 1970s, offering consistent, long-term datasets for studying environmental changes, land use, and resource management. The Landsat series, including Landsat 8 and Landsat 9, has been instrumental in capturing data in various spectral bands, each designed to monitor specific aspects of the Earth's surface.

Landsat 8, launched in 2013, includes two primary sensors: the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). The OLI captures data in visible, near-infrared (NIR), and shortwave infrared (SWIR) bands, useful for applications such as land cover classification, vegetation analysis, and water resource management. These spectral bands allow scientists to monitor natural phenomena like deforestation, urban growth, and coastal changes. The TIRS, specifically designed for thermal infrared data collection, measures the heat emitted from the Earth's surface, providing data on land surface temperature (LST). This sensor includes two thermal infrared bands: Band 10 and Band 11, which detect longwave radiation emitted by the Earth's surface. Band 10 is particularly essential for surface temperature analysis, helping in applications like monitoring urban heat islands, agricultural water stress, and fire detection.

Landsat 9, launched in 2021, continues the legacy of its predecessors by enhancing the resolution, accuracy, and consistency of Earth observation data. Like Landsat 8, Landsat 9 is equipped with OLI-2 and TIRS-2 sensors. The OLI-2 sensor offers improved radiometric resolution for visible, NIR, and SWIR bands, providing detailed imagery for monitoring land use and vegetation. TIRS-2 further advances thermal data collection, refining the accuracy of land surface temperature measurements by improving stray light correction and calibration processes. This ensures high-quality data for analyzing environmental phenomena, including urban heat islands and vegetation stress[21].

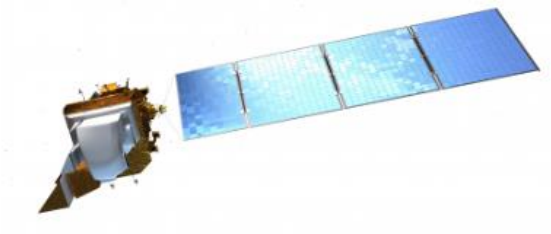

Figure 13. Landsat Satellite

PlanetScope (Planet Labs): PlanetScope, developed by Planet, is a satellite constellation designed to capture high-resolution, daily imagery of the Earth. The constellation comprises numerous small, low-cost, and lightweight satellites, known as "Doves," that operate in low Earth orbit. Each satellite captures images at a spatial resolution of approximately 3 meters per pixel, making PlanetScope ideal for monitoring rapid changes in various applications such as agriculture, urban development, and natural resource management.

With its ability to provide daily revisits over the same area, PlanetScope offers valuable data for tracking short-term events and assessing changes in the landscape. The satellites capture imagery across four spectral bands: red, green, blue, and near-infrared (NIR), making them well-suited for vegetation analysis, land cover classification, and environmental monitoring. The data provided by PlanetScope is often used to complement other satellite systems, offering a cost-effective solution for high-frequency, high-resolution Earth observation.

This accessibility and frequency make PlanetScope especially useful in sectors like precision agriculture, disaster response, and infrastructure monitoring, where timely data is critical[22].

Figure 14. PlanetScope satellite

2.8.2. Radar Platforms

Radar platforms, particularly those equipped with Synthetic Aperture Radar (SAR), use active sensors that emit microwave signals and measure the reflected energy. This capability allows radar systems to capture high-resolution imagery regardless of weather conditions or daylight, making them invaluable for monitoring surface changes, such as land deformation, flood mapping, and vegetation analysis. Unlike optical sensors, radar sensors can penetrate clouds and even vegetation, providing consistent data in challenging environments. Notable SAR platforms include **Sentinel-1 (ESA)** for wide-area monitoring and **RADARSAT (Canada)** for disaster management and agricultural applications.

Sentinel-1 (ESA): Sentinel-1 is a radar satellite equipped with Synthetic Aperture Radar (SAR) technology, providing crucial all-weather, day-and-night Earth observation capabilities. Unlike optical sensors, SAR can penetrate through clouds and work in complete darkness, making it invaluable for regions frequently covered by clouds. This makes Sentinel-1 highly effective in monitoring surface deformation (such as land subsidence and earthquake impacts), flood mapping, and assessing forest cover changes.

Sentinel-1's radar imaging is especially useful for detecting surface changes over time, including movements related to natural hazards like landslides and glacial shifts, or human activities like mining and urban expansion. Its ability to monitor large areas with consistent and frequent revisits is essential for applications like disaster response, environmental monitoring, and infrastructure management[23].

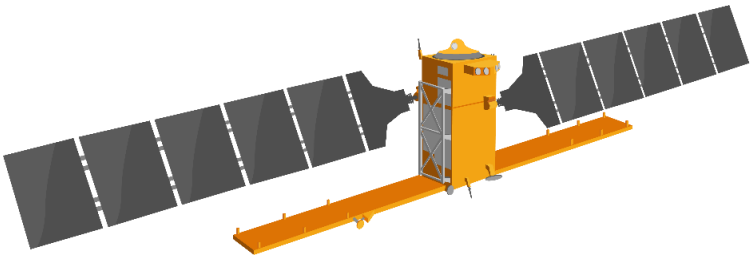

Figure 15. Sentinel-1 (ESA) Satellite

RADARSAT (Canada): RADARSAT is a powerful Synthetic Aperture Radar (SAR) platform designed primarily for all-weather, day-and-night monitoring. Developed by the Canadian Space Agency, RADARSAT specializes in capturing high-resolution radar imagery, which is instrumental in applications such as natural disaster response, ice monitoring, and agricultural management. Its ability to penetrate through clouds and operate in diverse weather conditions makes it particularly useful in regions with frequent cloud cover or for tracking changes in environments such as the Arctic, where it can monitor sea ice dynamics. RADARSAT's C-band radar offers precise surface measurements, which are vital for flood mapping, monitoring deforestation, and tracking soil moisture in agricultural lands.

Like Sentinel-1, RADARSAT's radar capabilities are key to observing surface deformation, flood-prone areas, and changes in forest cover, enabling timely data acquisition even in remote or cloud-covered regions. These features make SAR platforms like RADARSAT crucial for continuous monitoring and managing environmental hazards across the globe[24].

Figure 16. RADASAT Satellite

2.8.3. LiDAR Platforms

LiDAR (Light Detection and Ranging) platforms utilize laser pulses to measure distances and generate highly accurate 3D models of the Earth's surface. As an active sensor system, LiDAR is effective in capturing precise elevation data, making it essential for applications like topographic mapping, forest structure analysis, and urban planning. LiDAR's ability to measure both ground and canopy heights offers unique insights into terrain and vegetation structure. Prominent example include **ICESat-2 (NASA)** for ice sheet elevation and forest canopy measurements.

ICESat-2 (NASA): Using LiDAR technology, ICESat-2 is designed to measure ice sheet elevations, forest canopy heights, and land topography with remarkable precision. LiDAR's ability to generate detailed 3D maps makes it especially effective for accurately mapping elevation and analyzing vegetation structures, offering valuable data for studies on ice loss, deforestation, and terrain modeling.

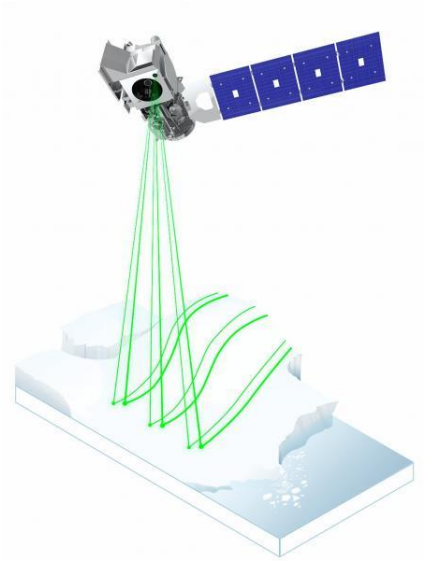

Figure 17. ICESat-2 (NASA) Satellite[25]

Combining data from optical, radar, and LiDAR platforms offers a comprehensive approach to Earth observation. Platforms like Sentinel, Landsat, PlanetScope, and ICESat-2 each have unique strengths that, when integrated, provide a more detailed and accurate understanding of environmental processes, land use changes, and disaster impacts. These technologies together empower better decision-making across sectors like agriculture, forestry, urban planning, and natural resource management.

2.9. Feature Extraction

Feature extraction in remote sensing refers to the process of transforming raw data from satellite images into meaningful information by identifying key characteristics (features) that help differentiate between various land cover types. Instead of using the raw pixel values alone, feature extraction allows the classification process to focus on the most relevant aspects of the data, such as color, texture, or patterns, which help in distinguishing one land cover type from another.

2.9.1. Spectral Features

In remote sensing, bands represent specific portions of the electromagnetic spectrum, such as visible light, near-infrared, or thermal infrared. These bands, often referred to as layers when imported into GIS, can be processed individually or combined to enhance analysis through various band combinations. Common bands include red, green, blue, near-infrared, and thermal infrared, which can be selectively combined depending on the feature of interest in the image. For example, to mimic what the human eye sees, bands in the visible spectrum (red, green, and blue) are combined to create a “true” or “natural” color image[26].

However, to highlight specific surface features, bands outside the visible range can be combined to produce “false” color images. These images are essential for emphasizing characteristics such as healthy vegetation, soil moisture, and heat anomalies. For instance, healthy vegetation reflects strongly in the near-infrared band, while water absorbs most NIR radiation, making these combinations highly effective for identifying vegetation health and differentiating between water bodies and other surfaces. This band combination technique is crucial for analyzing various environmental conditions, leveraging the unique reflectance properties of different materials across the electromagnetic spectrum[27].

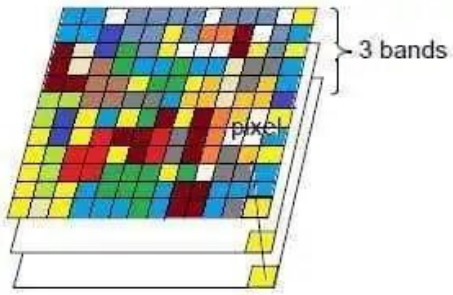

Figure 18. Band Combination

2.9.1.1. Normalized Difference Vegetation Index (NDVI)

To increase the precision of data interpretation, mathematical algorithms are used to create indices, such as NDVI (Normalized Difference Vegetation Index), which are calculated from these bands to highlight specific land cover types, particularly vegetation.

The Normalized Difference Vegetation Index (NDVI) is a widely used index in remote sensing for monitoring and assessing vegetation health and coverage. It leverages the unique spectral reflectance properties of vegetation in the red (RED) and near-infrared (NIR) bands. Vegetation strongly absorbs visible light (RED) for photosynthesis and reflects near-infrared light (NIR). The NDVI uses this difference to quantify vegetation density and health.

The NDVI is calculated using the following formula:

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad 2.1$$

Values of NDVI range from -1 to +1. Higher values indicate healthier and denser vegetation, while lower values may indicate sparse vegetation, bare soil, or non-vegetated surfaces.

While the human eye can only perceive wavelengths within the visible spectrum (390-700 nm), sensors in remote sensing are designed to capture a broader range, including near-infrared (700-1400 nm), where vegetation reflects strongly. This capability allows for more detailed analysis of features like vegetation, where band combinations incorporating near-infrared wavelengths, such as NDVI, are commonly used to assess plant health and classify land cover. The continuous development of new indices expands our understanding of environmental features and processes[28].

2.9.2. Spatial Features

These are derived from the spatial relationships between pixels. For example, built-up areas or urban infrastructure often form regular shapes, while natural landscapes (e.g., forests, rivers) might follow more irregular patterns. The size, shape, and arrangement of pixel groups can give valuable information about land cover types.

2.9.3. Textural Features

Texture describes the variation of pixel intensities in an image, giving information about surface roughness, smoothness, and structure. Texture features are particularly helpful when spectral features alone cannot clearly differentiate between land cover types. The Gray Level Co-occurrence Matrix (GLCM) is a popular technique for texture analysis.

2.9.3.1. Gray Level Co-occurrence Matrix (GLCM)

The Gray-Level Co-Occurrence Matrix (GLCM) is a powerful technique in remote sensing used for texture analysis, particularly in land cover classification.The GLCM is a statistical method that examines the spatial relationship between pixels in an image. It analyzes how often pairs of pixels with specific values (gray levels) occur in a specified spatial relationship, such as adjacent pixels in a given direction (e.g., horizontal, vertical, diagonal). By calculating these relationships, GLCM generates a matrix that describes the texture of the image.

GLCM computes several statistical features from the matrix to describe the texture of an image. Key GLCM features include contrast, which measures the intensity difference between a pixel and its neighbor; correlation, which assesses the predictability of texture based on pixel relationships; energy (or angular second moment), representing the uniformity or smoothness of the texture; and homogeneity, which evaluates how closely the distribution of elements in the matrix aligns with the diagonal, indicating texture consistency.

Haralick et al. (1973) introduced GLCM as a method for extracting second-order statistical texture features from images. These features have been extensively used in remote sensing applications to differentiate between various land cover types[29]. Hall-Beyer (2017) offers a practical tutorial on implementing GLCM, detailing the computation of texture measures and their application in image classification[30]. This resource is valuable for understanding how to extract and utilize GLCM features in eCognition software.

These features are essential in distinguishing land cover types with similar spectral responses but different textures, such as differentiating between dense forests and sparse vegetation or separating urban areas from bare soil.

2.10. Land Cover Classification

Land cover classification is a fundamental process in remote sensing and GIS, enabling the identification and mapping of various land cover and land use types, such as forests, urban areas, water bodies, and agricultural lands, based on spectral and spatial data captured by sensors. This classification has been a key area of research due to its significance in applications like urban

planning, agricultural monitoring, and environmental conservation. Techniques like Supervised Classification, Object-Based Image Analysis (OBIA), and texture analysis using Gray Level Co-occurrence Matrix (GLCM) have been widely employed to improve classification accuracy, particularly in complex environments like urban areas. These methods allow for more precise differentiation of land cover types by integrating spectral information with spatial context and texture[31].

Accurate land cover classification plays a crucial role in understanding and managing global environmental changes. Vitousek [32] in 1994 highlighted that land cover change is a significant driver of ecological system alterations, comparable to climate change. Reliable classification data, combined with advanced techniques, is essential for assessing and mitigating these impacts. By integrating these approaches with climate data, researchers can better analyze the relationship between land cover changes and environmental factors, providing insights critical for sustainable development and conservation.

In land cover classification, two main approaches are used: pixel-based and object-based classification. These can be further divided into supervised and unsupervised methods. The following sections will explain both approaches and their variants.Each approach has distinct methodologies and applications depending on the nature of the data and the specific goals of the analysis.

2.10.1. Pixel-Based And Object-Based Classification

Pixel-based classification focuses on analyzing individual pixels based solely on their spectral characteristics. While this approach has been widely used, it faces limitations when dealing with high-resolution data such as imagery from Unmanned Aerial Vehicles (UAVs) or high-resolution satellites. UAVs, which are remotely controlled or autonomous aerial devices, provide high spatial resolution data and allow for flexible flight planning and data acquisition, even in hard-to-reach or hazardous areas. However, pixel-based techniques struggle to handle the rich details available in this high-resolution imagery, often leading to inconsistent classification results and difficulty in extracting complex objects of interest.

In contrast, Object-Based Image Analysis (OBIA) is more effective for classifying high-resolution data. OBIA considers not only spectral characteristics but also the spatial and textural properties of

pixels. Instead of classifying individual pixels, OBIA begins with a segmentation process that groups neighboring pixels into meaningful objects based on their shapes, textures, and spectral information. This method enables the classification of homogeneous image objects rather than individual pixels, which significantly improves accuracy, especially in environments where detailed information is critical. OBIA is particularly beneficial when working with high-resolution data, making it a superior approach compared to pixel-based classification for extracting detailed features[33].

2.10.2. Supervised And Unsupervised Classification

Supervised classification involves the use of sample data provided by an analyst to guide the classification algorithm. The user selects sample pixels representing various land cover types, which are then used to "train" the algorithm. The algorithm identifies similar pixels across the entire image, categorizing them into specific land cover classes. Supervised classification requires initial input from the analyst and offers control over the classification process, ensuring that the results align with known land cover types. This method is commonly used for detailed classifications, as it can incorporate both spectral and ancillary data to improve accuracy, making it highly effective for well-defined categories.

Unsupervised classification, on the other hand, does not require input from the analyst prior to running the algorithm. Instead, the algorithm groups similar pixels into clusters, and the analyst labels these clusters afterward. This method allows the computer to autonomously create initial classifications, making it useful when the land cover types are unknown or when the goal is to identify natural groupings in the data. Unsupervised classification can be more exploratory, but it requires post-classification interpretation to label the clusters correctly. Both methods are widely used in remote sensing, with supervised classification offering more precision and control, while unsupervised classification is useful for discovering natural groupings without prior knowledge[34].

2.11. Classification algorithem

Classification algorithms in remote sensing can be broadly divided into parametric and non-parametric methods, many of which fall within the realm of machine learning. Machine learning is an advanced field of artificial intelligence that focuses on developing algorithms that can learn from and make predictions based on data. In remote sensing, machine learning has revolutionized land

cover classification by improving accuracy and efficiency. It includes a variety of algorithms, from simpler approaches to more complex ones, such as Decision Trees, Random Forests, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN)[35].

Parametric algorithms, like Maximum Likelihood Classification (MLC), rely on the assumption that data follows a particular statistical distribution, such as the normal distribution. While effective in simpler, controlled environments, they struggle with complex or heterogeneous datasets, making them less suitable for diverse land cover types. On the other hand, non-parametric algorithms, commonly used in machine learning, do not assume any specific distribution and thus offer greater flexibility. This makes them highly effective in handling complex and diverse datasets found in remote sensing applications.

Machine learning methods, particularly non-parametric algorithms, offer powerful solutions for land cover classification. Decision Trees, for example, split data based on feature thresholds to classify different land cover types. Random Forest improves upon this by building multiple decision trees and combining their results, which enhances accuracy and reduces overfitting. Support Vector Machines (SVM) excel in high-dimensional spaces, identifying the optimal hyperplane to separate land cover classes. Similarly, K-Nearest Neighbors (KNN) classifies pixels based on their proximity to other classified data points.

Machine learning methods, particularly non-parametric algorithms, offer powerful solutions for land cover classification. Decision Trees, for example, split data based on feature thresholds to classify different land cover types. Random Forest improves upon this by building multiple decision trees and combining their results, which enhances accuracy and reduces overfitting. Support Vector Machines (SVM) excel in high-dimensional spaces, identifying the optimal hyperplane to separate land cover classes. Similarly, K-Nearest Neighbors (KNN) classifies pixels based on their proximity to other classified data points[36].

Among the various algorithms, the Random Forest (RF) classifier has gained popularity for its robustness and accuracy. Breiman (2012) introduced RF as an ensemble learning method that constructs multiple decision trees and aggregates their predictions. This method has been applied successfully in remote sensing for land cover classification due to its resilience to overfitting and capacity to process high-dimensional data efficiently[37].

Ultimately, these algorithms—whether Decision Trees, Random Forest, SVM, or KNN—are integral parts of the machine learning toolkit. Machine learning includes both basic and complex models that can handle intricate patterns in remote sensing data, offering versatile and robust solutions for land cover classification. The adaptability and precision of these methods make them essential in modern remote sensing, where the ability to process and classify data with high accuracy is increasingly vital.

2.12. Accuracy Assessment

Accuracy assessment is a critical step in evaluating the effectiveness of land cover classification, ensuring that the results are both reliable and valid. This process involves the use of metrics like precision, recall, F1 score, and overall accuracy to measure how well the classification has performed. Confusion matrices are often generated to provide a detailed analysis, highlighting the model's ability to correctly classify different land cover types while identifying potential errors. These metrics, combined with others like the kappa coefficient, offer a comprehensive overview of the classification's performance and allow for meaningful comparisons across different approaches.

The emphasis on accuracy assessment in land cover classification highlights its importance in producing reliable maps that can be confidently used for various applications, from urban planning to environmental management.

Foody (2010) reviews various accuracy assessment methodologies, emphasizing the importance of ground truth data in validating remote sensing classifications[38]. This is particularly relevant for the thesis, as ground truth data will be necessary to assess the accuracy of classifications using OBIA, GLCM, and spectral data.

2.12.1. Confusion Matrix

A confusion matrix is a useful tool for evaluating the performance of a classification algorithm by providing a clear summary of its predictions compared to the actual reference data. The matrix displays the number of correct and incorrect classifications made by the model. It is structured to show four key outcomes: true positives (TP), where positive observations are correctly classified; true negatives (TN), where negative observations are correctly identified; false positives (FP), where negative observations are incorrectly classified as positive; and false negatives (FN), where positive

observations are incorrectly labeled as negative. These outcomes give a comprehensive overview of how well the model is performing across all possible scenarios. Congalton and Green (2009) emphasize the importance of the confusion matrix in evaluating the performance of classification algorithms. The overall accuracy is calculated as the ratio of correctly classified instances to the total number of instances, providing a straightforward measure of classification effectiveness[39].

By analyzing the confusion matrix, various accuracy metrics can be derived, such as precision, recall, F1 score, and overall accuracy. These metrics help in assessing not only how often the model is correct but also how it handles specific types of errors, such as false positives or false negatives. The insights gained from the confusion matrix are crucial for understanding the strengths and weaknesses of the classification model, guiding further refinement and improvement.

2.12.2. Precision

Precision, also known as the positive predictive value, measures the accuracy of the positive predictions made by the model. It is defined as the ratio of true positive predictions to the total number of positive predictions (both true positives and false positives).

$$precision = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad 2.2$$

A high precision score indicates that the model has a low false positive rate, meaning it rarely misclassifies non-vegetation areas as vegetation, while a low precision score suggests that the model frequently misclassifies non-vegetation areas as vegetation.

2.12.3. Recall

Recall, also known as sensitivity or the true positive rate, measures the model's ability to identify all relevant positive cases. It is defined as the ratio of true positive predictions to the total number of actual positive instances (both true positives and false negatives).

$$Recall = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad 2.3$$

A high recall score indicates that the model successfully identifies most vegetation areas, resulting in few missed detections (false negatives), whereas a low recall score suggests that the model frequently misses actual vegetation areas, leading to a higher number of false negatives.

2.12.4. F1 score

The F1 score is the harmonic mean of precision and recall. It provides a single metric that balances both precision and recall, especially useful when the class distribution is imbalanced.

$$F1\ score = 2 * \frac{\text{precision} * \text{Recall}}{\text{precision} + \text{Recall}} \quad 2.4$$

2.12.5. kappa statistic

The kappa statistic, often referred to as Cohen's kappa, is a valuable metric used in classification studies to assess the accuracy of a model or method. Unlike simple overall accuracy, which only measures the proportion of correctly classified instances, kappa provides a more nuanced evaluation by considering the agreement between the predicted classifications and the reference data (often called "ground truth") while also accounting for the possibility of chance agreement.

Kappa is a measure of the degree of agreement between two raters (or between a classification method and reference data in remote sensing) beyond what would be expected by random chance. The kappa value ranges from -1 to 1, with 1 indicating perfect agreement, 0 suggesting agreement no better than chance, and negative values indicating worse-than-chance agreement, suggesting systematic disagreement.

The kappa statistic is particularly useful in scenarios where the data is imbalanced, meaning some classes are more prevalent than others, which can lead to misleading accuracy measures. By correcting for these imbalances, kappa offers a more reliable indication of classification performance.

Precision and recall are essential metrics derived from the confusion matrix that provide insights into the performance of individual classes. Precision, also known as positive predictive value, measures the proportion of true positive predictions among all positive predictions made by the model. Recall, or sensitivity, assesses the proportion of true positives correctly identified by the model. The F1 score,

which is the harmonic mean of precision and recall, offers a balanced measure, especially useful in imbalanced datasets.

Recent advancements in accuracy assessment techniques have focused on improving the reliability and interpretability of the results. Pontius and Millones [40] in 2011 introduced the concept of Quantity Disagreement and Allocation Disagreement as complementary metrics to the confusion matrix, providing a more detailed analysis of classification errors. These metrics help to distinguish between errors due to quantity differences and spatial misallocations, offering deeper insights into the classification performance.

2.13. Land Surface Temperature (LST)

Surface temperature, a critical metric derived from remote sensing data, represents the thermal characteristics of the Earth's surface. It plays a significant role in understanding environmental changes, urban heat islands, and land-use dynamics. Land Surface Temperature (LST) is commonly extracted from satellite imagery, such as Landsat data, which records thermal emissions in specific spectral bands. However, these thermal bands do not directly provide temperature values; instead, they measure the intensity of radiation emitted from the surface as digital numbers (DN).

To convert these digital numbers into meaningful surface temperature values (in Kelvin), a series of mathematical transformations is required. The process typically involves:

1. **Converting DN to Radiance:**

The digital numbers from the satellite sensor are converted into Top-of-Atmosphere (TOA) radiance using the formula:

$$L = M * Q + A \quad (2.5)$$

Where:

- L is the TOA radiance (in Watts/meter^2/steradian/μm)
- M is the radiance multiplier (gain), a scaling factor available in the satellite metadata.
- Q is the digital number (DN) for a given pixel.
- A is the radiance add (offset), also provided in the metadata.

The radiance multiplier (M) and radiance add (A) can be found in the Landsat image metadata file. This step transforms the raw digital numbers into physical radiance values that correspond to the energy emitted from the Earth's surface.

2. **Radiance to Temperature in Kelvin:**

The derived TOA radiance is then converted to temperature using Planck's radiation law:

$$T = \frac{K_2}{\ln\left(\frac{K_1}{L} + 1\right)} \quad 2.6$$

Where:

- T is the temperature in Kelvin.
- K_1 and K_2 are calibration constants specific to the Landsat sensor, provided in the metadata.
- L is the TOA radiance (calculated in the previous step).

Using these equations, the LST for each pixel can be calculated, providing a spatial representation of surface temperature across the area of interest.

Guha and Govil (2020) explored the relationship between LST and the Normalized Difference Vegetation Index (NDVI) in a tropical city. Their study showed a consistent inverse relationship, where areas with higher vegetation density (indicated by higher NDVI values) exhibited lower surface temperatures. This correlation underscores the role of vegetation in mitigating surface heating, particularly in urban settings where impervious surfaces dominate[41].

Their findings, derived from analyzing satellite data across different seasons, revealed the importance of vegetation in regulating urban thermal environments. In dry seasons, sparse vegetation amplified surface temperatures, while in wetter seasons, increased vegetation coverage moderated thermal extremes.

In conclusion, the literature review highlights the diverse approaches and advancements in satellite sensors, data types, and classification methods used in land cover analysis. From optical and radar technologies to techniques like supervised classification, OBIA, and GLCM, these tools and methods provide a comprehensive framework for accurate and reliable land cover mapping. The integration of

various data sources and classification strategies plays a crucial role in understanding and monitoring complex landscapes.

With this foundation, the next chapter will introduce the study area, Tools and the data utilized for the analysis, setting the stage for the application of these methods in a real-world context.

3. Study Area,Tools and Data

The following chapter provides the specifics of the study area, the data characteristics, and the tools that facilitated the classification and analysis process.

3.1. Description of Study Area

Ouagadougou, the capital city of Burkina Faso, is situated in the central part of the country. It lies approximately between latitudes 12.36°N and longitudes 1.53°W(Figure 19). The city covers an area of about 219.3 square kilometers, making it one of the largest urban centers in Burkina Faso. Ouagadougou is a pivotal hub for political, cultural, and economic activities in the country.

The city is located in a relatively flat region, characterized by its low-lying topography with an average elevation of around 305 meters above sea level. This geographical setting places Ouagadougou within the Sudanian savanna zone, which transitions between the Sahelian and more humid savanna regions to the south. The terrain is predominantly composed of sandy and clayey soils, which influence both the urban infrastructure and the natural vegetation of the area.

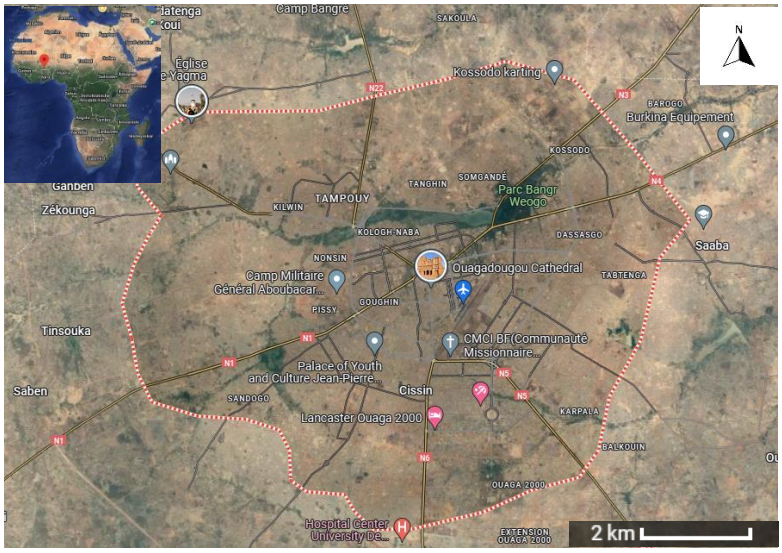

Figure 19. Ouagadougou, the capital city of Burkina Faso

Ouagadougou experiences a tropical wet and dry climate, classified as Aw under the Köppen-Geiger climate classification system. This climate is characterized by a distinct wet season and a dry season, each contributing uniquely to the environmental conditions of the city.

The wet season typically spans from June to September, with August being the peak month for rainfall. During this period, Ouagadougou receives an average annual precipitation of about 800-900 millimeters. These rains are essential for agricultural activities and contribute to the replenishment of local water resources.

Conversely, the dry season extends from October to May, with the Harmattan winds significantly influencing the climate from December to February. These winds, originating from the Sahara Desert, bring dry and dusty conditions, reducing humidity levels and sometimes causing visibility issues. Temperatures during the dry season can be quite extreme, often ranging from 25°C to 40°C, with the hottest months being March and April.

The climate of Ouagadougou plays a critical role in shaping the city's environment and influencing land cover patterns. The seasonal variability in rainfall and temperature affects vegetation growth, water availability, and urban development, making it a key factor to consider in land cover classification studies.

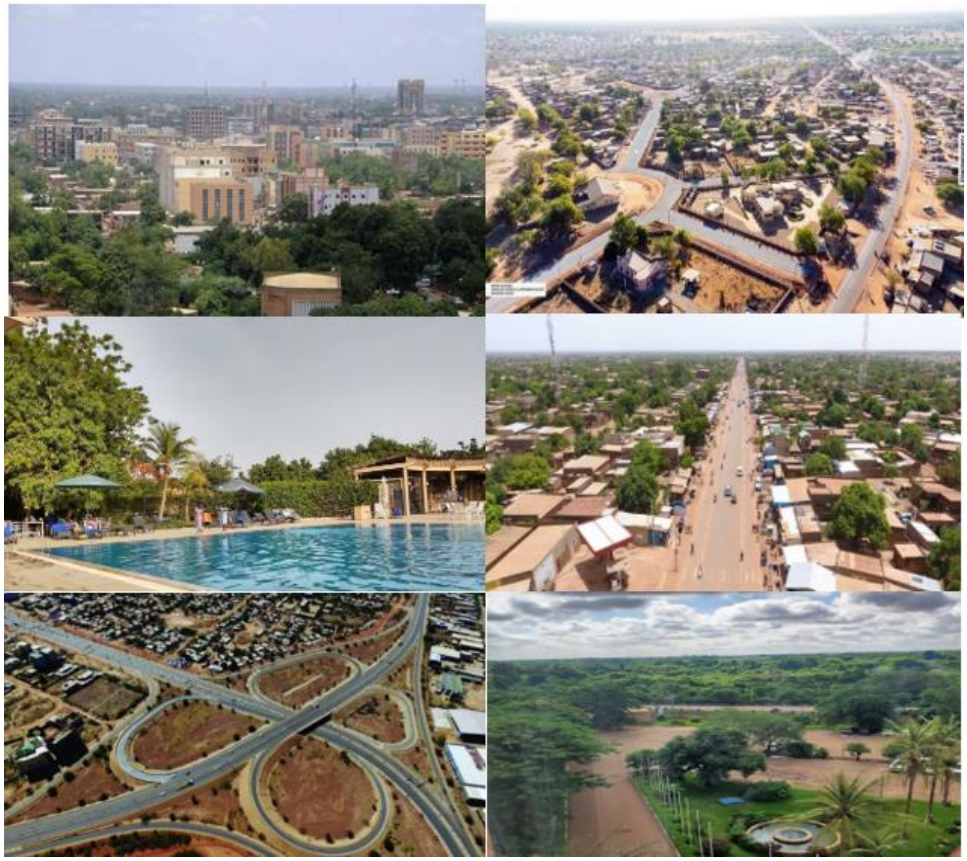

Figure 20. different part of the city

3.2. Tools

This section introduces the key tools used for data processing, classification, and analysis in this study.

3.2.1. QGIS

QGIS is an open-source Geographic Information System (GIS) software that provides tools for viewing, editing, and analyzing geospatial data. It supports a wide range of data formats and offers numerous plugins for specialized analyses. In this study, QGIS was used to compute NDVI values, create histograms, reclassify data, and assess classification accuracy. The integration of QGIS with various remote sensing and analysis tools made it possible to efficiently process and classify land cover data in Ouagadougou.

3.2.1.1. Raster Calculator

The Raster Calculator in QGIS is an essential tool for performing mathematical operations on raster datasets. In this study, it was used to calculate both vegetation indices and surface temperature from satellite imagery.

For vegetation analysis, the Raster Calculator was applied to compute the Normalized Difference Vegetation Index (NDVI), which is derived from the red and near-infrared (NIR) bands according to Formula 2.1 as described in Chapter 2. NDVI helps identify vegetation density, with higher values representing dense vegetation and lower values indicating bare soil or other non-vegetated surfaces.

In addition to NDVI, the Raster Calculator was also used to process thermal band data to calculate Land Surface Temperature (LST). This involved converting the Digital Number (DN) values to TOA Radiance and then using the appropriate formulas to derive the surface temperature, as outlined in Chapter 2 (Formula 2.5 & 2.6). These steps were key in preparing the data for analyzing the relationship between vegetation cover and surface temperature.

By using the Raster Calculator for these calculations, this study was able to extract meaningful insights about vegetation and temperature patterns in the study area.

3.2.1.2. Raster Histogram

Histograms are graphical representations of the distribution of pixel values in a raster image. In this study, the histogram was utilized to understand the distribution of NDVI values across the study area. By examining the peaks and ranges in the histogram, thresholds for different land cover classes were determined. The histogram revealed the frequencies of specific NDVI values, which corresponded to different land cover types, such as water, bare soil, vegetation, and buildings.

The histogram tool was used to visualize the NDVI layer and identify the spread of pixel values. The peaks and breaks in the graph were analyzed to define threshold values that separate different classes, such as vegetation, water, and bare soil. This allowed for a more informed choice of classification ranges, as pixel value frequency could indicate where significant land cover transitions occurred.

3.2.1.3. Reclassify by Table

The Reclassify by Table tool in QGIS allows the user to categorize raster data by assigning new values to specific ranges of pixel values. This tool was used to classify NDVI values into distinct land cover categories, such as water, bare soil, vegetation, and buildings, based on predefined thresholds derived from the histogram.

These thresholds are input into the Reclassify by Table tool, where specific ranges of NDVI values were assigned to different classes (e.g., water, vegetation, bare soil). This created a reclassified raster map, where each pixel was categorized into one of the six land cover classes.

3.2.1.4. Semi-Automatic Classification Plugin (SCP)

The Semi-Automatic Classification Plugin (SCP) for QGIS is an advanced tool designed to facilitate both supervised and unsupervised classifications of remote sensing imagery. It streamlines many steps in the classification process, making it easier to handle large datasets and perform various analyses, including accuracy assessments. SCP's user-friendly interface and automated workflows make it a valuable tool for researchers dealing with complex classification tasks.

Several key features were utilized in this study. One of the primary functions was accuracy assessment, where SCP was used to compute classification accuracy by comparing the reclassified map to reference data. This involved generating confusion matrices, which provided metrics such as

overall accuracy, precision, recall, and the kappa coefficient. Additionally, for supervised classifications, SCP allowed for the manual selection of training samples to represent different land cover types. These samples served as the basis for classifying the imagery, ensuring a more targeted and reliable classification process.

3.2.1.5. Zonal Statistics

The Zonal Statistics tool in QGIS is a useful feature for analyzing raster data within the boundaries of a vector layer. It provides summary statistics such as sum, count, and mean values for raster data, calculated for each defined zone in the vector layer. These zones can include administrative boundaries, land use polygons, or sampling grids.

In this study, the Zonal Statistics tool was used to calculate vegetation percentages by summing and counting NDVI values and to determine mean surface temperature for the defined zones. These results were instrumental in analyzing the correlation between vegetation density and surface temperature across the study area.

3.2.2. eCognition

eCognition is a powerful software platform primarily designed for Object-Based Image Analysis (OBIA). It allows for the segmentation and classification of high-resolution satellite imagery by combining spectral, spatial, and contextual information. This advanced tool is particularly effective in environments where pixel-based methods struggle, providing a more detailed and context-aware approach to image analysis and making it suitable for handling complex remote sensing datasets.

The software's ability to integrate machine learning, rule-based approaches, and object-based segmentation provides flexibility and precision in classifying complex areas like urban environments, vegetation, and water bodies.

In eCognition, various tools and workflows are used to manage the classification process efficiently.

3.2.2.1. Process Tree

The process tree is a central feature of eCognition where users can define and manage the sequence of algorithms applied to the dataset. It allows for building workflows in a structured and logical

manner. Each step in the classification process, from segmentation to feature extraction, classification, and accuracy assessment, is included in the process tree, providing an overview of the entire workflow. Users can easily adjust parameters and test different methods by modifying the process tree.

3.2.2.2. Class Hierarchy

The class hierarchy defines the different land cover or land use types that will be classified in the image. This can include water, bare soil, buildings, roads, vegetation, and sparse vegetation. Each class is assigned a specific label and color, making it easier to organize and visualize the results. The hierarchy helps in structuring the classification process and understanding the relationships between different classes.

3.2.2.3. View Features

The "View Features" tool in eCognition allows users to apply specific features such as NDVI or GLCM (Gray-Level Co-Occurrence Matrix) to improve classification accuracy. For example, NDVI helps to distinguish between vegetation and non-vegetation areas, while GLCM provides textural information that can help differentiate between objects with similar spectral properties, such as bare soil and buildings. By applying these features, users can refine their classification and improve accuracy.

3.2.2.4. Image Object Information

This tool provides detailed numerical information on the segmented objects and the features applied to them. For instance, after applying NDVI or GLCM features, users can examine the values of those features for specific objects in the image. This helps to verify that the features are working as expected and provides insight into how the features are influencing classification.

Each of these components plays a crucial role in the object-based image analysis workflow within eCognition, allowing for the integration of spectral, spatial, and textural information to produce more accurate land cover classifications. In this study, eCognition was used to perform classifications based on NDVI thresholds, GLCM, and supervised classification with Random Tree.

3.3. Data Collection

The dataset used in this project was sourced from PlanetScope, a high-resolution satellite imagery platform operated by Planet Labs. The PlanetScope imagery offers a spatial resolution of 3 meters per pixel, which is ideal for detailed land cover classification, particularly in urban environments and areas with complex land use patterns. The imagery was captured on the 10th of May, ensuring that the data reflects the conditions during the late dry season, which is crucial for identifying distinct land cover features. The PlanetScope data consists of three spectral bands within the visible spectrum, specific wavelengths correspond to the colors red, green, and blue (RGB).

Table 1. Vegetation and Land Cover Data from PlanetScope Imagery

Attribute	**Description**	**PlanetScope Data**
File Name	The name of the .tif file	ouagadougou_image_2024.tif
Satellite Name	The name of the satellite or sensor	PlanetScope
Acquisition Date	The date when the image was captured	10/05/2024
Spatial Resolution	The ground sampling distance, or pixel size	3 meters per pixel
Number of Bands	The number of spectral bands included in the image	4 bands: Red, Green, Blue, NIR
Coordinate Reference System (CRS)	The projection system used for the image	EPSG:32630 (WGS 84 / UTM zone 30N)
File Format	The format of the data file	.tif
Sensor Type	The type of sensor used (if specified)	Optical
Data Provider	The organization or company that provided the data	Planet Labs

This high-quality dataset allowed for precise differentiation between land cover classes such as water, bare soil, buildings, roads, sparse vegetation, and dense vegetation, which are key to understanding the study area's spatial composition. The choice of this dataset aligns with the previously discussed concepts of high-resolution, multi-spectral imagery highlighted in the literature review, offering a solid foundation for the classification and analysis performed in this project.

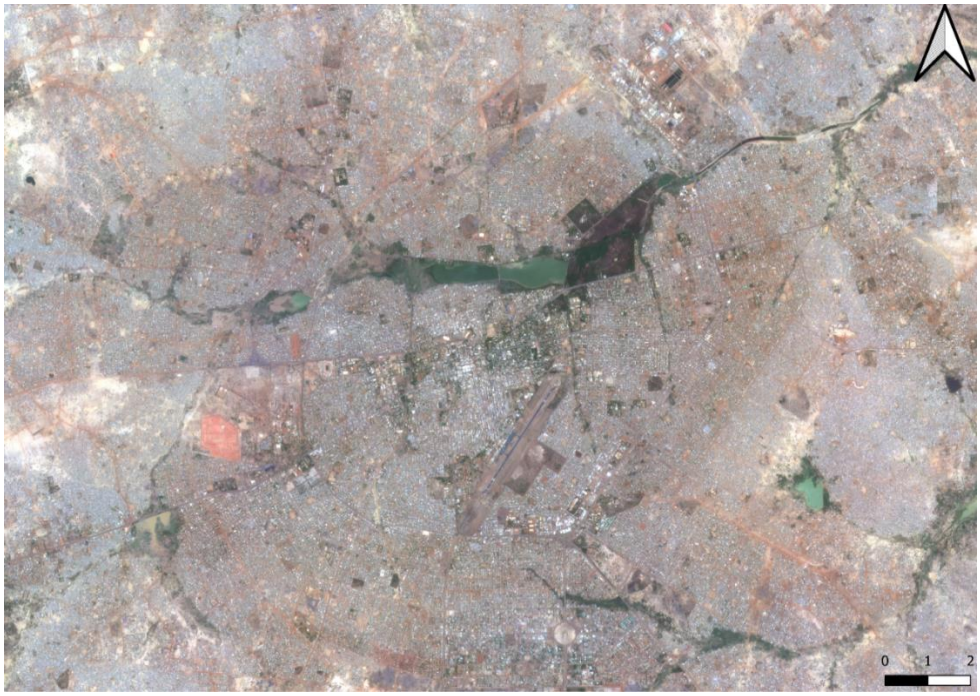

Figure 21.True color visualization of Ouagadougou city in 4 bands of RGB and NIR

In addition to the PlanetScope dataset used for vegetation analysis, this project also utilized Landsat 9 Level-2 Surface Reflectance (LC09_L2SP_195051_20230524_20230601_02_T1) data for calculating surface temperature. While the PlanetScope imagery was acquired on the 10th of May 2024, the Landsat 9 image was selected from 24th of May 2023, as it was the closest available imagery with minimal cloud cover (0.67%) to ensure clear and reliable data. The time gap between the datasets is acceptable for this analysis, as it allows for a meaningful comparison of vegetation percentages derived from the PlanetScope image with surface temperature data from the Landsat 9 image.

Landsat 9 provides medium-resolution multispectral imagery with a spatial resolution of 30 meters per pixel, suitable for analyzing broader environmental patterns such as surface temperature variability. For LST analysis, Band 10, corresponding to the thermal infrared (TIRS) range, was used. This band's radiometric calibration coefficients, specifically the Radiance Multiplier (ML) and

Radiance Add (AL), were extracted from the metadata file to convert raw digital number (DN) values into Top of Atmosphere (TOA) Radiance. Subsequently, the TOA Radiance was converted to temperature in Kelvin using pre-defined thermal constants K1 and K2, also provided in the metadata.

Both datasets share the same Coordinate Reference System (CRS): WGS 84 / UTM Zone 30N (EPSG:32630), ensuring spatial consistency and enabling comparative analysis. The integration of high-resolution PlanetScope imagery with Landsat 9 thermal data facilitated the study of vegetation and temperature correlation.

Table 2. Surface Temperature Data from Landsat 9 Imagery

Attribute	Description	Landsat 9 Data
File Name	The name of the .tif file	LC09_L2SP_195051_20230524_20230601_02_T1.tif
Satellite Name	The name of the satellite or sensor	Landsat 9
Acquisition Date	The date when the image was captured	24/05/2023
Spatial Resolution	The ground sampling distance, or pixel size	30 meters per pixel
Cloud Cover	The percentage of cloud cover in the scene	0.67%
Number of Bands	The number of spectral bands included in the image	11 bands total (Band 10 used for surface temperature)
Coordinate Reference System	The projection system used for the image (CRS)	EPSG:32630 (WGS 84 / UTM Zone 30N)
File Format	The format of the data file	.tif
Sensor Type	The type of sensor used	Thermal Infrared (TIRS) for Band 10
Radiance Multiplier (ML)	Calibration coefficient for converting DN to TOA Radiance	3.8 *10^-4
Radiance Add (AL)	Offset value for TOA Radiance calculation	0.1
K1 Constant	Thermal constant for converting radiance to temperature in Kelvin	799.0284
K2 Constant	Thermal constant for converting radiance to temperature in Kelvin	1329.2405

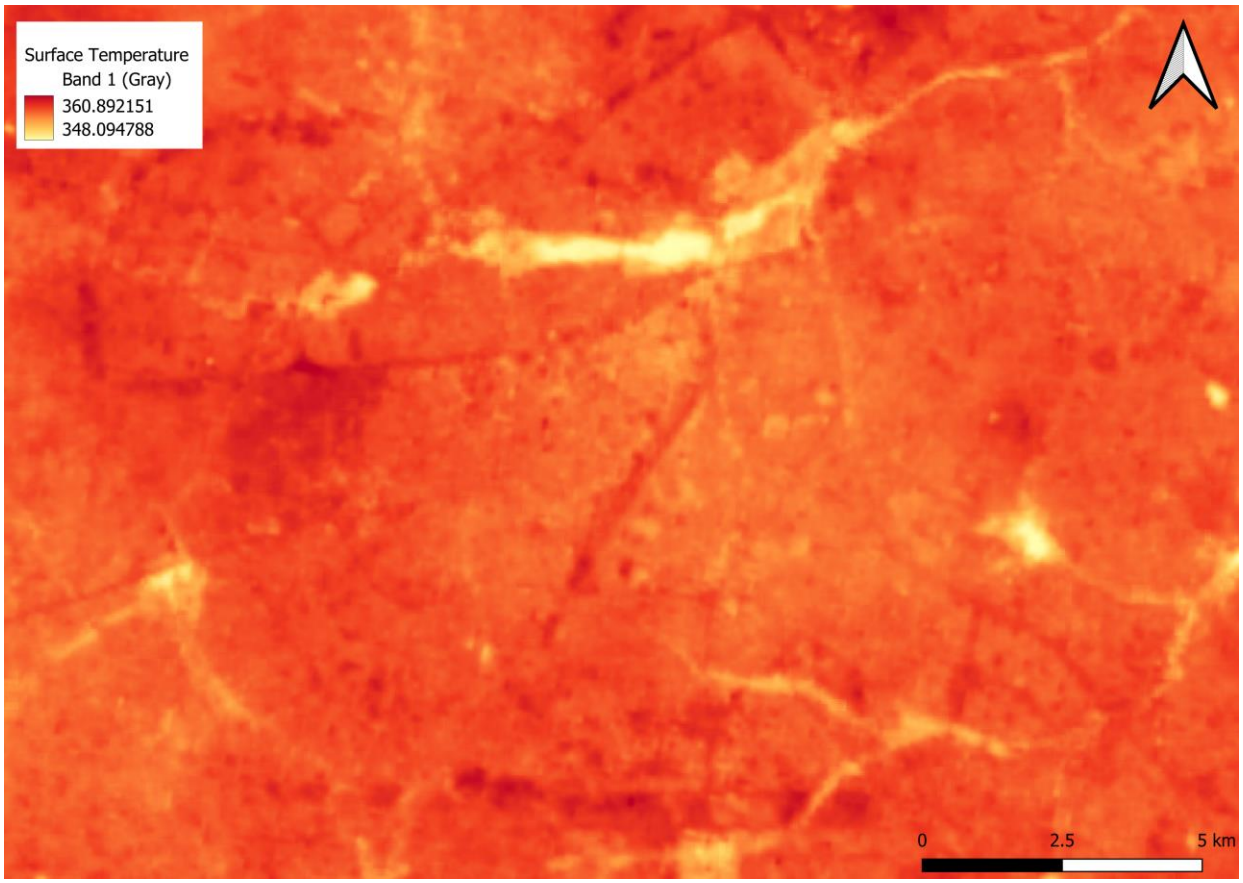

Figure 22. Raw Surface Temperature (Colorized)

In this study, RGB and NIR data were used to classify and assess the accuracy of various land cover types. By analyzing the spectral signatures and using specific algorithms, we can better understand vegetation health and other environmental factors. Additionally, surface temperature data were incorporated to explore its relationship with vegetation cover, providing deeper insights into the thermal and ecological dynamics of the study area.

4. Methodology

This chapter outlines the tests conducted for land cover classification in Ouagadougou, using both QGIS and eCognition as shown in the table 3. The process began in QGIS by creating NDVI thresholds, which were then applied in eCognition for further analysis through Object-Based Image Analysis (OBIA). The chapter details different testing scenarios that were explored using both software platforms.

Various tests used initial and refined NDVI thresholds. In eCognition, these thresholds were combined with texture analysis techniques, such as the Gray Level Co-occurrence Matrix (GLCM), to better distinguish between similar-looking areas. Additionally, a machine learning approach using the Random Forest algorithm was applied to enhance land cover classification based on manually selected training samples.

Each method's accuracy was evaluated by comparing results using confusion matrices in QGIS, to determine the most effective classification technique. This overview introduces the diverse tests performed, setting the stage for a detailed discussion on each method's application and results.

In addition to classification, surface temperature data was analyzed to explore its relationship with vegetation density. By using NDVI and thermal bands from satellite imagery, correlations between vegetation percentages and surface temperature were examined. This analysis aimed to provide insights into how vegetation impacts local temperature patterns, particularly in urbanized and sparsely vegetated areas.

Table 3. Classification Tests Overview in QGIS and eCognition

Software	Method	Test	Name of Test	Descriotion
QGIS	Computing NDVI	**Test 1**	First NDVI Thresholds	Used 'Reclassify by Table' for initial NDVI thresholds to differentiate land covers.
		Test 2	Second NDVI Thresholds	Used 'Reclassify by Table' for Refined NDVI thresholds to differentiate land covers.
eCognition	**Object Based Image Analysis** (Image segmentation)	**Test 3**	First OBIA Classification	Implements first NDVI thresholds using OBIA
		Test 4	First OBIA Classification _GLCM	Combines second NDVI thresholds with GLCM contrast in all directions for green layer
		Test 5	Second OBIA Classification	Implements Second NDVI thresholds using OBIA
		Test 6	Second OBIA Classification_GLCM	Combines second NDVI thresholds with GLCM contrast in all directions for green layer
		Test 7	Supervised Classification	Training sample selection, Executes supervised classification using machine learning, Random Forest 'classifier'

4.1. QGIS Software (TEST 1 & 2)

4.1.1. Computing NDVI

After loading the TIFF data file, high-resolution satellite imagery with RGB and NIR bands in QGIS, the NDVI was computed using the Raster Calculator tool. This allowed for the combination of the red and NIR bands according to the NDVI formula previously outlined in the literature review (Formula 2.1). The resulting NDVI values ranged from -1 to 1, where higher values indicated denser vegetation, and lower values represented non-vegetation features like water or urban infrastructure.

Once the NDVI layer was generated, the Properties option was accessed to view detailed metadata and other relevant information, such as pixel values and statistics, providing further insights into the distribution and characteristics of NDVI values across the study area.

4.1.2. calculate NDVI thresholds

In the process of calculating NDVI Thresholds for land cover classification, six main land cover classes were considered: water, bare soil, buildings, roads, sparse vegetation, and dense vegetation. two different methods were used to calculate NDVI thresholds. The first method **(Test1)** relied on the manual selection of threshold values by using the "Identify Feature" tool and involved trial and error to distinguish between the six classes. The second method **(Test2)**, however, utilized the histogram tool in QGIS, which allowed for a more systematic approach in identifying the threshold values, particularly for water and vegetation classes. The results of these different threshold approaches will be discussed in more detail in the next chapter.

4.2. eCognition Software Preparation (Common Steps for Tests 3 to 7)

4.2.1. Loading the Data

To perform Object-Based Image Analysis (OBIA) using eCognition, the process begins by importing the high-resolution imagery and adding the computed NDVI as a new layer. Once the data is loaded, the NDVI layer is integrated as part of the feature space, allowing it to be used in the classification process. To do this, NDVI is calculated in QGIS and then imported into eCognition.

4.2.2. Object-Based Image Analysis (OBIA) Segmentation

The first step is to segment the image using the "multiresolution segmentation" algorithm available in the "Process Tree" of eCognition. This algorithm groups pixels into meaningful objects based on both spectral and spatial characteristics, such as color, shape, and texture. For this project, specific parameters were set to optimize the segmentation process: the shape factor was set to 0.1 and compactness to 0.9. These values were chosen to prioritize compact, well-defined segments with minimal emphasis on shape, allowing the algorithm to focus more on the spectral homogeneity of the objects rather than their geometric properties.

This approach helps in accurately reflecting the various land cover types, particularly in areas where the spectral characteristics are more important than the shape, such as in the classification of vegetation.

To ensure a consistent basis for comparison across all five classification methods, the same image segmentation process was used.

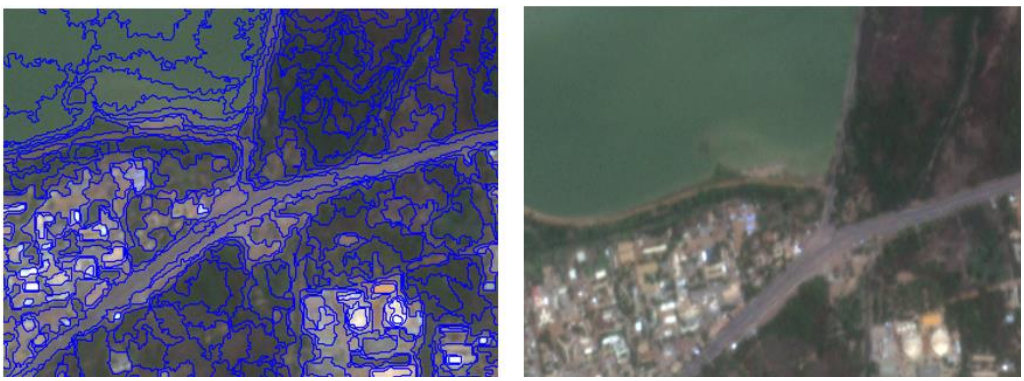

Figure 23. Example of segmentations

4.2.3. Class Hierarchy

After segmentation, one of the essential steps in eCognition is defining the "Class Hierarchy". This step is crucial because it provides a structure that organizes the different land cover classes into categories. In this study, the "Class Hierarchy" was created to define six land cover classes: water, bare soil, buildings, roads, sparse vegetation, and dense vegetation.

By defining the hierarchy, it ensures that the software can classify each object (created during segmentation) into one of these predefined classes. The "Class Hierarchy" acts as a framework where all subsequent classifications, whether based on NDVI thresholds or texture analysis using GLCM, are aligned with the defined classes.

4.3. eCognition Software Specific Analyses

4.3.1. Assigning Classes (For TEST 3-6)

After segmentation and defining the Classes, the classification process proceeds by utilizing the "Assign Class" algorithm in the "Process Tree" of eCognition. The segmented objects are classified based on the NDVI thresholds that were previously determined in QGIS. Given that two sets of NDVI thresholds were developed during the study, this classification process was carried out twice—once for each set of thresholds **(Test 3 & 5)**. The classes defined for both classifications include water, bare soil, buildings, roads, sparse vegetation, and dense vegetation.

In addition to NDVI-based classification, GLCM contrast was used to further refine the classification results. To incorporate GLCM, texture analysis was applied through the Feature View and Texture tools in eCognition. GLCM contrast was calculated for all directions and applied to the green layer. The choice of the green layer was due to its ability to capture more fine details of texture, while the use of all directions in the GLCM analysis helped ensure that the classification accounted for textural variations from different orientations, enhancing accuracy in complex urban environments.

For each class, the NDVI thresholds were entered into the "Condition" icon within the "Assign Class" algorithm. Additionally, for classes where textural information was relevant, a second "Assign Class" algorithm was added, incorporating the defined GLCM thresholds as an extra condition **(Test 4 & 6)**.

The GLCM thresholds were established based on an analysis of different regions in the image, observing texture variations that best corresponded to each land cover class.

This combined approach of applying NDVI and GLCM thresholds produced two sets of classification results. These will be compared in the results chapter to assess the impact of incorporating textural data into the classification process.

To improve the classification results, the "Merge" algorithm is used to combine adjacent segments of the same class. This step is particularly useful for reducing the number of small, fragmented segments that can occur during segmentation, thereby enhancing both the visual clarity and analytical accuracy of the final classified map.

4.3.2. Training Samples (Test 7)

supervised classification is an essential technique in remote sensing, where known samples of specific land cover Classes are used to train the classifier to categorize the entire image based on spectral signatures. In this study, supervised classification is performed, utilizing manual sampling to ensure accurate land cover identification. After performing image segmentation and defining the Class Hierarchy as previously explained, the following steps was applied in a new eCognition window for the supervised classification process is Training samples which were manually selected for each land cover class, including water, bare soil, soil and sand, sparse vegetation, and dense vegetation. Careful attention was given to select representative objects that capture the variability within each class to improve the classifier's accuracy. table 4 shows the number of objects which selected as samples for each land cover class.

Table 4. Number of samples and classified objects in eCognition

Class	Water	Sparse Vegetation	Vegetation	Baresoil	Building	Road	total
The number of Training samples	86	3992	86	8301	6664	539	19668

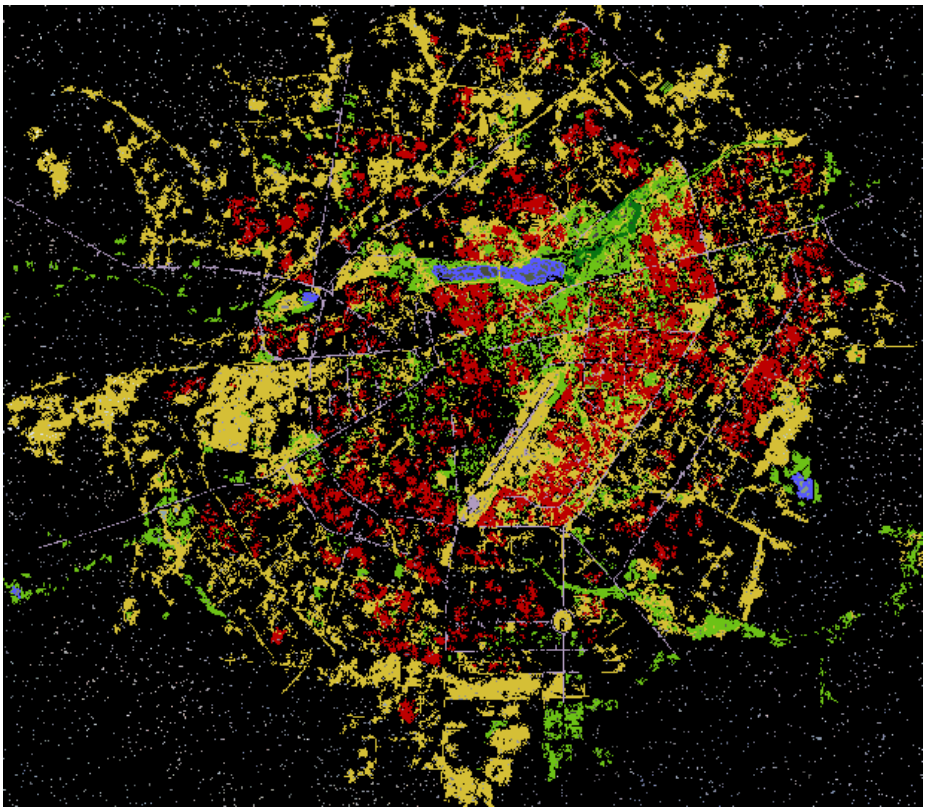

Figure 24. Manually Selected Training Samples for Supervised Classification.(This map highlights the locations of the manually selected training samples used in the supervised classification process, representing various land cover types including water , bare soil , buildings , roads , sparse vegetation , and dense vegetation)

4.3.3. Classifier Algorithm(Test 7):

eCognition offers a variety of supervised classification algorithms. To classify the map using machine learning, the "Random Forest" is implemented from the "Classifier" algorithm within the process tree. the process tree. Here, the classifier utilizes the training samples previously defined, analyzing their spectral properties to model classification rules. The Random Forest setup in eCognition includes parameters such as a depth of 150 trees, a minimum sample count of 10, and a maximum tree count of 50, which are configured to enhance the classifier's ability to accurately segment and classify the landscape into the designated land cover classes. The features used for classification included **mean spectral values** from the Red, Green, Blue (RGB), and Near-Infrared (NIR) bands. These bands provide key information about vegetation, water, and built-up areas. This setup helps to optimize performance by balancing the model's complexity with the need to avoid overfitting.

After training, the classifier was applied to the entire image to categorize each object based on the learned spectral signatures. The "Classifier" algorithm was used to execute this process, resulting in a thematic map that clearly delineates the different land cover types.

4.4. Accuracy Assessment

To perform the accuracy assessment and calculate metrics such as F1 score, recall, and precision, the classified images from each method used in eCognition and QGIS were first exported as raster files and then loaded as layers in QGIS. These classified images represented the results from the different classification methods employed in the study. The manually created samples from eCognition were used as the reference layer (Figure 25). These samples have been carefully chosen to reflect the true characteristics of each class, ensuring that they accurately depict the various land cover categories in the study area.

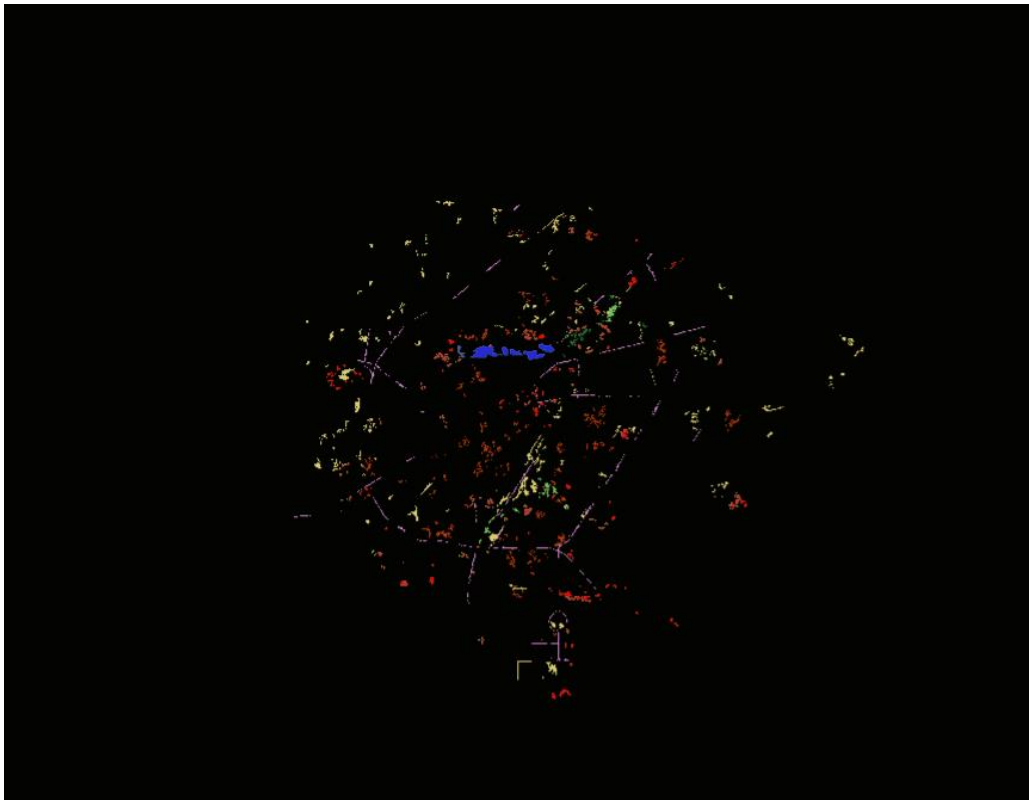

Figure 25. Manually Selected forGround truth data from eCognition including water , bare soil , buildings , roads , sparse vegetation , and dense vegetation)

When it comes to assessing the accuracy of the classification, these samples provide a reliable "ground truth" against which the classified image can be compared. The process involves comparing the classified image to the reference layer to calculate key metrics such as precision, recall, and the F1 score. These metrics provide insights into the accuracy of the classification, identifying where the model has performed well and where it may have made errors.

Unlike the training samples used for supervised classification, this ground truth dataset was smaller in size, containing fewer samples. While these samples were fewer in number, they were carefully selected to represent the true characteristics of each land cover class within the study area.

In QGIS, the Semi-Automatic Classification Plugin (SCP) was utilized to conduct the accuracy assessment. Within the accuracy section of SCP, the reference data and the classified images from each method were entered for comparison. This process generated a confusion matrix for each classification method, which served as the basis for calculating the recall, precision, and F1 score. By comparing these metrics across the different methods, a comprehensive evaluation of the classification performance was achieved, highlighting the strengths and weaknesses of each approach.

Table 5.The number of samples for each class for Ground truth data

Class	Water	Sparse Vegetation	Vegetation	Baresoil	Building	Road	total
The number of samples	14	123	158	232	240	87	854

4.4.1. Challenges in Computing Accuracy in QGIS

During the process of computing accuracy in QGIS, several challenges were encountered that required specific steps to address. These issues were critical to ensuring the validity and reliability of the accuracy assessment and are detailed below:

1. **Unclassified Pixels and NoData Values**: A significant challenge was the presence of unclassified pixels, and marked as NoData. These pixels were influencing the accuracy results negatively, as they were not part of the intended classification but were still being included in the calculations. To mitigate this, the NoData value was specified in the accuracy assessment tool, effectively excluding these pixels from the analysis. This step was crucial in improving the overall accuracy by focusing only on the relevant, classified data.

2. **Non-Matching Values in Attribute Tables**: The final issue involved discrepancies in the attribute tables of the reference and classified layers, where the class values did not match. This mismatch required reclassification of the raster data to align the class values between layers. Using the Raster Calculator in QGIS, a formula, that multiplies or adds values in the raster to adjust them to the desired class values, was applied to adjust the class values, ensuring that the reference classes matched the predicted classes in the attribute table. This alignment was essential for generating a valid confusion matrix and for accurately assessing the classification performance.

   Here's a general formula that adapted for reclassifying values in a raster layer using the QGIS Raster Calculator:

   (("layer_name"= original_value_1) * new_value_1) + (("layer_name" = original_value_2) *new_value_2) + ... (("layer_name"=original_value_n)*new_value_n)+(("layer_name"!=original_value_1)*("layer_name"!=original_value_2)*...*("layer_name"!= original_value_n) *"layer_name")

Table 6. formula guide

Word	Description
Original value	The class number identified in the Classified Layer properties
Layer_name	The Classified layer

After addressing these challenges—extracting the correct thematic raster file, ensuring consistent CRS across all layers, excluding unclassified pixels, and aligning attribute table values—the accuracy assessment could be successfully performed. The resulting confusion matrix provided a reliable comparison between the reference classes and the predicted classes, forming the basis for evaluating the classification's effectiveness.

4.5. Surface Temperature and Vegetation Percentage Calculation

To calculate the vegetation percentage and analyze the surface temperature, the following steps were performed using QGIS and its associated tools:

4.5.1. Vegetation Percentage Calculation

To begin, a vegetation mask was generated from the NDVI (Normalized Difference Vegetation Index) layer, using a threshold value of 0.3. This threshold was selected to classify areas with vegetation, as NDVI values greater than 0.3 typically correspond to vegetation cover. Any pixels with an NDVI value above this threshold were considered vegetation, while those below were categorized as non-vegetation.

Figure 26. vegetation mask (threshold value of 0.3)

Next, an area boundary was created in vector format, which defined the extent of the study area. A 500 m x 500 m grid was then overlaid on the vector boundary to divide the study area into smaller sections for analysis. The grid was created using polygon type features, ensuring that each grid cell represented a spatial unit of 500 meters by 500 meters.

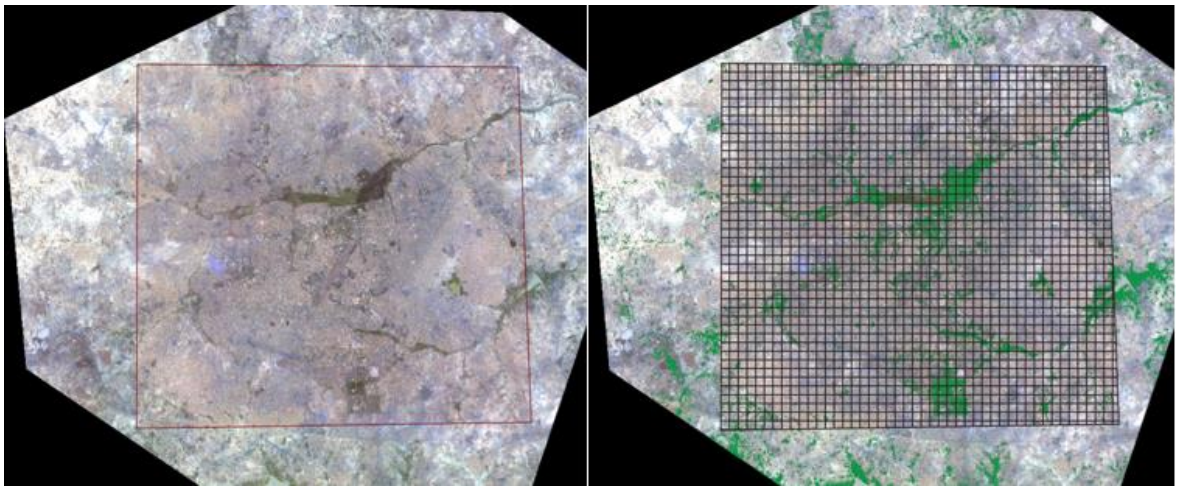

Figure 27. Creating study area and Grid over the area

The vegetation percentage within each grid cell was calculated using the Zonal Statistics tool in QGIS. This tool computes summary statistics for each grid zone. For vegetation percentage, the sum and count values were calculated:

- **Sum** refers to the total number of pixels within each grid zone that meet the vegetation criteria (NDVI > 0.3).
- **Count** refers to the total number of pixels in each grid cell.

To calculate the vegetation percentage for each grid, the sum of vegetation pixels is divided by the total count of pixels in the grid cell and then multiplied by 100. This provides the percentage of the grid cell area covered by vegetation.

$$Vegetation\ percentage = (\frac{sum\ of\ vegetation\ pixels}{count\ of\ total\ pixels}) \times 100 \quad 4.1$$

4.5.2. Surface Temperature Calculation

For surface temperature, the digital numbers (DN) from the Landsat thermal band were first converted to radiance using the formula provided in Chapter 2 (2.5,2.6). This conversion is essential to translate raw satellite data into physical values, such as temperature, that are meaningful for analysis. Afterward, the surface temperature raster layer was clipped to match the extent of the area defined by

the vector boundary, ensuring that only relevant data within the study area was included in the analysis.

The mean surface temperature for each grid cell was then calculated using the Zonal Statistics tool in QGIS. This tool provided the average (mean) surface temperature for each grid, which was used for further analysis.

5. Results

5.1. NDVI-Based Classification Results in QGIS

The NDVI classification maps for Ouagadougou, produced using two different sets of thresholds, reveal significant variations in land cover classification.

The analysis of NDVI classification based on two different sets of thresholds reveals key insights into the performance of each land cover class in terms of precision, recall, and F1-score. These metrics provide a comprehensive evaluation of the accuracy and effectiveness of the classification process for both the initial and refined NDVI threshold sets which is based on histogram Tool.

The NDVI map generated for Ouagadougou offers a comprehensive analysis of the city's land cover, categorizing it into these six distinct classes<

Water bodies, such as rivers, lakes, and ponds, are identified by their negative NDVI values, which result from the low reflectance of water in the near-infrared spectrum. These areas are clearly delineated on the NDVI map, reflecting both natural and artificial water sources within the city.

Bare soil areas, which encompass urban infrastructure like construction sites and exposed soil surfaces, are prominent in the urbanized regions of Ouagadougou. The classification of bare soil was particularly challenging due to the spectral similarities between bare soil, buildings, and roads, especially in areas where local materials like clay are used in construction. These similarities led to some overlap in classification, making it difficult to distinctly separate these classes.

Buildings and **roads** are classified as separate entities on the NDVI map, reflecting the extensive urban development across Ouagadougou. These classes are critical in understanding the city's infrastructure and built environment, though their close spectral characteristics to bare soil required careful threshold adjustments to achieve the most accurate classification.

Sparse vegetation, identified by low to moderate NDVI values, is primarily found in suburban areas where vegetation cover is present but not densely packed. These regions may include grasslands, shrublands, and areas with scattered trees, contributing to the overall greenery of the city without forming dense vegetation clusters.

In contrast, **dense vegetation** areas, which exhibit high NDVI values, are concentrated in parks, agricultural lands, and other green spaces that are vital for the city's ecological balance. These areas are clearly distinguishable on the NDVI map, highlighting the importance of vegetation in maintaining environmental health within the urban landscape.

5.1.1. Test 1: First NDVI Thresholds

the initial NDVI thresholds were applied to classify the study area into six distinct land cover classes: water, bare soil, buildings, roads, sparse vegetation, and dense vegetation. The NDVI values calculated from the satellite imagery were used to define threshold ranges for each class, as outlined in Table 7. These thresholds were manually selected through trial and error using the "Identify Feature" tool in QGIS to distinguish between the different land cover types.

Table 7. the first data of NDVI-Based Land Cover Classification Thresholds

classes	Min NDVI value	Max NDVI value	Color
Water	-1	0.09	1
Building	0.09	0.12	2
Road	0.12	0.14	3
Bare Soil	0.14	0.34	4
Sparse Vegetation	0.34	0.65	5
Dense Vegetation	0.65	1	6

The results of the initial NDVI classification can be seen in Figure 28, which visually represents the distribution of the six land cover classes across the study area based on the applied thresholds.

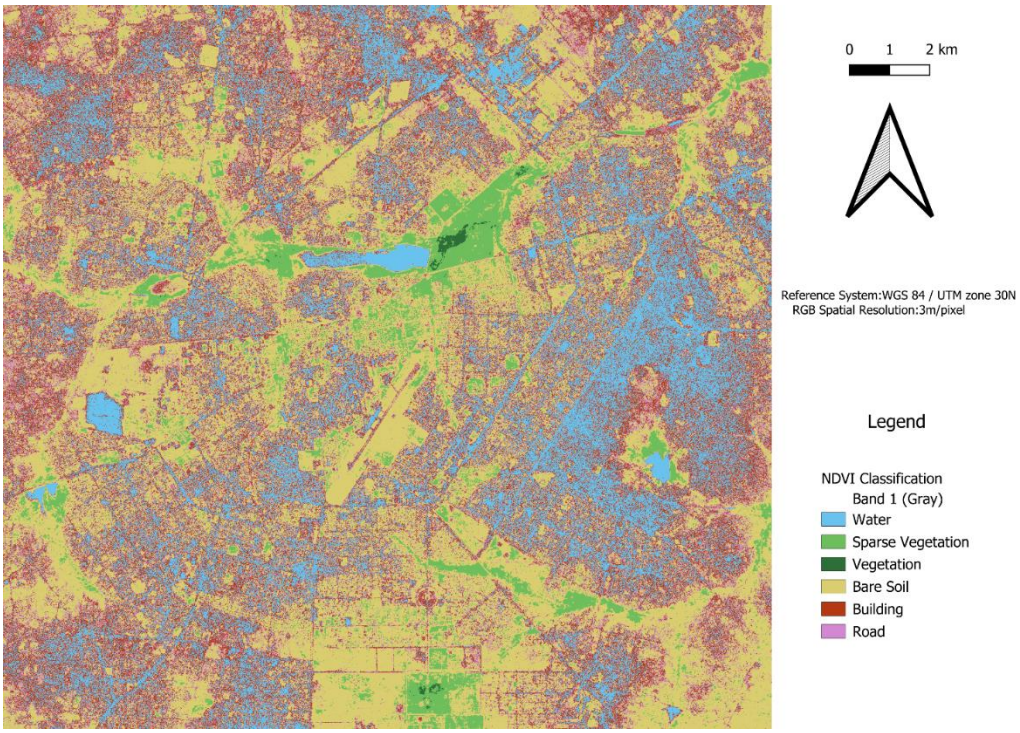

Figure 28. The Classification Of Test 1 (the first NDVI Classification Result in QGIS based on (table 7))

Table 8 shows the confusion matrix for the initial NDVI classification using the first set of thresholds. It provides the counts of pixels classified into each land cover category compared to the reference data.

Table 8. Confusion Matrix for Test 1

Reference \ Classified		water	sparse	vegetation	baresoil	building	road	Total
		1	2	3	4	5	6	
Water	**1**	45530	0	0	378	4278	944	51130
Sparse Vegetation	**2**	6	62732	381	35990	77	249	99435
Dense Vegetation	**3**	2928	61517	16221	9287	1452	1082	92487
Baresoil	**4**	47571	3380	0	78982	58792	31405	220130
Building	**5**	48625	57	0	33122	36420	18316	136540
Road	**6**	40941	398	0	21294	29076	14084	105793
Total		185601	128084	16602	179053	130095	66080	705515

The confusion matrix reveals key patterns in classification accuracy and areas of significant misclassification. Water and sparse vegetation stand out as the most accurately classified categories,

with minimal errors in distinguishing them from other classes. Water, in particular, was well-identified, though there is some overlap with sparse vegetation and roads, reflecting minor confusion. Sparse vegetation performed well but exhibited limited misclassification with baresoil, likely due to similarities in spectral reflectance. In contrast, baresoil and buildings showed notable misclassification, with baresoil often confused with sparse vegetation and buildings, while roads struggled the most, frequently being misclassified as baresoil or buildings. These results emphasize the challenges of differentiating urban features, particularly in environments with overlapping spectral characteristics.

Table 9 shows the TP, FP, FN, and TN values for each land cover class based on the initial NDVI classification using the first set of thresholds.

Table 9. True Positive, False Positive, False Negative, and True Negative Values for Test 1

Classes	TP	FP	FN	TN
Water	45530	140071	5600	120586454
Sparse Vegetation	62732	65352	36703	120612868
Dense Vegetation	16221	381	76266	120684787
Baresoil	78982	100071	141148	120457454
Building	36420	93675	100120	120547440
Road	14084	51996	91709	120619866

The true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values confirm these patterns, providing a detailed breakdown of classification errors. Water showed a strong true positive count, reflecting accurate identification, but had some false positives, indicating minor confusion with non-water classes. Sparse vegetation demonstrated a high TP and relatively low FP and FN, confirming its reliable classification. Dense vegetation had moderate performance, with some areas missed (FN) or misclassified into sparse vegetation or baresoil. Baresoil and buildings faced challenges with overlapping features, resulting in higher FP and FN values. Roads had the lowest TP count and the highest FP, indicating significant difficulty in correctly classifying road surfaces.

Table 10 presents the precision, recall, and F1-score metrics for each land cover class based on the initial NDVI classification.

Table 10. Precision, Recall, and F1-Score for Test 1

Classes	Precision	Recall	f1score
Water	0.25	0.9	0.39
Sparse Vegetation	0.49	0.64	0.56
Dense Vegetation	0.98	0.18	0.3
Baresoil	0.45	0.36	0.4
Building	0.28	0.27	0.28
Road	0.22	0.14	0.17

For the performance metrics, water exhibited excellent recall and strong precision, leading to a high F1-score, showcasing its overall classification reliability. Sparse vegetation had a good balance between precision and recall, resulting in a robust F1-score, indicating consistent identification of this class. Dense vegetation showed moderate precision but a lower recall, reflecting its susceptibility to being misclassified, which impacted its F1-score. Baresoil had balanced but moderate precision and recall, suggesting that while classification was reasonable, there was still noticeable confusion with other classes. Buildings struggled significantly, with both precision and recall being low, leading to a poor F1-score and highlighting the difficulty in separating buildings from urban and baresoil areas. Roads had the weakest performance, with extremely low precision, recall, and F1-score, reflecting substantial misclassification and difficulty in distinguishing linear or narrow features from surrounding surfaces.

5.1.2. Test 2: Second NDVI Thresholds

To explain the second set of NDVI thresholds for land cover classification, we used the raster histogram (Figure 29) to visualize the distribution of NDVI pixel values across the study area. This histogram shows the frequency of pixel values ranging from -0.3 to approximately 0.8, which represent various land cover types.

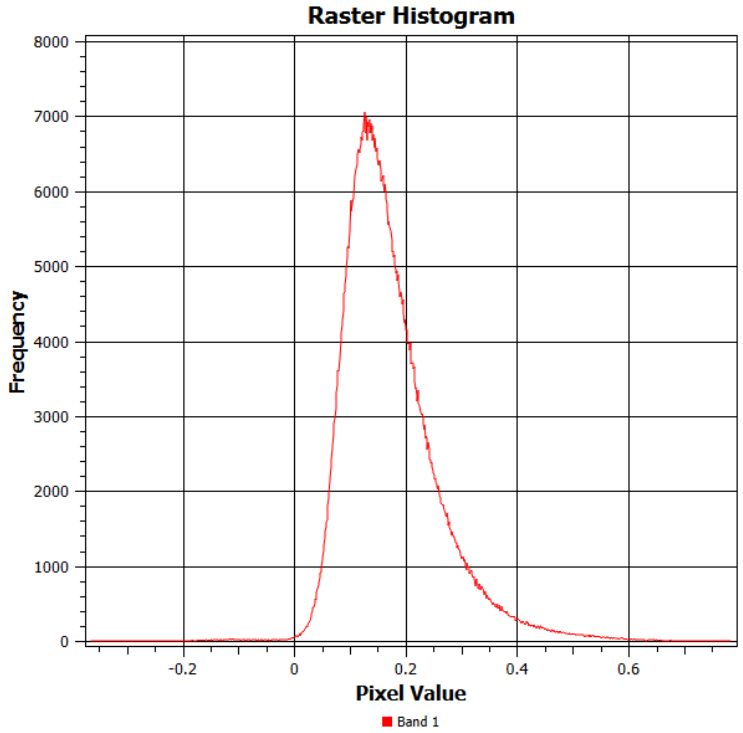

Figure 29. Raster Histogram for Second NDVI Classification (Refined Thresholds)

The selection of NDVI threshold values for the six classes in Table 11 was based on both the shape of the histogram and the interpretation of the land cover types from the NDVI values:

Table 11. the second data of NDVI-Based Land Cover Classification Thresholds

classes	Min NDVI value	Max NDVI value	Color
Water	-1	-0.06	1
Road	-0.06	0.05	2
Building	0.05	0.15	3
Baresoil	0.15	0.26	4
Sparse Vegetation	0.26	0.58	5
Dense Vegetation	0.58	1	6

Water

The NDVI values for water typically fall below 0 due to the low reflectance of water in the near-infrared spectrum. Therefore, we set the water class with a minimum NDVI value of -1 and a maximum value of -0.06. This range captures the negative NDVI values associated with water bodies, which are evident in the lower left section of the histogram.

Buildings

Buildings often show NDVI values slightly above zero, as they reflect some visible light but not near-infrared radiation. Based on the histogram curve and testing, the range for buildings was set between 0.05 and 0.15, representing the flatter region on the histogram that corresponds to built-up areas.

Roads

Roads typically have reflectance properties similar to buildings but often fall within a lower NDVI range. We set the NDVI values for roads between -0.06 and 0.05. This range captures areas that are not vegetated and have relatively low reflectance, likely representing paved or bare road surfaces.

Bare Soil

Bare soil exhibits NDVI values that are generally higher than built-up areas but lower than vegetated surfaces. From the histogram, bare soil can be seen in the section between 0.15 and 0.26. This range represents exposed soil surfaces, which are commonly found in both urban and rural areas, particularly in the dry season.

Sparse Vegetation

Sparse vegetation appears in the mid-range of NDVI values, reflecting the presence of low-density plant cover such as grasslands or shrubs. Based on the histogram, the NDVI range for sparse vegetation was set between 0.26 and 0.58, capturing the rise and fall of the histogram in the middle section where less dense vegetation is present.

Dense Vegetation

Dense vegetation, such as forests and agricultural lands, is characterized by high NDVI values due to the strong reflectance of near-infrared radiation by healthy vegetation. We defined dense vegetation as having NDVI values between 0.58 and 1.0, representing the peak and rightmost section of the histogram, where the highest NDVI values are found.

The histogram was essential for identifying these thresholds, as it provided a visual representation of the NDVI distribution and helped refine the boundaries between classes, particularly for complex land cover types like buildings, roads, and bare soil.

The second set of thresholds (Figure30) was developed to refine the classification. This new set of thresholds aimed to better distinguish between the various land cover types by adjusting the NDVI values used to define each class.

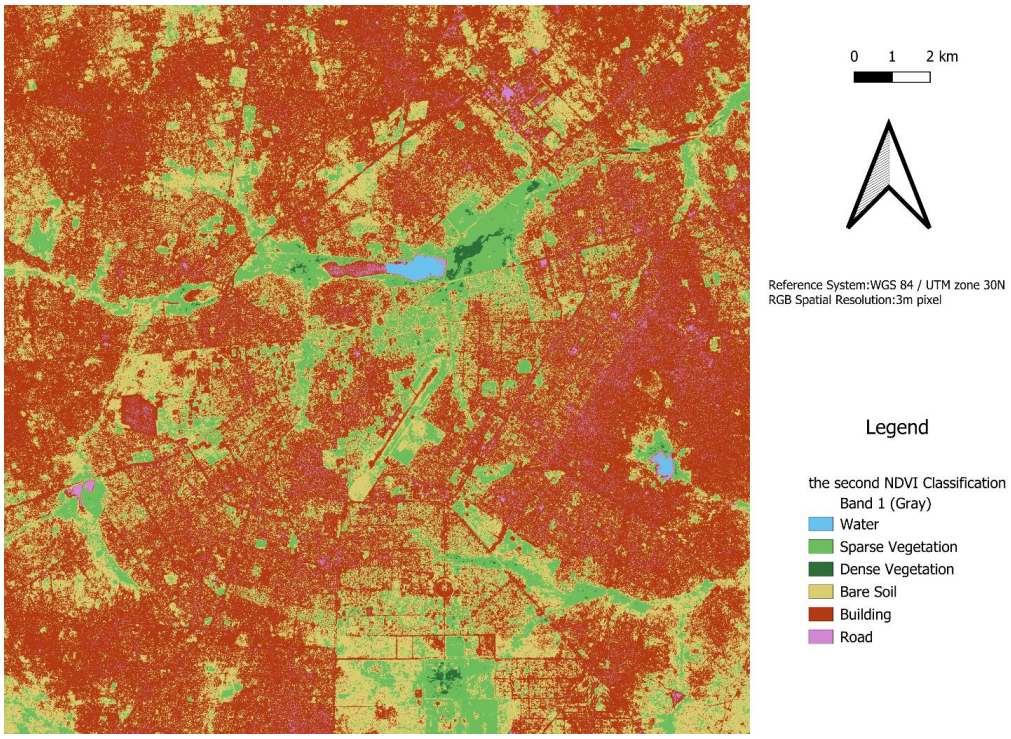

Figure 30. The Classification of Test 2 (the second NDVI Classification Result in QGIS based on (table 11))

Table 12 displays the confusion matrix for the refined NDVI classification using the second set of thresholds. It compares the classification results against the reference data.

Table 12. Confusion Matrix for Test 2

Reference \ Classified		water	sparse	vegetation	baresoil	building	road	Total
		1	2	3	4	5	6	
Water	1	25069	0	0	173	16519	9369	51130
Sparse Vegetation	2	0	83648	3480	11742	564	1	99435
Dense Vegetation	3	4	56711	25793	3922	4995	1062	92487
Baresoil	4	4	21262	0	50559	143371	4934	220130
Building	5	3	1453	0	24850	101118	9116	136540
Road	6	0	1055	0	15602	83230	5906	105793
Total		25080	164129	29273	106848	349797	30388	705515

The confusion matrix highlights the performance of the classification in terms of correctly and incorrectly classified land cover types. Water was well-classified with minimal misclassification, though there were slight confusions with roads and sparse vegetation, likely due to spectral similarities near water bodies. Sparse vegetation was another strong performer, with most pixels correctly classified, but some overlap with baresoil reflects challenges in distinguishing vegetation in

arid or semi-arid conditions. Misclassification was most pronounced in urban categories, such as buildings and roads, which were frequently confused with baresoil due to similar spectral reflectance. Roads, in particular, showed significant errors, being incorrectly labeled as baresoil or buildings, underscoring the difficulty of distinguishing narrow, linear features in a mixed urban landscape.

Table 13 provides the TP, FP, FN, and TN values for each land cover class based on the refined NDVI classification using the second set of thresholds.

Table 13. True Positive, False Positive, False Negative, and True Negative Values for Test 2

Classes	TP	FP	FN	TN
Water	25069	11	26061	120726514
Sparse Vegetation	83648	80481	15787	120597739
Dense Vegetation	25793	3480	66694	120681688
Baresoil	50559	56289	169571	120501236
Building	101118	248679	35422	120392436
Road	5906	24482	99887	120647380

The TP, FP, FN, and TN values shed light on specific classification patterns. Water achieved a strong TP value, confirming accurate identification, but the presence of some FP indicates occasional misclassification with non-water classes. Sparse vegetation also had high TP values and relatively low FN and FP, reinforcing its reliable classification. Dense vegetation, while moderately classified, had noticeable FN, indicating many areas were missed and instead classified as sparse vegetation or baresoil. Misclassification challenges were most evident for buildings and roads, with low TP and high FP/FN values. Baresoil, despite moderate TP values, exhibited frequent confusion with buildings and roads, showing overlap in urban areas. Roads had the lowest TP and highest FP values, reflecting the most significant classification difficulty.

Table 14 provides the precision, recall, and F1-score metrics for each land cover class based on the refined NDVI classification.

Table 14. Precision, Recall, and F1-Score for Test 2

Classes	Precision	Recall	f1score
Water	1	0.5	0.66
Sparse Vegetation	0.51	0.85	0.64
Dense Vegetation	0.89	0.28	0.43
Baresoil	0.48	0.23	0.31
Building	0.29	0.75	0.42
Road	0.2	0.06	0.09

The precision, recall, and F1-scores reveal performance nuances for each class.Water achieved excellent recall and strong precision, resulting in a high F1-score, confirming its consistent identification. Sparse vegetation also had high precision and recall, producing a robust F1-score, highlighting its reliable classification. Dense vegetation, however, had moderate precision and low recall, leading to a lower F1-score due to misclassification as sparse vegetation or baresoil. Baresoil had balanced but moderate precision and recall, with confusion mainly arising from urban features. Buildings performed poorly, with both precision and recall being low, and the resulting F1-score underscored the difficulty in distinguishing buildings from surrounding surfaces. Finally, roads showed the weakest performance, with extremely low precision, recall, and F1-scores, highlighting their frequent misclassification as baresoil or buildings.

5.2. OBIA Classification Results in eCognition

To effectively assess the land cover classification of Ouagadougou, two different sets of NDVI thresholds were defined in QGIS, each representing distinct approaches to classify the land cover types within the study area. These thresholds were used as the foundation for subsequent analysis in eCognition, where Object-Based Image Analysis (OBIA) was employed. To enhance the accuracy of these classifications, GLCM (Gray Level Co-occurrence Matrix) contrast was applied specifically to the green layer across all directions. The integration of GLCM texture analysis aimed to refine the differentiation between similar land cover classes, such as bare soil, roads, and buildings, which often exhibit overlapping spectral characteristics. The results demonstrate that incorporating GLCM contrast into the classification process significantly improved the overall accuracy, providing a more reliable representation of the diverse land cover types in Ouagadougou. The following sections detail the outcomes of these classifications, comparing the results before and after the application of GLCM, and highlight the improvements in accuracy achieved through this advanced analytical approach.

5.2.1. Test 3: First OBIA classification (based on initial NDVI threshold):

The figure 31 illustrates the OBIA classification map of Ouagadougou, created using the initial NDVI threshold (Table 7). The map displays the spatial distribution of various land cover types, including water, sparse vegetation, dense vegetation, bare soil, buildings, and roads, across the study area. The classification provides a visual representation of the land cover categories and their respective extents within the city.

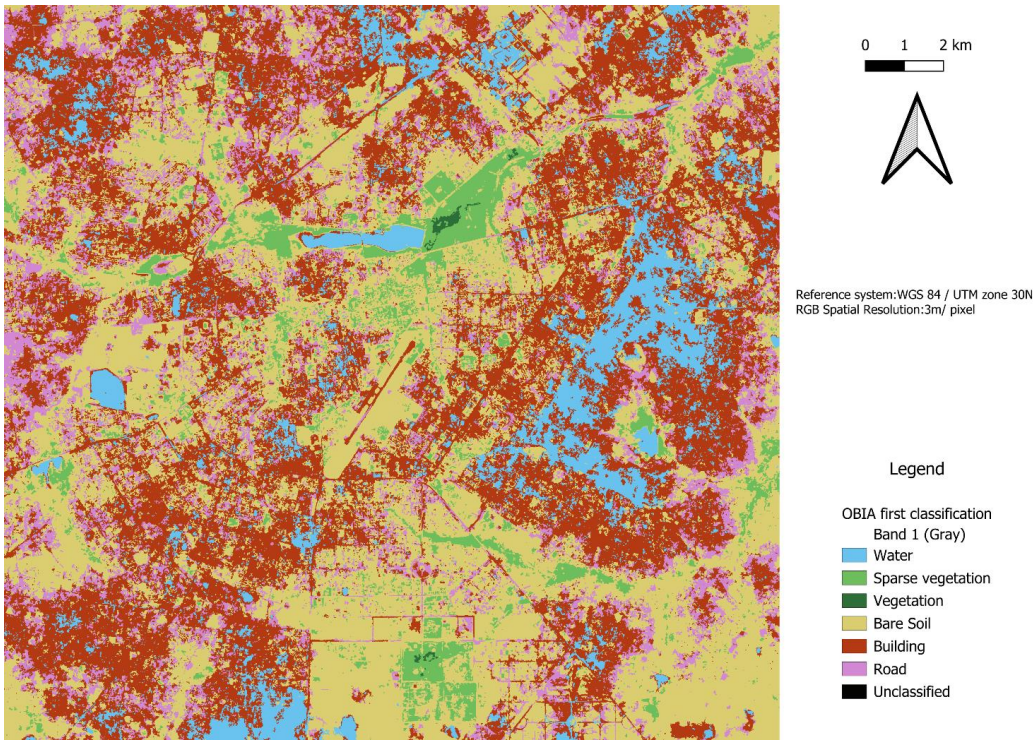

Figure 31. The Classification of Test 3(based on initial NDVI threshold)

Table 15 presents the confusion matrix for the OBIA classification based on the initial NDVI threshold. It shows the number of correctly and incorrectly classified pixels for each land cover type, providing insights into the classification accuracy and the distribution of classification errors across different classes.

Table 15. Confusion Matrix for Test 3 (based on initial NDVI threshold)

Reference \ Classified		water	sparse	vegetation	baresoil	building	road	Total
		1	2	3	4	5	6	
Water	**1**	50630	0	0	0	500	0	51130
Sparse Vegetation	**2**	0	62879	0	36556	0	0	99435
Dense Vegetation	**3**	1707	60746	18185	7850	3999	0	92487
Baresoil	**4**	21887	3807	0	69288	87991	37157	220130
Building	**5**	24511	0	0	29089	71130	11810	136540
Road	**6**	6927	431	0	14238	65385	18812	105793
Total		105662	127863	18185	157021	229005	67779	705515

The confusion matrix reveals significant patterns in classification performance and areas of misclassification. Water and sparse vegetation emerge as the most accurately classified categories, with minimal errors. Water, in particular, shows a strong correct classification rate, though there is

minor confusion with sparse vegetation and roads, likely due to spectral overlap. Sparse vegetation also performs well but exhibits notable misclassification into baresoil, which may be attributed to similarities in spectral reflectance. On the other hand, dense vegetation shows moderate classification accuracy, with significant misclassification into sparse vegetation and baresoil. Urban classes, such as baresoil and buildings, face more considerable challenges, with baresoil frequently misclassified into buildings and roads, highlighting issues in distinguishing these overlapping features. Roads struggle the most, with significant misclassification into buildings and baresoil, reflecting difficulty in identifying linear features.

The table 16 provides a breakdown of the true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) for each land cover class in the OBIA classification based on the first NDVI threshold.

Table 16. True Positive, False Positive, False Negative, and True Negative Values For Test 3 (based on initial NDVI threshold)

Classes	TP	FP	FN	TN
Water	50630	55032	500	120671493
Sparse Vegetation	62879	64984	36556	120613236
Dense Vegetation	18185	0	74302	120685168
Baresoil	69288	87733	150842	120469792
Building	71130	157875	65410	120483240
Road	18812	48967	86981	120622895

The true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values reinforce these findings. Water demonstrates a high TP value, indicating reliable classification, though its moderate FP count suggests occasional confusion with non-water classes. Sparse vegetation maintains strong TP and TN values, confirming its robust classification, while its moderate FP and FN highlight limited misclassification. Dense vegetation, however, shows lower TP and high FN values, revealing that a significant portion of dense vegetation is either missed or misclassified into other classes. Urban features like baresoil and buildings face considerable challenges, with high FP and FN counts reflecting frequent misclassification into overlapping classes. Roads, with the lowest TP and highest FP values, confirm the difficulty in accurately classifying narrow and linear surfaces.

Table 17 summarizes the precision, recall, and F1 score for each land cover class in the OBIA classification using the first NDVI threshold.

Table 17. Precision, Recall, and F1-Score for Test 3 (based on initial NDVI threshold)

Classes	Precision	Recall	f1score
Water	0.48	1	0.65
Sparse Vegetation	0.5	0.64	0.56
Dense Vegetation	1	0.2	0.33
Baresoil	0.45	0.32	0.37
Building	0.32	0.53	0.39
Road	0.28	0.18	0.22

The precision, recall, and F1-score metrics provide a deeper understanding of classification reliability. Water exhibits excellent recall and moderate precision, leading to a high F1-score, reflecting its overall robust classification.Sparse vegetation balances precision and recall effectively, resulting in a strong F1-score and demonstrating consistent classification performance. Dense vegetation, despite high precision, suffers from low recall, indicating susceptibility to being missed, which reduces its F1-score. Urban features like baresoil and buildings show moderate performance, with balanced but lower precision and recall scores, reflecting ongoing challenges in separating these classes. Roads have the weakest performance, with very low precision, recall, and F1-scores, underscoring the significant difficulty in accurately distinguishing road surfaces from surrounding features.

5.2.2. Test 4: First OBIA classification (based on initial NDVI threshold) With GLCM

The integration of the GLCM (Gray Level Co-occurrence Matrix) contrast as an additional threshold in the Object-Based Image Analysis (OBIA) process was aimed at refining the classification results obtained from the initial NDVI thresholding. The GLCM feature was used to enhance the differentiation between classes that were challenging to separate using NDVI alone, particularly in urban environments where bare soil, buildings, and roads often share similar spectral characteristics.

As shown in Table 18, these thresholds reflect the unique textural characteristics of each land cover type. For example, water bodies, which exhibit low variability in texture, were classified using a GLCM contrast range of 0 to 9.4, while buildings, with their complex structures, were categorized within a higher contrast range of 70 and above. By incorporating these GLCM thresholds alongside the NDVI values already established, the OBIA classification became more precise.

Table 18.GLCM thresholds

Classes	GLCM Contrast Min	GLCM Contrast Max
Water	0	9.4
Vegetation	9.4	13
Sparse Vegetation	13	17
Bare Soil	17	40
Road	40	70
Building	70	100+

The figure 32 shows the OBIA classification map of Ouagadougou after applying the initial NDVI threshold and integrating GLCM contrast. The map illustrates the distribution of the six land cover classes, with improved differentiation due to the added texture information.

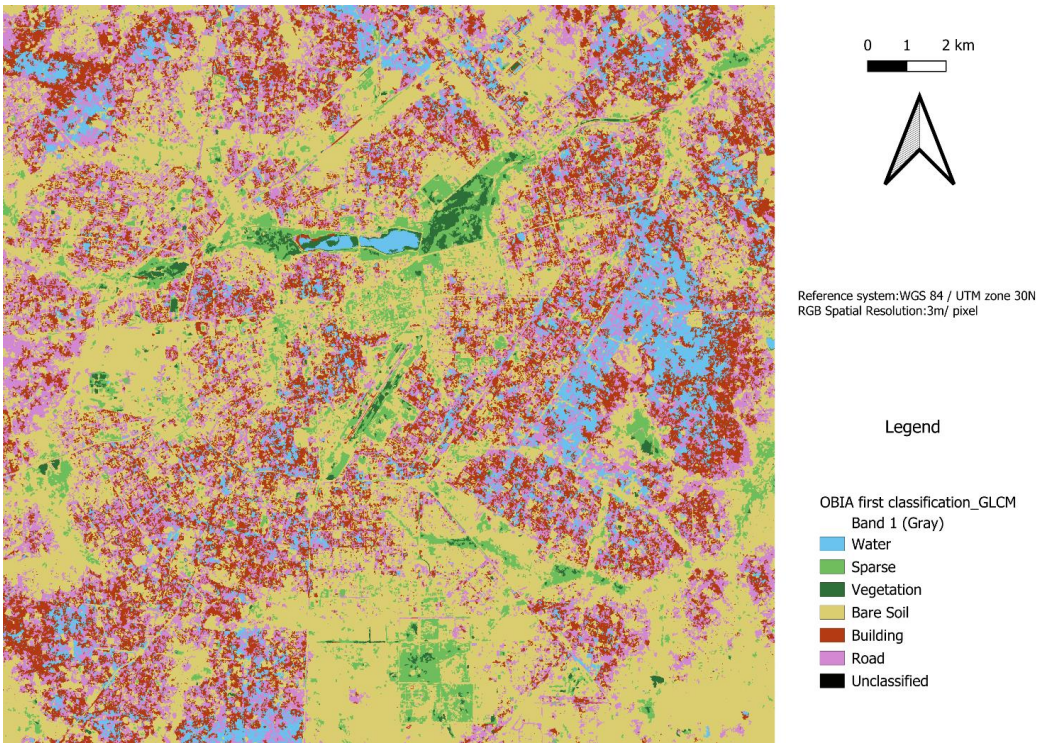

Figure 32. The Classification of Test 4 (based on initial NDVI threshold and considering GLCM)

This table provides the confusion matrix, detailing the number of correctly and incorrectly classified pixels for each class. The inclusion of GLCM improved the classification performance for bare soil and buildings but still showed challenges for roads.

Table 19. Confusion Matrix for Test 4 (based on initial NDVI threshold and considering GLCM)

Reference \ Classified		water	sparse	vegetation	baresoil	building	road	Total
		1	2	3	4	5	6	
Water	1	38093	0	13037	0	0	0	51130
Sparse Vegetation	2	0	57667	17634	24134	0	0	99435
Dense Vegetation	3	0	53337	29670	5822	3658	0	92487
Baresoil	4	5087	47803	22864	110138	9225	25013	220130
Building	5	28285	0	0	29089	63118	16048	136540
Road	6	2133	9890	947	82121	5072	5630	105793
Total		73598	168697	84152	251304	81073	46691	705515

In the confusion matrix table for Test 4, water remains a relatively well-classified category, but its accuracy is lower compared to previous tests, with increased confusion with dense vegetation. Sparse vegetation shows moderate performance but struggles with notable misclassification into baresoil, likely due to the added complexity introduced by texture-based features. Dense vegetation faces

substantial misclassification, particularly into sparse vegetation and baresoil, emphasizing the challenges in distinguishing vegetation types. Urban features, such as baresoil and buildings, show significant overlap, with baresoil often confused with sparse vegetation and roads, while buildings are frequently misclassified into baresoil. Roads remain the most challenging category to classify, with significant errors and frequent misclassification into urban features like buildings and baresoil.

Table 20 shows the True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN) for each land cover class. These values are crucial for calculating Precision, Recall, and F1 scores, which measure the performance of the classification:

Table 20. True Positive, False Positive, False Negative, and True Negative Values for Test 4 (based on initial NDVI threshold and considering GLCM)

Classes	TP	FP	FN	TN
Water	38093	35505	13037	120691020
Sparse Vegetation	57667	111030	41768	120567190
Dense Vegetation	29670	54482	62817	120630686
Baresoil	110138	141166	109992	120416359
Building	63118	17955	73422	120623160
Road	5630	41061	100163	120630801

The true positive, false positive, false negative, and true negative values illustrate these trends. Water demonstrates strong recall but reduced precision due to misclassification into non-water classes. Sparse vegetation maintains reasonable classification performance but sees increased overlap with baresoil. Dense vegetation struggles the most, with many instances misclassified or missed, reflecting difficulties in identifying this class amidst spectral and textural overlaps. Baresoil and buildings face notable challenges, with baresoil often confused with vegetation classes and buildings overlapping with baresoil. Roads perform the weakest, with widespread misclassification and minimal true positive identification.

Table 21 presents the Precision, Recall, and F1 score for each land cover class. The results highlight the benefits and limitations of integrating GLCM into the classification process, with some classes showing significant improvement in accuracy and others still facing challenges.

Table 21. Precision, Recall, and F1-Score for Test 4 (based on initial NDVI threshold and considering GLCM)

Classes	Precision	Recall	f1score
Water	0.52	0.75	0.62
Sparse Vegetation	0.35	0.58	0.44
Dense Vegetation	0.36	0.33	0.34
Baresoil	0.44	0.51	0.47
Building	0.78	0.47	0.59
Road	0.13	0.06	0.08

Performance metrics highlight the overall impacts of the GLCM-based approach. Water retains good recall, reflecting its ability to correctly identify most water instances, though its precision is reduced due to increased overlap with other classes. Sparse vegetation shows a moderate drop in precision and recall, indicating challenges in distinguishing it from baresoil. Dense vegetation has low precision and recall, reflecting its susceptibility to misclassification and reducing its overall reliability. Baresoil shows balanced but moderate precision and recall, revealing confusion with vegetation and urban features. Buildings perform better in precision than recall, indicating frequent correct identification but also significant misclassification. Roads, with very low precision, recall, and F1-score, underscore the difficulty in identifying linear features with this classification method.

5.2.3. Test 5: Second OBIA classification (based on Refined NDVI threshold):

The figure 33 represents the land cover classification of Ouagadougou based on the refined NDVI thresholds (Table 11). The classification categories include water, sparse vegetation, dense vegetation, bare soil, building, and road. The refined NDVI thresholds provide a more accurate distinction between these classes compared to the initial thresholds.

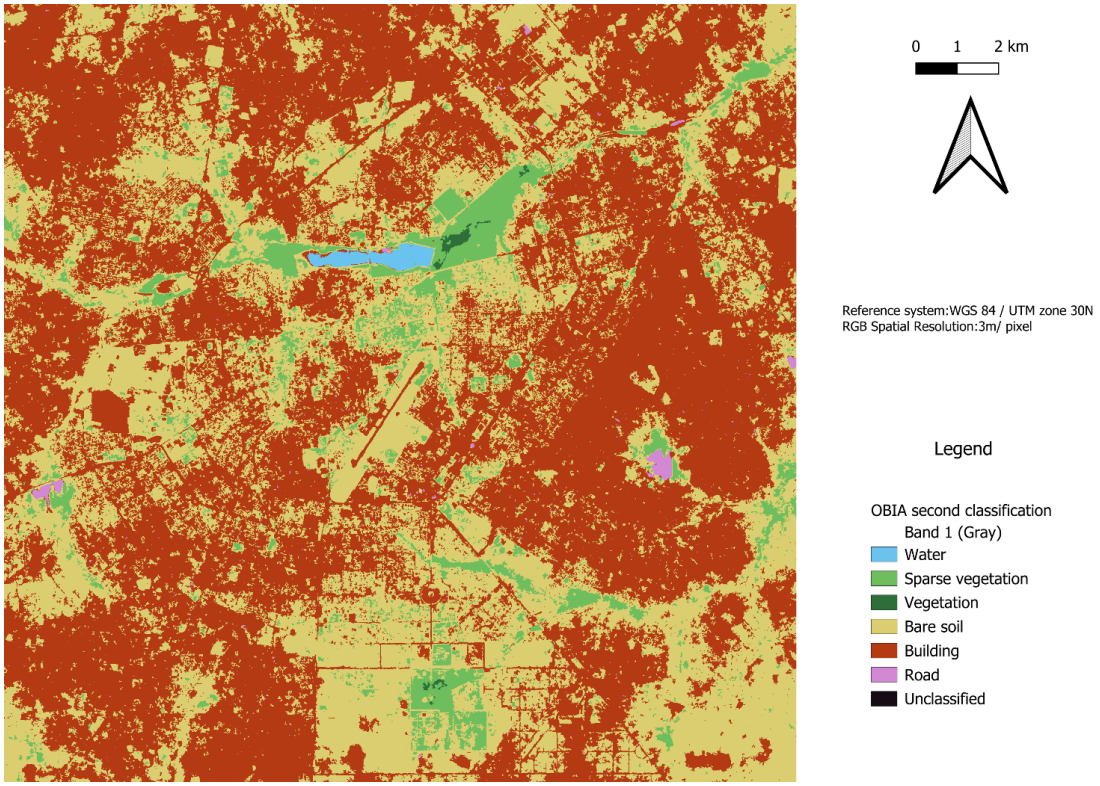

Figure 33. The Classification of Test 5 (based on refined NDVI threshold)

The table 22 shows the confusion matrix for the second OBIA classification based on the refined NDVI threshold. The rows represent the reference data, while the columns represent the classified data. The matrix provides detailed insights into how accurately each land cover type was classified, with numbers indicating the count of pixels for each class comparison.

Table 22. Confusion Matrix for Test 5 (based on refined NDVI threshold)

Reference \ Classified		water	sparse	vegetation	baresoil	building	road	Total
		1	2	3	4	5	6	
Water	**1**	50630	0	0	0	500	0	51130
Sparse Vegetation	**2**	0	62879	0	36556	0	0	99435
Dense Vegetation	**3**	1121	60952	18185	7644	4585	0	92487
Baresoil	**4**	0	6868	0	66227	147035	0	220130
Building	**5**	0	0	0	28352	107430	758	136540
Road	**6**	0	431	0	14238	91124	0	105793
Total		51751	131130	18185	153017	350674	758	705515

In the confusion matrix table for Test 5, water shows excellent classification accuracy, with almost no confusion into other classes, highlighting its distinct spectral characteristics and the effectiveness of the refined NDVI threshold. Sparse vegetation also performs well but shows notable misclassification into baresoil, which continues to be a recurring challenge. Dense vegetation exhibits significant overlap with sparse vegetation and baresoil, showing ongoing difficulty in distinguishing between these vegetative classes. Urban features such as baresoil and buildings demonstrate substantial confusion, with baresoil misclassified into buildings and sparse vegetation, while buildings are often mistaken for baresoil. Roads remain a problematic category, with almost no correct classifications and widespread errors across all other classes.

The table 23 presents the True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) values for each land cover class in the refined NDVI classification.

Table 23.True Positive, False Positive, False Negative, and True Negative Values for Test 5 (based on refined NDVI threshold)

Classes	TP	FP	FN	TN
Water	50630	1121	500	120725404
Sparse Vegetation	62879	68251	36556	120609969
Dense Vegetation	18185	0	74302	120685168
Baresoil	66227	86790	153903	120470735
Building	107430	243244	29110	120397871
Road	0	758	105793	120671104

The true positive, false positive, false negative, and true negative values highlight these results. Water stands out as a reliably classified category with minimal false positives and negatives. Sparse vegetation maintains relatively strong performance but faces challenges with significant false

negatives, indicating areas where the model failed to identify sparse vegetation correctly. Dense vegetation continues to struggle, with a high number of false negatives, reflecting a large portion of missed dense vegetation pixels. Baresoil shows moderate classification performance but suffers from substantial false positives and negatives, indicating confusion with both vegetation and urban features. Buildings, despite having one of the highest true positive counts, still experience considerable false negatives, reflecting difficulties in separating them from baresoil. Roads have the weakest performance, with no meaningful correct classifications and widespread misclassification into other categories.

The Table 24 summarizes the precision, recall, and F1 score for each land cover class in the second OBIA classification. Precision represents the accuracy of the class predictions, recall indicates how well the model identifies all the pixels of a given class, and the F1 score provides a balanced metric of precision and recall.

Table 24. Precision, Recall, and F1-Score for Test 5 (based on refined NDVI threshold)

Classes	Precision	Recall	f1score
Water	0.98	1	0.99
Sparse Vegetation	0.48	0.64	0.55
Dense Vegetation	1	0.2	0.33
Baresoil	0.44	0.31	0.36
Building	0.31	0.79	0.45
Road	0	0	NAN

Performance metrics underline these observations. Water achieves near-perfect precision, recall, and F1-score, showcasing its clear separability and effective identification. Sparse vegetation demonstrates a balanced precision and recall, resulting in a strong F1-score, though misclassification into baresoil remains a concern. Dense vegetation, while achieving perfect precision, has low recall, indicating that although it is rarely confused with other classes, many of its instances are missed, leading to a low F1-score. Baresoil shows moderate but balanced precision and recall, reflecting ongoing confusion with urban and vegetative classes. Buildings, while achieving high recall, suffer from lower precision, indicating frequent misclassification into baresoil. Roads are the weakest category, with no measurable performance, highlighting the continued challenge in identifying linear features with this classification approach.

5.2.4. Test 6: Second OBIA classification (based on Refined NDVI threshold) with GLCM

The figure 34 illustrates the OBIA classification results based on the refined NDVI thresholds, incorporating GLCM texture analysis (Table 14) to enhance accuracy. The map visually represents the distribution of the six land cover classes: water, sparse vegetation, dense vegetation, bare soil, buildings, and roads.

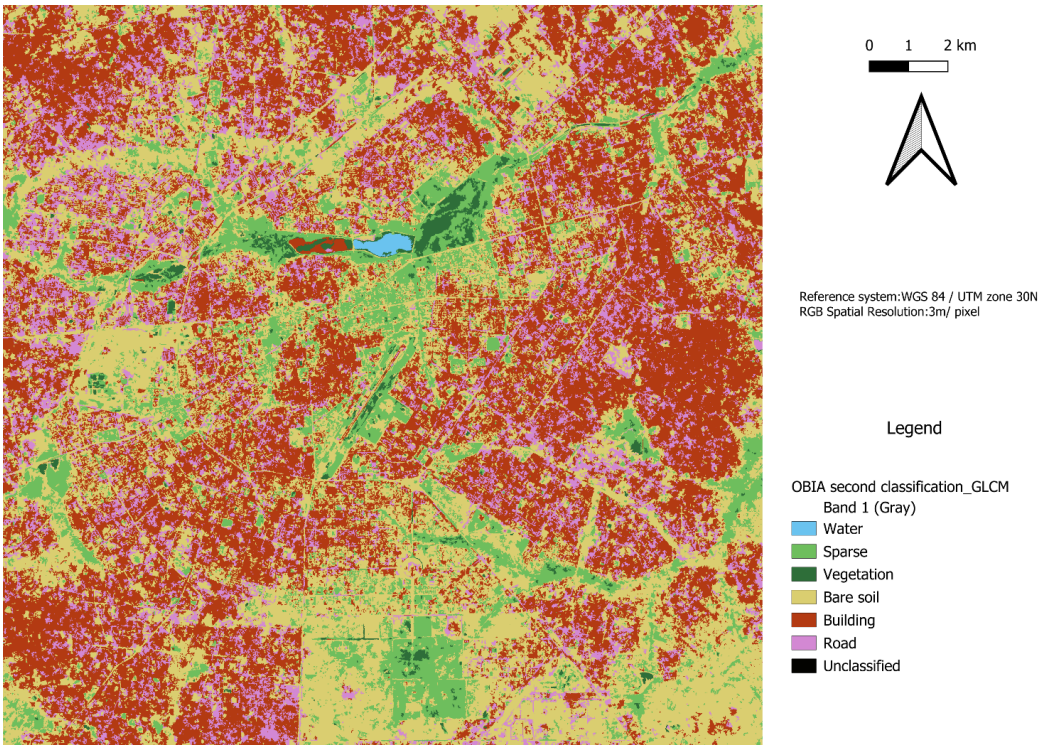

Figure 34. The Classification of Test 6 (based on refined NDVI threshold and considering GLCM)

This confusion matrix details the classification outcomes by comparing the predicted classes against the actual reference classes.

Table 25. Confusion Matrix for Test 6 (based on refined NDVI threshold and considering GLCM)

Reference \ Classified		water	sparse	vegetation	baresoil	building	road	Total
		1	2	3	4	5	6	
Water	1	25163	0	13037	0	12930	0	51130
Sparse Vegetation	2	0	72485	19261	7689	0	0	99435
Dense Vegetation	3	0	53142	32991	2696	3658	0	92487
Baresoil	4	0	50931	22864	104264	24776	17295	220130
Building	5	0	393	0	18179	113846	4122	136540
Road	6	0	9890	947	82121	7841	4994	105793
Total		25163	186841	89100	214949	163051	26411	705515

In the confusion matrix table for Test 6, water shows a decline in classification accuracy compared to previous tests, with increased misclassification into dense vegetation and buildings. Sparse vegetation performs well in terms of identification but struggles with notable misclassification into baresoil and dense vegetation, highlighting the challenges introduced by the integration of GLCM features. Dense vegetation continues to face significant overlap with sparse vegetation and baresoil, suggesting difficulty in distinguishing these categories based on texture. Urban features such as baresoil and buildings display considerable misclassification, with baresoil often confused with buildings and roads, while buildings are frequently misclassified into baresoil. Roads remain the most problematic class, with very few correct classifications and high rates of misclassification into buildings and baresoil.

The table 26 provides a detailed breakdown of the True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN) for each class

Table 26. True Positive, False Positive, False Negative, and True Negative Values for Test 6 (based on refined NDVI threshold and considering GLCM)

Classes	TP	FP	FN	TN
Water	25163	0	25967	120726525
Sparse Vegetation	72485	114356	26950	120563864
Dense Vegetation	32991	56109	59496	120629059
Baresoil	104264	110685	115866	120446840
Building	113846	49205	22694	120591910
Road	4994	21417	100799	120650445

The true positive, false positive, false negative, and true negative values reveal these issues in greater detail. Water demonstrates strong precision, as it is rarely confused with other classes, but its recall is limited due to many missed instances. Sparse vegetation maintains high recall, indicating that most sparse vegetation is correctly identified, but its precision is reduced due to confusion with baresoil. Dense vegetation shows low recall and precision, reflecting widespread misclassification and missed instances. Baresoil achieves moderate recall and precision but struggles with substantial overlap into buildings and roads. Buildings exhibit the best balance among the urban categories, with relatively high precision and recall, though misclassification into baresoil remains a challenge. Roads, once again, exhibit the weakest performance, with minimal correct identification and substantial overlap with other urban classes.

This table summarizes the precision, recall, and F1 scores, providing insight into the effectiveness of the classification for each class

Table 27. Precision, Recall, and F1-Score for Test 6 (based on refined NDVI threshold and considering GLCM)

Classes	Precision	Recall	f1score
Water	1	0.5	0.66
Sparse Vegetation	0.39	0.73	0.51
Dense Vegetation	0.38	0.36	0.37
Baresoil	0.49	0.48	0.48
Building	0.7	0.84	0.77
Road	0.19	0.05	0.08

Performance metrics reinforce these observations. Water retains perfect precision, underscoring its separability from other classes, but its moderate recall lowers the F1-score. Sparse vegetation demonstrates a reasonable balance of precision and recall, resulting in a moderate F1-score, though misclassification into baresoil and vegetation reduces its effectiveness. Dense vegetation has poor precision and recall, leading to a low F1-score, reflecting its susceptibility to misclassification. Baresoil achieves a balanced performance with moderate precision and recall, but its F1-score is impacted by confusion with other urban features. Buildings perform strongly, with the highest F1-score among urban features, showcasing improved classification reliability. Roads remain the weakest class, with the lowest precision, recall, and F1-score, highlighting ongoing difficulties in accurately identifying linear features in this classification scenario.

These results indicate that while the introduction of GLCM texture analysis has improved the classification in some areas, particularly in distinguishing vegetation and bare soil, challenges remain, especially in accurately classifying water, dense vegetation, and roads. The overall accuracy shows that while improvements were made, certain classes still suffer from significant misclassification.

5.2.5. Test 7: Supervised Classification Results

The supervised classification of Ouagadougou was performed using manually selected training samples to represent distinct land cover types within the study area. The training samples included categories such as water, bare soil, buildings, roads, sparse vegetation, and dense vegetation.

Figure 35, illustrates the output of the supervised classification process. The map displays a detailed land cover classification segmented into the predefined categories based on the spectral and spatial characteristics of the image data. The classification map highlights the following:

The manually selected training samples, as shown in Figure 26, played a crucial role in guiding the classification process. Each sample provided a reference point for the algorithm, enabling it to accurately classify similar pixels across the entire image. The samples were distributed across different land cover types to ensure a representative training dataset.

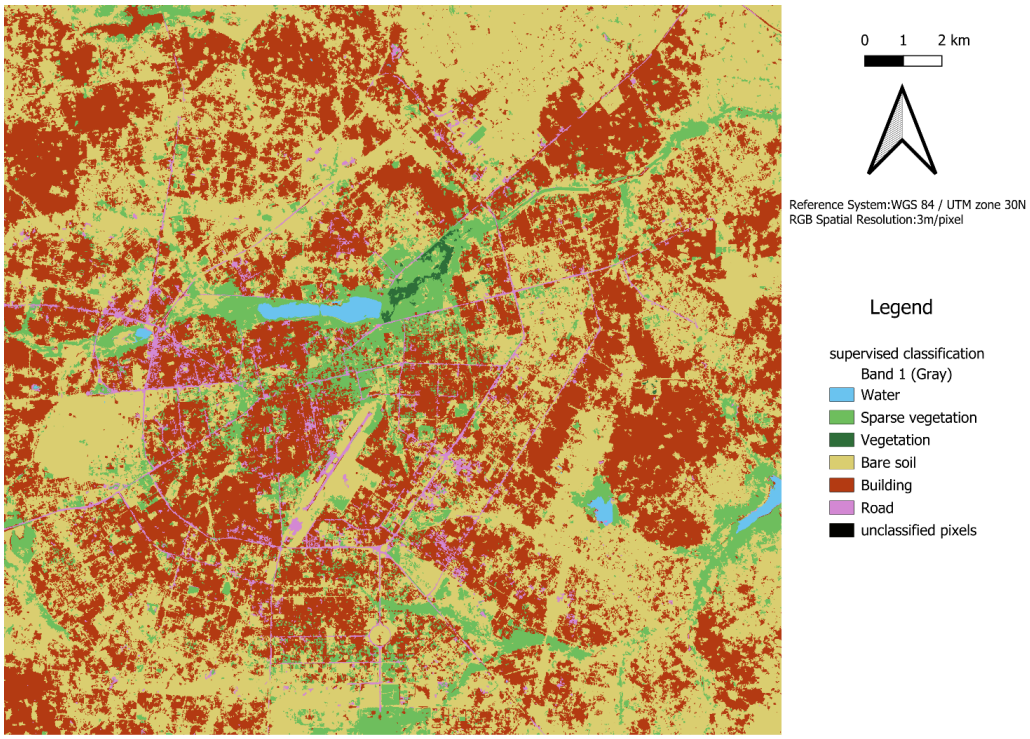

Figure 35. The Classification of Test 7(This map displays the results of the supervised classification using manually selected training samples. The classification differentiates between water, bare soil, buildings, roads, sparse vegetation, and dense vegetation, reflecting the land cover distribution in Ouagadougou.)

The supervised classification results for the NDVI-based land cover mapping in Ouagadougou show significantly improved accuracy and precision compared to previous classifications. The following analysis outlines the precision, recall, and F1-score for each land cover class based on the provided confusion matrix.

The table 28 presents the confusion matrix for the supervised NDVI classification. It shows the number of pixels correctly and incorrectly classified into each land cover category compared to the reference data. It provides a detailed breakdown of how well the classification algorithm performed across different land cover types.

Table 28. Confusion matrix for the supervised classification of land cover(Test7)

Reference \ Classified		water	sparse	vegetation	baresoil	building	road	Total
		1	2	3	4	5	6	
Water	**1**	45630	0	3686	0	0	5000	54316
Sparse Vegetation	**2**	5500	85185	30647	2441	10722	19769	154264
Dense Vegetation	**3**	0	3894	58013	0	0	0	61907
Baresoil	**4**	0	7145	141	215623	7862	1019	231790
Building	**5**	0	0	0	2066	114050	814	116930
Road	**6**	0	3211	0	0	3906	79191	86308
Total		51130	99435	92487	220130	136540	105793	705515

In the confusion matrix table for Test 7, water demonstrates strong classification performance with minimal misclassification into vegetation, highlighting its distinct spectral characteristics. Sparse vegetation also shows significant accuracy but faces notable misclassification into dense vegetation and baresoil, likely due to overlapping spectral properties. Dense vegetation, while showing improvement, still has instances being classified as sparse vegetation, indicating ongoing challenges in differentiation. Baresoil achieves excellent classification, with minimal misclassification into buildings, reflecting its distinct features. Buildings also perform well, though some overlap with baresoil is evident. Roads show considerable improvement, with reduced misclassification into other classes.

Table 29. True positives, true negatives, false positives, and false negatives for each land cover class in the supervised classification of Ouagadougou.

Classes	TP	FP	FN	TN
Water	45630	5500	8686	11742313
Sparse Vegetation	85185	14250	69079	11633615
Vegetation	58013	34474	3894	11705748
Bare soil	215623	4507	16167	11565832
Building	114050	22490	2880	11662709
Road	79191	26602	7117	11689219

The true positive, false positive, false negative, and true negative values provide further clarity. Water has a high true positive count, with very few false positives, indicating its reliable identification. Sparse vegetation, while achieving a high number of true positives, experiences significant false negatives due to overlap with dense vegetation. Dense vegetation has improved true positive rates but still faces challenges with false negatives and misclassification. Baresoil and buildings show robust classification, with high true positives and minimal false positives. Roads, despite showing improvement, still have room for enhancement, with some false negatives and overlap with baresoil.

Overall, while supervised classification shows good performance, particularly for water and sparse vegetation, challenges remain in distinguishing between classes like bare soil, roads, and buildings due to their spectral similarities. The overall accuracy would benefit from refining the thresholds and introducing additional data layers like texture analysis (e.g., GLCM) to improve the classification of complex urban environments.

The table 30 lists the True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) values for each land cover class. These values are crucial for calculating the precision, recall, and F1-scores, helping to understand the classification accuracy and potential areas of confusion between classes.

Table 30. Precision, recall, and F1 score for each land cover class in the supervised classification of Ouagadougou.

Classes	precision	recall	f1score
Water	0.9	0.85	0.87
Sparse Vegetation	0.86	0.56	0.68
Vegetation	0.63	0.94	0.76
Bare soil	0.98	0.94	0.96
Building	0.84	0.98	0.9
Road	0.75	0.92	0.83

Performance metrics indicate water's excellent precision and recall, leading to a high F1-score. Sparse vegetation demonstrates good precision but lower recall, resulting in a moderate F1-score, indicating its susceptibility to confusion with dense vegetation. Dense vegetation achieves strong recall but lower precision, reflecting its challenges with misclassification. Baresoil achieves one of the highest F1-scores, showcasing its reliable classification. Buildings also achieve high precision and recall, resulting in a strong F1-score. Roads demonstrate substantial improvement in both precision and recall, achieving a respectable F1-score, underscoring significant progress in its classification accuracy.

The supervised classification results demonstrate a high level of accuracy across all land cover classes, with precision, recall, and F1-scores significantly better than the previous NDVI classifications using thresholds. This method provides a robust and reliable classification, particularly excelling in distinguishing between challenging classes such as bare soil, buildings, and roads. The overall performance underscores the effectiveness of supervised classification in capturing the diverse land cover types within the urban environment of Ouagadougou.

5.3. Vegetation Percentage and Surface Temperature calculation

In the following section, the results of surface temperature and vegetation percentage calculations are presented and analyzed. These findings provide insights into the spatial patterns and their implications for the study area.

5.3.1. Vegetation Percentage

The vegetation percentage map (Figure 36) represents the spatial distribution of vegetation across the study area. Areas with high vegetation percentages, shown in darker green (79–99%), are mainly located in regions with dense vegetation cover, such as parks, green spaces, or agricultural areas. Conversely, zones with low vegetation percentages, depicted in lighter shades, correspond to built-up areas, roads, or barren lands. This map provides critical insight into the uneven distribution of vegetation within the study area and highlights the urbanized sections where vegetation is sparse.

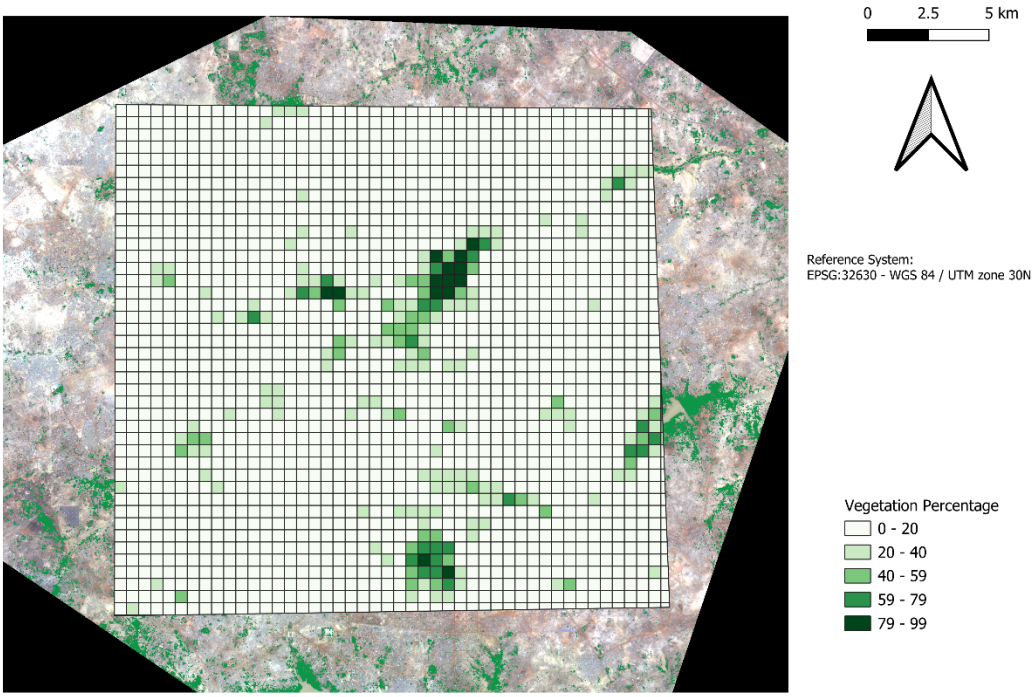

Figure 36. Vegetation Percentage

5.3.2. Surface Temperature

The surface temperature map (Figure 37) shows the spatial distribution of land surface temperature across the study site. Higher temperatures, represented in red (357.4–359.6 K), are predominantly found in areas with little to no vegetation, such as urban centers, industrial zones, or bare soil. Conversely, cooler regions, shown in lighter colors, are associated with vegetated areas. The stark contrast between vegetated and non-vegetated regions reflects the significant role of vegetation in regulating land surface temperatures through processes like evapotranspiration and shading.

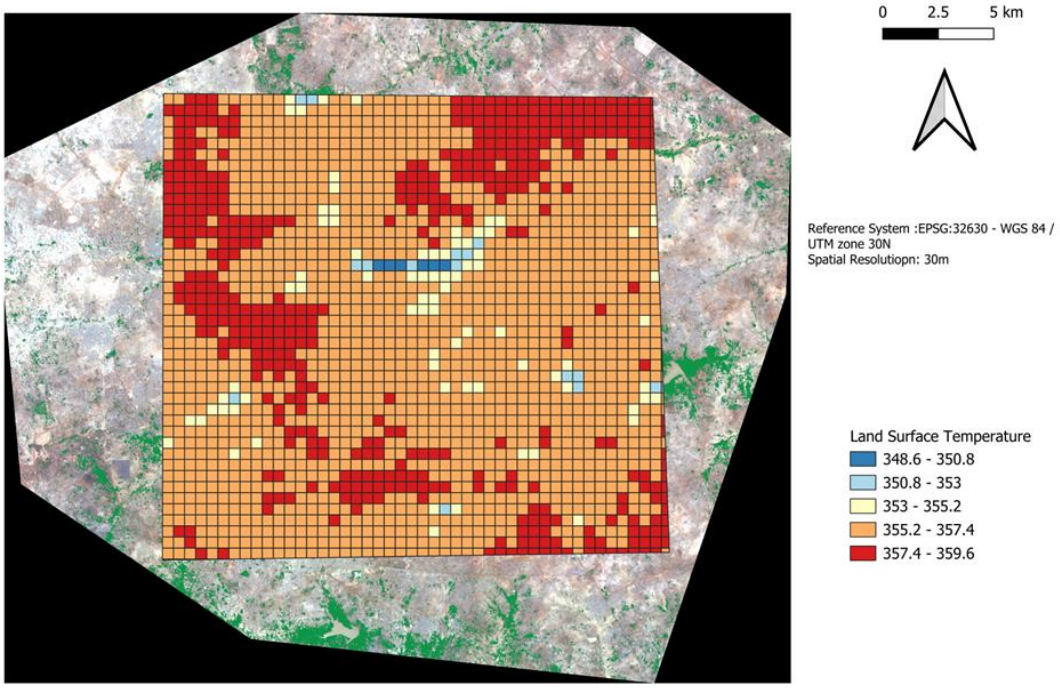

Figure 37. Surface temperature (Kelvin)

5.3.3. Correlation Analysis

Finally, the mean surface temperature and vegetation percentage for each grid cell were extracted and recorded. A correlation analysis was performed using Excel to identify any potential relationships between vegetation percentage and surface temperature across the study area.

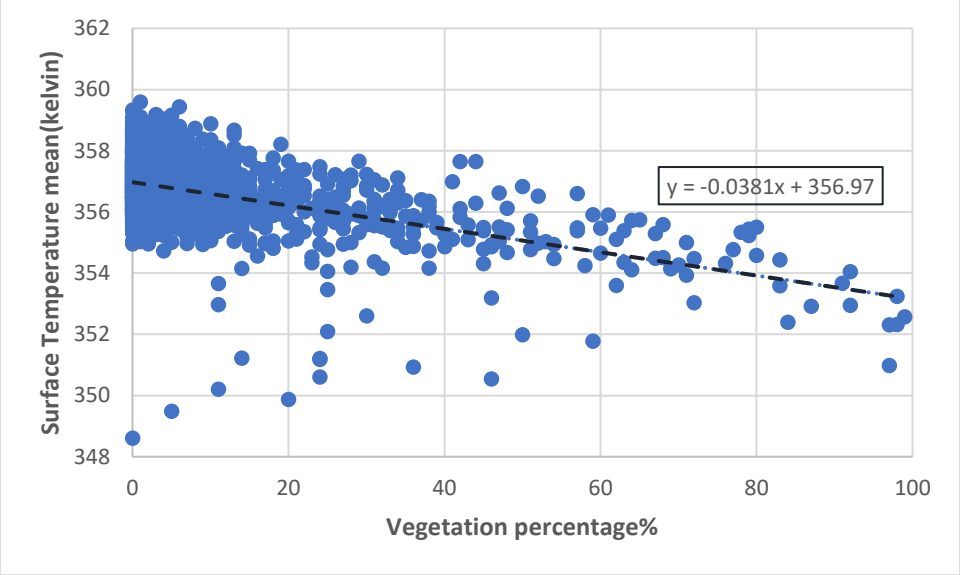

Figure 38. The scatterplot of the relationship between vegetation percentage and mean surface temperature

The scatterplot (Figure 38) presents the relationship between vegetation percentage and mean surface temperature, revealing a clear inverse correlation. The linear regression equation, $y=-0.0381x+356.97$, indicates that for every 1% increase in vegetation coverage, the surface temperature decreases by approximately 0.038 Kelvin. This relationship is consistent with the understanding that vegetation mitigates land surface temperatures through evapotranspiration and shading effects.

6. Discussion

Based on the results from various classification methods and accuracy assessments, the following conclusions can be drawn regarding the effectiveness of the different approaches used in this study for land cover classification in Ouagadougou:

Table 31.comparing Overall Accuracy between different methods of classification

Methods	overall Accuracy
Test 1	0.36
Test 2	0.41
Test 3	0.41
Test 4	0.43
Test 5	0.43
Test 6	0.5
Test 7	0.84

The table 31 comparing overall accuracy across different classification methods shows a progressive improvement in classification performance. Test 1 begins with the lowest overall accuracy, indicating significant challenges in accurately identifying land cover classes. Tests 2 and 3 demonstrate marginal improvements, reflecting adjustments in the classification approach but still showing limitations in handling misclassifications.

By Test 4 and Test 5, a further increase in accuracy is observed, attributed to the refinement of methods, including the incorporation of more sophisticated features like GLCM and improved thresholds. Test 6 marks a substantial improvement, achieving 0.50 overall accuracy, indicating better handling of spectral overlaps and class distinctions. Finally, Test 7 achieves the highest overall accuracy of 0.84, demonstrating the success of the supervised classification method in resolving previous challenges and distinguishing land cover classes with a high degree of reliability. This progression underscores the effectiveness of refining methodologies and incorporating advanced features to enhance classification performance.

Overall, the data supports a trend where more complex classification approaches, particularly those incorporating advanced machine learning models, significantly outshine simpler methods. Future research could benefit from focusing on refining these advanced techniques, potentially applying them to broader urban areas to further validate their effectiveness and adaptability.

Table 32. Comparing F1 Score of all method for each class

methods	Water	Sparse Vegetation	Dense Vegetation	Bare Soil	Building	Road
Test1	0.39	0.56	0.3	0.4	0.28	0.17
Test 2	0.66	0.64	0.43	0.31	0.42	0.09
Test3	0.65	0.56	0.33	0.37	0.39	0.22
Test4	0.62	0.44	0.34	0.47	0.59	0.08
Test5	0.99	0.55	0.33	0.36	0.45	NAN
Test6	0.66	0.51	0.37	0.48	0.77	0.08
Test7	0.87	0.68	0.76	0.96	0.9	0.83

The table shows how the F1-scores for each land cover class improved across the tests. Water starts with low accuracy but improves steadily, reaching its best in Test 5 with nearly perfect classification. Sparse vegetation also gets better over time, with its best performance in Test 7, although it still faces some challenges.

Dense vegetation has struggled in most tests, showing only a small improvement in later tests, but it performs much better in Test 7. Baresoil improves steadily, with a big jump in Test 7, where it is classified very accurately. Buildings show good progress, especially in Tests 6 and 7, with much better accuracy as the tests refine the method. Roads, which performed poorly in earlier tests, show major improvement in Test 7, proving that the advanced methods worked well for identifying them. This table makes it clear that improvements in methods led to better classification for almost all classes.

Moreover, this study explores the inverse correlation between vegetation coverage and surface temperature, a relationship supported by previous research. Vegetation is instrumental in mitigating urban heat island effects, primarily through mechanisms like evapotranspiration and shading. These findings underscore the significant role that dense vegetation plays in regulating surface temperatures, particularly in rapidly urbanizing tropical regions such as Ouagadougou.

7. Conclusion

7.1. Challenges and Limitations

Despite the advancements made through various classification techniques, several challenges were encountered. The spectral similarity between urban features such as Bare Soil, Roads, and Buildings was a persistent issue across all methods, leading to misclassification and reduced accuracy.

classifying these classes proved challenging due to their similar spectral properties. In Ouagadougou, many buildings and roads are constructed from locally sourced materials like clay, which have similar reflectance values to bare soil. This similarity, coupled with the effects of the dry season, which increases the reflectance of bare soil and alters the spectral signatures of sparse vegetation, made it difficult to distinguish these classes accurately using NDVI alone (figure 39). Despite careful threshold adjustments through tools like the "Reclassify by Table" tool in QGIS, accurately separating bare soil from built-up areas was a persistent challenge.

Figure 39. Example of local architecture

Moreover, while the satellite data used in this project is high-resolution, it still faces limitations in capturing the subtle differences between these surfaces. As a result, NDVI, which is typically

effective at differentiating vegetation from non-vegetation, struggles to accurately distinguish bare soil from built-up areas like buildings and roads. This overlap causes the NDVI threshold for bare soil to be less precise, leading to potential misclassifications in the land cover mapping.

Figure 40. Example of Satellite Data

7.2. Implications and Future Directions

This study highlights important insights for improving land cover classification in urban areas. The use of GLCM texture analysis with OBIA proved effective for handling complex urban landscapes, while supervised classification, supported by quality training data, remains the most accurate method. These findings emphasize the value of combining traditional and advanced techniques for better results.

Future research could explore machine learning methods like random forests or support vector machines to automate the classification process without losing accuracy. Refining texture metrics and integrating them with spectral indices, such as NDVI, may further enhance the ability to classify diverse land cover types.

The relationship between vegetation density and surface temperature found in this study shows the importance of vegetation in regulating urban heat. Expanding future studies to include factors like soil moisture, building materials, and seasonal changes could offer a deeper understanding of what influences surface temperature in cities.

In summary, combining NDVI with advanced methods like OBIA and GLCM analysis offers a practical way to improve land cover mapping and understand environmental challenges in urban areas. These findings can guide urban planning and environmental management in Ouagadougou and similar regions.

8. Recommendations

8.1. Summary of Findings

This study on urban land cover classification in Ouagadougou utilized high-resolution PlanetScope satellite imagery to explore various classification techniques. NDVI thresholding, Object-Based Image Analysis (OBIA), and machine learning approaches including Random Forest were employed to distinguish between land cover types such as water, vegetation, bare soil, and urban structures. The integration of these techniques aimed to harness the strengths of each method to improve the accuracy of urban land cover mapping.

The findings revealed that the machine learning approach, specifically the Random Forest algorithm, achieved the highest overall accuracy and reliability. This method outperformed simpler methods like NDVI thresholding and traditional OBIA, highlighting its efficacy in handling complex urban environments. Additionally, the use of Gray Level Co-occurrence Matrix (GLCM) texture analysis alongside NDVI thresholds enhanced the differentiation of land covers that are spectrally similar, thus refining the classification outcomes.

Furthermore, the analysis extended to the correlation between vegetation cover percentages and land surface temperature (LST). Using Landsat 9 thermal data, the study confirmed an inverse relationship, demonstrating the cooling effect of vegetation in urban areas. This highlights the critical role of vegetation in mitigating urban heat islands and supports the integration of thermal data into urban planning and environmental monitoring.

In conclusion, the study underscores the importance of selecting appropriate classification techniques based on specific project needs and the characteristics of the data. Advanced methods like machine learning proved superior in complex settings, offering robust and accurate classifications. These insights suggest that future research could focus on further refining these advanced techniques and exploring their applicability to other challenging environments.

8.2. Answering Research Questions or Hypotheses

Q 1:Which classification method proves most effective at distinguishing between different types of urban land cover?

Test 7, utilizing supervised classification with the Random Forest algorithm, demonstrates the highest F1 scores across nearly all land cover classes, including water, dense vegetation, buildings, and roads. This suggests that this method is most effective at distinguishing between different types of urban land cover

Q 2: To what extent does incorporating NDVI thresholding improve the accuracy of land cover classification in urban environments?

The improved performance in Tests 3, 4, 5, and 6, which incorporated NDVI thresholding into the OBIA framework, indicates that NDVI thresholding enhances the accuracy of land cover classification in urban environments by providing a solid preliminary layer for further detailed analysis.

Q 3: Does NDVI thresholding provide a reliable preliminary layer for further analysis in OBIA workflows?

NDVI thresholding proved to be a reliable preliminary layer for further analysis in OBIA workflows, as evidenced by its repeated use in multiple tests to enhance the segmentation and initial classification phases.

Q 4: How does the integration of Gray Level Co-occurrence Matrix (GLCM) texture analysis influence the precision of land cover classifications obtained from OBIA?

The integration of Gray Level Co-occurrence Matrix (GLCM) texture analysis, particularly in Tests 4 and 6, enhanced the precision of land cover classifications. This improvement is particularly noticeable in the classification of sparse and dense vegetation, where texture features help differentiate between similar spectral signatures.

Q 5: How does the use of a machine learning approach, specifically the random forest algorithm, compare with traditional supervised classification methods in terms of classifying urban land cover?

Test 7, which employed the Random Forest machine learning algorithm, outperformed traditional supervised methods (as seen in the direct comparisons in Tests 1 and 2). This suggests that machine learning approaches can offer superior accuracy and robustness in classifying complex urban land cover.

Q 6: What are the benefits and limitations of employing machine learning algorithms like random forest in urban land cover analysis?

Benefits: Machine learning algorithms, like Random Forest, provide robust classification results, handle large datasets effectively, and are less prone to overfitting.
Limitations: These methods require significant training data to perform optimally and can be computationally intensive, which might be a limitation in resource-constrained settings.

Q 7: What are the synergistic effects, if any, of combining NDVI thresholding, GLCM texture analysis, and machine learning techniques on the overall accuracy of land cover classification?

The project's results suggest synergistic effects from combining NDVI thresholding, GLCM texture analysis, and machine learning techniques. This combination led to enhanced overall accuracy in land cover classification by leveraging the strengths of each method to address different aspects of the classification process.

Q 8: What is the relationship between vegetation cover percentage and surface temperature in the study area, and how does this correlation inform our understanding of urban heat distribution?

The study shows an inverse relationship between vegetation cover and surface temperature, with higher vegetation reducing heat. This highlights the role of green spaces in managing urban heat distribution

8.3. Recommendations for Future Research

the following recommendations for future research are proposed to enhance our understanding of urban land cover classification and improve methodological approaches:

- Investigate the use of other advanced machine learning models like deep learning neural networks which might provide improved classification accuracies over Random Forest for certain types of land cover, especially in highly urbanized areas with complex features.
- Consider using multi-temporal satellite data to monitor changes over time. This could help in understanding the dynamics of urban land cover changes and improve the predictive accuracy of classification models by incorporating temporal variations.
- Apply the developed methodologies to different geographic regions with diverse ecological and urban characteristics to validate the robustness of the classification methods across different environments.
- Utilize higher resolution data or integrate multiple data sources, such as LiDAR and SAR data, to enhance the classification framework. This can help in better distinguishing between land cover classes that are spectrally similar but structurally different.
- Develop methods for automatic tuning of classification parameters, such as the thresholds for NDVI and parameters for GLCM and machine learning algorithms. This could reduce the manual effort needed and potentially increase the objectivity of the classifications.
- Include socio-economic factors which could influence land cover, such as population density or land usage regulations, to provide a more comprehensive analysis of urban land cover changes.
- Examine the impact of urban planning policies on land cover changes by integrating policy data into the classification process. This could provide insights into how urban planning decisions influence land cover dynamics.
- Enhance the validation process by increasing the amount and diversity of ground truth data used. This could involve collaborations with local authorities or the use of crowd-sourced data verification techniques to improve the accuracy assessments.

- Investigate the incorporation of additional environmental parameters, such as soil moisture or urban material composition, to better understand their combined influence on surface temperature variations.
- Explore seasonal or diurnal variations in LST data to gain insights into how temporal patterns affect the relationship between vegetation and urban heat.
- Integrate thermal data from multiple sensors or platforms to improve spatial and temporal resolution for more detailed urban heat mapping.
- Develop models that link vegetation percentage directly with LST to predict urban heat scenarios under different greening strategies, aiding climate-resilient urban planning efforts.

These recommendations aim to build on the current project's successes and address potential areas for improvement, paving the way for more detailed and accurate land cover classification studies in the future.

9. References

[1] C. Zhang and X. Li, “Land Use and Land Cover Mapping in the Era of Big Data,” Land, vol. 11, no. 10, p. 1692, Sep. 2022, doi: 10.3390/land11101692.

[2] E. F. Lambin et al., “The causes of land-use and land-cover change: moving beyond the myths,” Global Environmental Change, vol. 11, no. 4, pp. 261–269, Dec. 2001, doi: 10.1016/S0959-3780(01)00007-3.

[3] EO4GEO. 2023. “Urban Heat Islands – Basic GIS Knowledge Vector and Raster Data.” Accessed November 20, 2024. http://www.eo4geo.eu/training/urban-heat-islands-basic-gis-knowledge-vector-and-raster-data/.

[4] D. J. Marceau and G. J. Hay, “Remote Sensing Contributions to the Scale Issue,” Canadian Journal of Remote Sensing, vol. 25, no. 4, pp. 357–366, Oct. 1999, doi: 10.1080/07038992.1999.10874735.

[5] S. A. Diddams, K. Vahala, and T. Udem, “Optical frequency combs: Coherently uniting the electromagnetic spectrum,” Science, vol. 369, no. 6501, p. eaay3676, Jul. 2020, doi: 10.1126/science.aay3676.

[6] National Aeronautics and Space Administration (NASA). "Electromagnetic Spectrum, NASA Illustration." Accessed November 20, 2024. https://imagine.gsfc.nasa.gov/science/toolbox/emspectrum1.html.

[7] ResearchGate. 2024. "Reflectance of Light of Spectral Signatures of Earth." Accessed November 20, 2024. https://www.researchgate.net/figure/Reflectance-of-light-of-spectral-signatures-of-Earth-fig11_279943423.

[8] NASA. 2024. "Electromagnetic Spectrum." NASA. Accessed November 20, 2024. https://imagine.gsfc.nasa.gov/science/toolbox/emspectrum1.html.

[9] HAIP Solutions. 2024. "What is Hyperspectral Imaging?" Accessed November 20, 2024. https://www.haip-solutions.com/library/what-is-hyperspectral-imaging/.

[10] Atkinson, Peter M., and Paul Aplin. “Spatial Variation in Land Cover and Choice of Spatial Resolution for Remote Sensing.” International Journal of Remote Sensing 25, no. 18 (September 20, 2004): 3687–3702. https://doi.org/10.1080/01431160310001654383

[11] Behask Geo. 2024. "What Do We Mean by Resolution in Remote Sensing? Part 1." Medium. Accessed November 20, 2024. https://medium.com/@behasg_geo/what-do-we-mean-by-resolution-in-remote-sensing-part-1-87e7bb24e514.

[12] Martin Herold, Student Member, IEEE, Margaret E. Gardner, and Dar A. Roberts, Member, IEEE. Spectral Resolution Requirements for Mapping Urban Areas. IEEE TRANSACTIONS ON GEOSCIENCE ANDREMOTE SENSING, VOL. 41, NO. 9, SEPTEMBER 2003, DOI:10.1109/TGRS.2003.815238

[13] Pinterest. 2024. "Spectral Signatures of Earth." Pinterest. Accessed November 20, 2024. https://mx.pinterest.com/pin/539235755364713538/.

[14] Chengbin Deng and Zhe Zhu, "Continuous Subpixel Monitoring of Urban Impervious Surface Using Landsat Time Series," Remote Sensing of Environment 238 (2019): 111011, https://doi.org/10.1016/j.rse.2018.10.011.

[15] J. Aschbacher and M. P. Milagro-Pérez, “The European Earth monitoring (GMES) programme: Status and perspectives,” Remote Sensing of Environment, vol. 120, pp. 3–8, May 2012, DOI: 10.1016/j.rse.2011.08.028.

[16] Equator Studios. 2024. "Remote Sensing." Accessed November 20, 2024. https://equatorstudios.com/remote-sensing/.

[17] Rignot, E., and J. Mouginot (2012), Ice flow in Greenland for the International Polar Year 2008–2009, Geophys. Res. Lett., 39, L11501, doi:10.1029/2012GL051634.

[18] Grind GIS. 2024. "Active and Passive Remote Sensing." Accessed November 20, 2024. https://grindgis.com/remote-sensing/active-and-passive-remote-sensing.

[19] Old Dominion University, "Vegetation Classification: Chapter 5.3," SEES - Significant Opportunities in Atmospheric Research and Science (SOARS), http://www.ccpo.odu.edu/SEES/veget/class/Chap_5/5_3.htm.

[20] D. Phiri, M. Simwanda, S. Salekin, V. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 Data for Land Cover/Use Mapping: A Review," Remote Sensing, vol. 12, no. 14, p. 2291, Jul. 2020, doi: 10.3390/rs12142291.

[21] H. Van Der Werff and F. Van Der Meer, "Sentinel-2A MSI and Landsat 8 OLI Provide Data Continuity for Geological Remote Sensing," Remote Sensing, vol. 8, no. 11, p. 883, Oct. 2016, doi: 10.3390/rs8110883.

[22] E. Javadnia, A. Abkar, and P. Schubert, "Estimation of High-Resolution Surface Shortwave Radiative Fluxes Using SARA AOD over the Southern Great Plains," Remote Sensing, vol. 9, no. 11, p. 1146, Nov. 2017, doi: 10.3390/rs9111146.

[23] ESA. 2024. "Sentinel-1." European Space Agency. Accessed November 20, 2024. https://www.esa.int/About_Us/ESA_Publications/ESA_Publications_Brochures/ESA_BR-313_Sentinel-1.

[24] ESA. 2024. "Radarsat." European Space Agency. Accessed November 20, 2024. https://earth.esa.int/eogateway/missions/radarsat.

[25] NASA. 2024. "ICESat-2." NASA. Accessed November 20, 2024. https://nasa3d.arc.nasa.gov/detail/icesat2.

[26] Zhao, Zheng, and Huan Liu. "Spectral Feature Selection for Supervised and Unsupervised Learning." Proceedings of the 24th International Conference on Machine Learning (ICML '07). Department of Computer Science and Engineering, Arizona State University. New York: ACM, 2007. https://doi.org/10.1145/1273496.1273641.

[27] EOS. 2024. "Make an Analysis." EOS Data Analytics. Accessed November 20, 2024. https://eos.com/make-an-analysis/.

[28] Huang, Sha, Lina Tang, Joseph P. Hupy, Yang Wang, and Guofan Shao. "A Commentary Review on the Use of Normalized Difference Vegetation Index (NDVI) in the Era of Popular Remote Sensing." Journal of Forestry Research (2020). Published online May 31, 2020. https://doi.org/10.1007/s11676-020-01155-1.

[29] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Image Classification,” IEEE Trans. Syst., Man, Cybern., vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.

[30] M. Hall-Beyer, “Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales,” International Journal of Remote Sensing, vol. 38, no. 5, pp. 1312–1338, Mar. 2017, doi: 10.1080/01431161.2016.1278314.

[31] E. Belcore, M. Piras, and A. Pezzoli, “Land Cover Classification from Very High-Resolution UAS Data for Flood Risk Mapping,” Sensors, vol. 22, no. 15, p. 5622, Jul. 2022, doi: 10.3390/s22155622.

[32] P. M. Vitousek, “Beyond Global Warming: Ecology and Global Change,” Ecology, vol. 75, no. 7, pp. 1861–1876, Oct. 1994, doi: 10.2307/1941591.

[33] H. I. Sibaruddin, H. Z. M. Shafri, B. Pradhan, and N. A. Haron, “Comparison of pixel-based and object-based image classification techniques in extracting information from UAV imagery data,” IOP Conf. Ser.: Earth Environ. Sci., vol. 169, p. 012098, Jul. 2018, doi: 10.1088/1755-1315/169/1/012098.

[34] Enderle, Donald I.M., and Robert C. Weih Jr. "Integrating Supervised and Unsupervised Classification Methods to Develop a More Accurate Land Cover Classification." Journal of the Arkansas Academy of Science 59 (2005): Article 10. Available at: https://scholarworks.uark.edu/jaas/vol59/iss1/10.

[35] D. C. Duro, S. E. Franklin, and M. G. Dubé, “A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery,” Remote Sensing of Environment, vol. 118, pp. 259–272, Mar. 2012, doi: 10.1016/j.rse.2011.11.020.

[36] M. Pal and P. M. Mather, “An assessment of the effectiveness of decision tree methods for land cover classification,” Remote Sensing of Environment, vol. 86, no. 4, pp. 554–565, Aug. 2003, doi: 10.1016/S0034-4257(03)00132-9.

[37] Bernard, Simon, Sébastien Adam, and Laurent Heutte. "Dynamic Random Forests." Pattern Recognition Letters 33, no. 12 (2012): 1580–1586. https://doi.org/10.1016/j.patrec.2012.04.003.

[38] Foody, G. M. "Assessing the Accuracy of Land Cover Change with Imperfect Ground Reference Data." Remote Sensing of Environment 114 (2010): 2271–85. https://doi.org/10.1016/j.rse.2010.05.003.

[39] MacLean, Meghan Graham, and Russell G. Congalton. "Map Accuracy Assessment Issues When Using an Object-Oriented Approach." Department of Natural Resources & the Environment, University of New Hampshire, Durham, NH.

[40] R. G. Pontius and M. Millones, “Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment,” International Journal of Remote Sensing, vol. 32, no. 15, pp. 4407–4429, Aug. 2011, doi: 10.1080/01431161.2011.552923.

[41] Guha, Subhanil, and Himanshu Govil. "Land surface temperature and normalized difference vegetation index relationship: a seasonal study on a tropical city." *Springer Nature Switzerland AG*, 2020. DOI: 10.1007/s11676-020-01155-1.