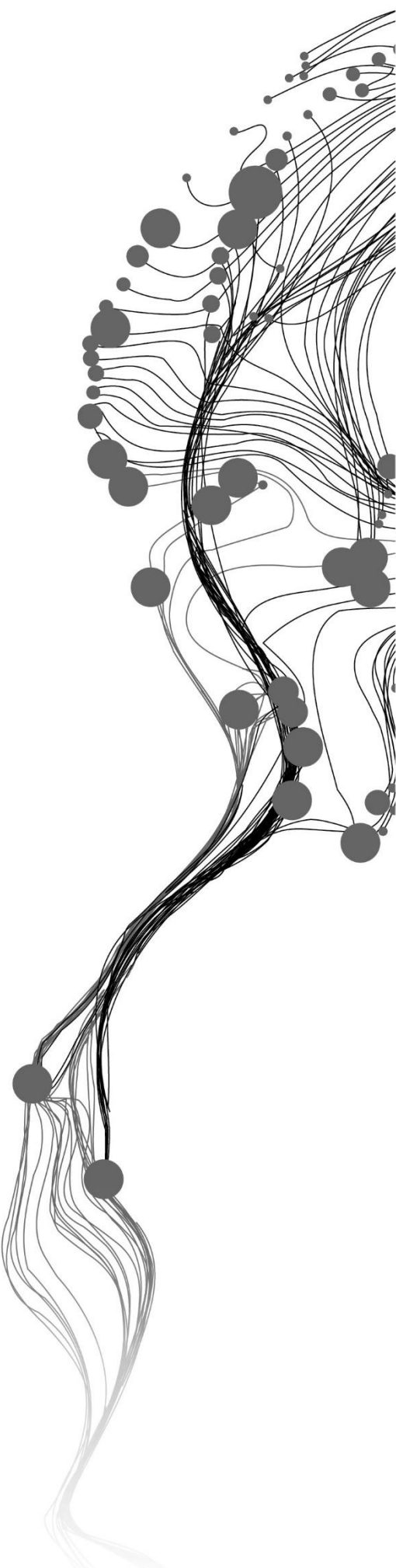# MAPPING DEFORESTATION IN THE GAZA PROVINCE(MOZAMBIQUE) WITH RANDOM FOREST MACHINE LEARNING ALGORITHM ON HISTORICAL LANDSAT SATELLITE IMAGERY.

PRINCE BOATENG
AUGUST 2024

SUPERVISORS:

Dr. RER. NAT. Florian.J. Ellsäßer
Dr. M. Huesca Martinez

MAPPING DEFORESTATION IN THE GAZA PROVINCE(MOZAMBIQUE) WITH RANDOM FOREST MACHINE LEARNING ALGORITHM ON HISTORICAL LANDSAT SATELLITE IMAGERY

PRINCE BOATENG
Enschede, The Netherlands, AUGUST 2024

SUPERVISORS:
Dr. RER. NAT. Florian.J. Ellsäßer
Dr. M. Huesca Martinez

THESIS ASSESSMENT BOARD:
Dr. R. Darvishzadeh Varchehi Roshanak (Chair)
Dr. Alexander Roll (External Examiner, University of Bonn)

ABSTRACT

This study provides analysis of deforestation dynamics and forest regeneration in Gaza Province, Mozambique, spanning the 30-year period from 1993 to 2023. Utilizing a Random Forest machine learning algorithm applied to historical Landsat satellite imagery, the research integrates a diverse range of spectral bands, spectral indices, topographic features, tasselled cap and texture measures to map and quantify landcover changes. The analysis reveals a substantial 37% decrease in forest cover equating to a total loss of 7,288.8 km² with an average annual loss rate of 242.96 km². In contrast, forest regeneration efforts resulted in the recovery of only 2,456 km² of forest averaging 81.9 km² per year. This significant disparity between forest cover loss and regeneration underscores the inadequacy of current reforestation and conservation initiatives in the region. The temporal breakdown of the data highlights critical periods where forest cover loss intensified. Between 1993 and 2004 forest regeneration accounted for 1,085 km² representing 10% of the forest cover while the loss during the same period amounted to 2,323.6 km² of forest cover. In the subsequent periods, forest regeneration rates diminished considerably with 655 km² (6.9%) regenerated from 1998 to 2008, 282 km² (3.3%) from 2004 to 2013, and a mere 152 km² (2%) from 2013 to 2023. Forest cover loss fluctuated but remained substantial peaking at 2,280.2 km² (29%) between 2008 and 2018 marking the most significant reduction in forested land during the study period.

Geographically, the study identifies Mabalane district as experiencing the highest levels of forest regeneration particularly in the earlier periods compared to Chókwè and Bilene districts. Despite these localized efforts the overall trend points to a persistent and severe reduction in forest cover across the province. This trend is further exacerbated by the expansion of cropland and built-up areas driven by agricultural development and urbanization which have significantly altered the landscape.The Random Forest algorithm employed in this study proved highly effective in classifying landcover types, achieving overall classification accuracies between 80% and 86% across the different years analysed. The model was particularly accurate in distinguishing forest and cropland categories with user and producer accuracies consistently exceeding 80%. However, challenges arose in accurately classifying built-up and other land categories reflecting the complexity of these landcover types. The study's use of spectral indices like NDVI as well as topographic and texture measures contributed to the robustness of the model's performance. These findings underscore the critical need for more effective and scalable conservation strategies in Gaza Province.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. BACKGROUND

Tropical forests located between the Tropic of Capricorn (23°26'14"N) and the Tropic of Cancer (23°26'14"S) include both tropical rainforests and tropical dry forests (Roberts et al., 2016). Tropical rainforests experience high rainfall and a warm climate year-round while tropical dry forests have a distinctly seasonal climate with less rainfall (Pennington et al., 2018). These forests are found in the Neotropical (Latin America), Afrotropical (Sub-Saharan Africa) and Indo-Malayan (Southeast Asia) regions (Raven et al., 2020). The Amazon rainforest in Brazil covering 5.5 million square kilometres is the largest tropical forest (Das & Saha, 2021) followed by the Congo Basin in Sub-Saharan Africa which spans 3.7 million square kilometres (Modu et al., 2014). In Southeast Asia, tropical forests are found in countries like Malaysia, Thailand, and Indonesia (Estoque et al., 2019). These tropical forests are epicentres of deforestation.

Sub-Saharan Africa's tropical forests like the other tropical forests face threats of deforestation resulting from rapid population growth, logging, agricultural expansion, collection of firewood and production of charcoal (Arko et al., 2024; Chirwa & Adeyemi, 2020; Maishanu et al., 2019; Martino, 2022). This region is crucial for biodiversity conservation due to endemic species like the *African mahogany* which have traditional and pharmaceutical applications (Abrefa Danquah et al., 2019). Fauna includes species such as the forest elephant, gorillas and the *Okapia johnstoni*, or "forest giraffe" (Williams et al., 2000). Conservation efforts in this region include integrated projects that enhance local livelihoods alongside forest protection such as agroforestry initiatives combined with sustainable farming practices (Franks et al., 2017).

Deforestation is a widespread issue across many Sub-Saharan African countries. Mozambique's tropical forests are experiencing alarming rates of deforestation (Mucova et al., 2018a). The loss of these forests threatens biodiversity and reduces the natural carbon sequestration capacity contributing to climate change (Edwards et al., 2014). Deforestation affects local communities that rely on forests for their livelihoods leading to increased poverty and food insecurity (Chirwa & Adeyemi, 2020a). The high deforestation rates combined with the critical importance of its forests for biodiversity and local communities' well-being make the need for effective mapping and conservation strategies particularly urgent (Tokura, Matimele, Smit, & Hoffman, 2020). Accurately mapping deforestation particularly in Mozambique is crucial for several reasons; Detailed maps help identify critical habitats and biodiversity hotspots allowing for targeted conservation efforts to protect endangered species and preserve ecological balance (Gou & Gou, 2016). In addition, as forests play an important role in sequestering carbon dioxide, understanding deforestation patterns enables better climate change mitigation strategies by highlighting areas that need reforestation and conservation (Yosef, 2014). Furthermore, accurate landcover maps provide essential data for sustainable land-use planning as well as suitability mapping. This helps balance development needs with conservation goals ensuring that agricultural expansion and infrastructure development do not irreversibly damage forested areas. Lastly, reliable data on forest cover changes support policymakers in drafting and implementing effective environmental regulations and conservation policies. It also aids in monitoring compliance with these regulations.

Currently, accurately measuring changes in forest cover is challenging due to the difficulty of accessing remote areas and obtaining historical data. While manual techniques like handheld GPS devices offer precision but time-consuming and are affected by factors such as canopy closure, receiver quality and topography (Valbuena et al., 2010). However, Remote Sensing (RS) offers an efficient alternative with satellite imagery providing spatial and temporal coverage (Kadhim et al., 2016). Satellite sensors like Sentinel, Landsat, and MODIS (Moderate Resolution Imaging Spectroradiometer) are instrumental in mapping forests and extracting crucial information for monitoring forest health, crop yield, soil moisture and urban expansion among others (Radeloff et al., 2024).

The Landsat program is a series of satellites that includes Landsat 5 which was launched by the US Geological Survey in 1984. Landsat 5 operated for 29 years making it the longest-operating Earth Observation satellite in history (Wulder et al., 2022). Its Thematic Mapper (TM) sensor captured detailed multispectral images across seven spectral bands providing valuable data for monitoring landcover, vegetation health, and geological formations (Wulder et al., 2022). With a 16-day revisit cycle Landsat 5 was crucial for tracking environmental changes such as deforestation and urban expansion (Zhou et al., 2019). For example, Bullock et al. (2020) used Landsat 5 to monitor deforestation in Brazil's Rondônia state from 1990 to 2013 achieving high accuracy using spectral mixture analysis and the Normalized Degradation Fraction Index (NDFI). Landsat 7 launched with the Enhanced Thematic Mapper Plus (ETM+) introduced a 15-meter panchromatic band and improved radiometric calibration (Sutradhar, 2018). However, the failure of the scan-line corrector (SLC) in 2003 created data gaps complicating monitoring efforts (Trigg et al., 2006). Despite this, Trigg et al. (2006) documented deforestation linked to industrial activities in Gunung Palung National Park, Indonesian Borneo using Landsat 7. Landsat 8, launched in 2013 brought significant advancements with its Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) offering enhanced imaging capabilities and additional spectral bands such as coastal aerosol and cirrus bands (Roy et al., 2014). These improvements enable better monitoring of atmospheric and surface conditions. Sothe et al. (2017) demonstrated that combining Landsat 8 data with Sentinel-2 significantly improved the accuracy of classifying forest successional stages.

Sentinel-2 with its higher spatial resolution and frequent revisit time offers valuable data for detailed monitoring of landcover and vegetation (Phiri et al., 2020). The combination of Sentinel-2's multispectral data with other sources such as Landsat enhances the accuracy and reliability of environmental monitoring. However, despite its advantages Sentinel-2 shorter time series compared to Landsat limits its utility in long-term change detection studies. PlanetScope known for its high spatial and temporal resolution provides daily imagery at a resolution of 3-5 meters (Houborg & McCabe, 2018). This allows for detailed and frequent monitoring of small-scale changes. However, the shorter historical archive of PlanetScope data can limit its use in long-term environmental monitoring compared to the decades-long data series offered by Landsat.

MODIS known for its high temporal resolution and broad spectral coverage is versatile in tracking spatial and temporal changes (Wei et al., 2019). However, its coarse spatial resolution (250m to 1km) makes it less suitable for detecting fine-scale changes compared to higher-resolution sensors like Landsat (Wei et al., 2019; Zhan et al., 2002). Despite these limitations MODIS's frequent, broad-scale observations are often integrated with higher-resolution data to provide a comprehensive understanding of forest dynamics (Portillo-Quintero & Sánchez-Azofeifa, 2010).

In the field of RS, supervised classification algorithms are essential for accurate and efficient landcover mapping. Various machine learning (ML) algorithms are employed for this purpose each with its unique strengths and weaknesses. For example, Support Vector Machines (SVM) are known for their effectiveness in high-dimensional spaces and are used for classifying complex datasets like satellite imagery (Cervantes et al., 2020). The main advantages of SVM include their ability to manage large feature spaces and their robustness against overfitting especially in high-dimensional data (Bolón-Canedo et al., 2016). However, SVMs can be computationally intensive especially with large datasets and require careful tuning of parameters such as the kernel type and regularization parameter (Horn et al., 2018; Landeros & Lange, 2023).

Another popular ML algorithm used for supervised classification in remote sensing are decision trees (DT). DT are known for their simplicity and interpretability making them a favoured choice for many classification tasks (Costa & Pedreira, 2022). They work by repeatedly splitting the data into subsets based on the value of input features creating a tree-like model of decisions (Priyanka & Kumar, 2020; Thomas et al., 2020a). The main strengths of DT include their ease of understanding and interpretation as they provide a clear visual representation of the decision-making process (Huysmans et al., 2011). Additionally, DT are computationally efficient and can manage large datasets effectively (Thomas et al., 2020b). However, DT have several weaknesses. One significant weakness is their tendency to overfit the training data especially if the tree is allowed to grow without constraints (Tamiminia et al., 2022). This overfitting can lead to poor generalization on unseen data (Thomas et al., 2020b). Additionally, DT can be unstable meaning that slight changes in the data can result in different tree structures(Thomas et al., 2020b).

To address the drawbacks of DT ensemble methods such as Random Forests (RF) are frequently employed. RF is a machine learning algorithm composed of numerous individual DT that work together as an ensemble (Parmar et al., 2019). Each tree in the forest makes a class prediction and the class with the most votes becomes the model's final prediction (Parmar et al., 2019). The benefits of RF include their high accuracy and robustness particularly when working with large and complex datasets (Belgiu & Drăgu, 2016). RF is capable of handling missing data and can provide reliable estimates of feature importance which is useful for understanding the importance of various variables in the classification process (Jordanov et al., 2018). Additionally, they are less sensitive to noise in the data compared to single DT (Belgiu & Drăgu, 2016). However, RF have some drawbacks. They can be computationally intensive requiring more processing power and memory especially with large datasets (Belgiu & Drăgu, 2016; Jordanov et al., 2018). The complexity of RF can also make them more difficult to interpret than single DT as the ensemble approach aggregates results from multiple trees to make the final decision (Aria et al., 2021).

From 2011 to 2020, numerous studies focused on landcover classification in Mozambique aiming to understand environmental changes and inform sustainable land management practices (Assede et al., 2023; Carrilho et al., 2024; Macarringue et al., 2023). A study by Macarringue et al., (2023) highlighted RF effectiveness in processing multi-temporal datasets for accurate landcover classification. The integration of various spectral bands and derived indices notably enhances classification accuracy. Key variables like Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and the Land Surface Water Index (LSWI) are critical in differentiating vegetation types and conditions (Macarringue et al., 2023). The combination of Landsat-8 and Sentinel-2, along with advanced classification algorithms like RF resulted in highly accurate landcover maps with 87% accuracy, essential for agricultural management and land-use planning in the Zambezi River basin (Bofana et al., 2020). Past research has demonstrated

that integrating multi-spectral data and derived indices processed through powerful machine learning algorithms on cloud computing platforms could significantly enhance the accuracy and efficiency of landcover classification in diverse landscapes. For instance, Nogueira Lisboa et al. (2024a) studied the landcover dynamics of the Miombo Landscape in the Beira corridor of Central Mozambique using NDVI, Normalized Difference Water Index (NDWI), and Normalized Burn Ratio (NBR) over two decades. NDVI was particularly effective in yielding accuracies of 88% and 90% for 2000 and 2020 respectively (Nogueira Lisboa et al., 2024). Similarly, Masolele et al. (2022) utilized RF to analyse forest cover loss in Mozambique incorporating terrain attributes and vegetation indices processed on the GEE platform significantly enhancing the study's capability to manage large-scale datasets.

1.2.   Problem Statement

Accurate landcover mapping is important for managing natural resources in Gaza Province, Mozambique, due to its dynamic landscape. These maps are important in tracking environmental changes over time which is vital for sustainable resource management and development. Reliable landcover data informs policy decisions such as forest conservation and management which directly influence the livelihoods of people.

Forest cover loss in Gaza Province demands urgent attention. Forests in this region face continuous threats from illegal logging, urban expansion and land conversion for agriculture, heightening the need for precise monitoring and effective conservation strategies (Aquino et al., 2018). These forests are critical ecological assets that provide significant environmental, economic and climate-related benefits. They serve as vital ecological buffers protecting coastlines from erosion and storm surges while stabilizing sediments and improving water quality through their complex root systems. Also, they support biodiversity by offering habitat and nursery grounds for a diverse range of marine and terrestrial species which is crucial for maintaining ecological balance and supporting local fisheries. Forests are remarkable for their ability to sequester carbon, a feature that positions them as key players in climate change mitigation strategies. Economically, forests are indispensable to local communities in Gaza Province. They provide critical resources such as timber and non-timber products that support local livelihoods. These products include wood for fuel and construction and other materials used in daily community life. The economic reliance on forests underscores the need for sustainable management and conservation practices to preserve these resources for future generations. Conservation and restoration efforts within these forests are vital. Research suggests that while replanted forests are beneficial, natural forests exhibit superior carbon storage and structural attributes, emphasizing the importance of preserving existing forest ecosystems alongside restoration efforts (Chazdon et al., 2016; Stanturf & Mansourian, 2020; Waring et al., 2020).

While mapping forests is essential, there is a gap in detailed landcover classification data at the provincial level, although such mapping has been conducted at the national level. Therefore, estimating changes in forest cover over the past decades is crucial as these changes affect both ecological and human community sustainability. This involves assessing the extent of forest cover loss and the regeneration of forest cover which are important for informed environmental management and policy development.

1.3.   Research objectives and questions

The main objective of this study is to assess the effectiveness of various metrics in differentiating forest cover from other landcover types, evaluate landcover change and estimate the rate of forest cover loss and regenerated forest cover in the study area.

Research Objective 1: To assess the importance of different metrics (spectral bands, vegetation indices, topographic, textural features and tasselled cap features) in distinguishing forest cover from other landcover types.

Research Question 1: Which metrics are relevant for distinguishing forest cover class from other landcover types?

Hypothesis: Among different metrics, spectral bands and vegetation indices will be the most important in distinguishing forest cover from other landcover types due to their relationship with vegetation properties compared to topographic, textural features and tasselled cap features.

Research Objective 2: To estimate the rate of forest cover loss and regenerated forest cover in the study area.

Research Question 2: What are the rates of forest cover loss and forest regeneration in the study area over time?

Hypothesis: Over time, the rate of forest cover loss in the study area is expected to decrease while the rate of regenerated forest cover is expected to increase

2.   METHODS

This chapter explains the study area, data, preparation of remote sensing data, implementation of random forest algorithm, variable importance assessment and change detection.

2.1.   Study area

The Gaza province extends to an area of 75,709 km² and has a population of 1,422,460 (Twumasi et al., 2022). Gaza province is one of the 10 administrative provinces of Mozambique with its capital city located in Xai-Xai. The Gaza province borders Inhambane province to the east, Manica to the north, and Maputo to the south. The province is subdivided into 14 districts and experiences a tropical climate with distinct wet and dry seasons (Salite, 2019). The wet season from November to April is characterized by heavy rainfall and high humidity while the dry season from May to October brings cooler temperatures and minimal precipitation (Mavume et al., 2021). The vegetation in Gaza is savannah featuring a mix of grasslands and scattered trees that thrive under the province's climatic conditions (Massingue, 2019). These forests are primarily made up of miombo woodlands characterized by a variety of tree species adapted to the drier sandy soil prevalent in the area (Massingue, 2019). The topography of the Gaza Province varies from coastal lowlands in the east including floodplains of rivers flowing into the Indian Ocean to higher plateaus in the western parts (Briggs, 2014). The Limpopo River which is one of the major geographical features runs through the province influencing the landscape and soil types across different districts (Briggs, 2014). Due to the vast area of the Gaza Province, the following districts were selected for their unique characteristics (Bilene Macia, Chokwe and Mabalane) as shown below in Figure 1.
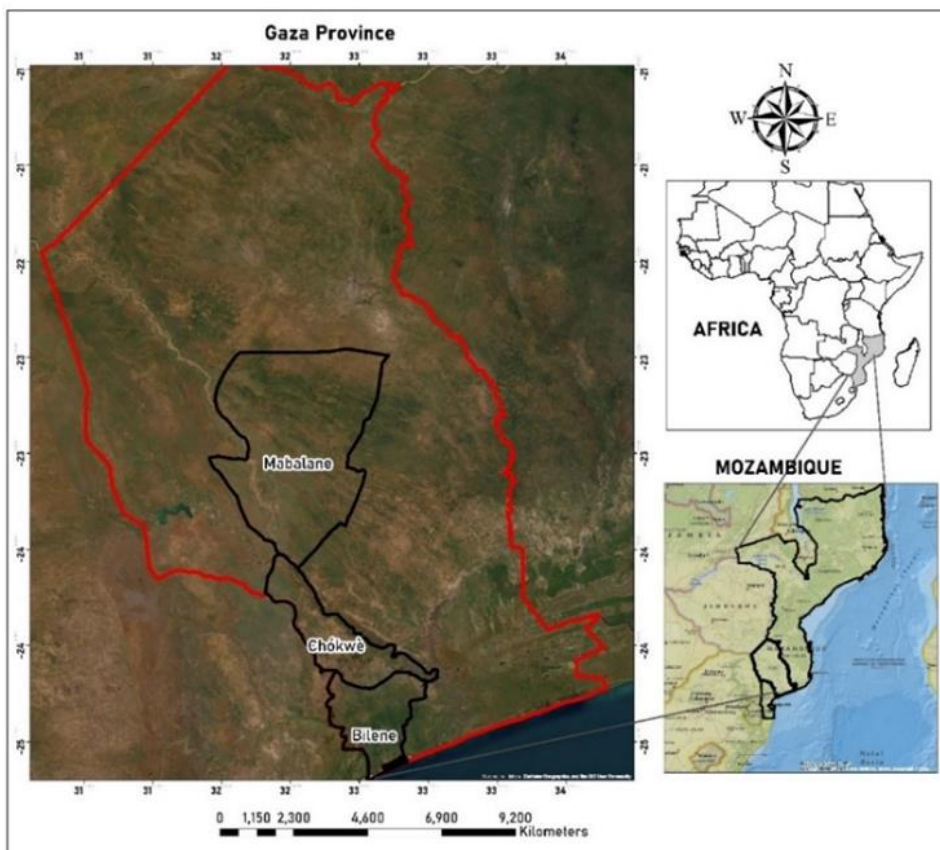
Figure 1: Map of the study area in Gaza Province, Mozambique, highlighting Mabalane, Chókwè, and Bilene districts.

Bilene's urban expansion driven by tourism and agriculture is putting pressure on forest resources leading to habitat and forest loss (Pihale, 2003). Tourism infrastructure development and agricultural growth exacerbate deforestation and land conversion. Chokwe known for its agricultural productivity particularly in rice and maize has seen important forest loss due to agricultural expansion (Ezeokoli et al., 2021). Balancing agricultural development with forest conservation in Chokwe is crucial for sustainable land management. Mabalane with its extensive miombo woodlands is rich in biodiversity and ecological significance but faces threats from illegal logging and charcoal production (Woollen et al., 2016). To map deforestation trends in Bilene, Chokwe, and Mabalane, the years 1993, 1998, 2004, 2008, 2013, 2018, and 2023 were selected. The year 1993 serves as a baseline period before global and regional policy changes regarding environmental conservation and sustainable development widely implemented. By starting in 1993, the study captures the state of forests before major international environmental agreements, such as the Kyoto Protocol which began influencing national policies and practices (McDermott et al., 2010). The year 2008 represents a period shortly after the 2007-2008 global financial crisis which had widespread economic effect that could have influenced deforestation rates and landcover changes (Antonarakis et al., 2022). By 2013, important global initiatives such as the United Nations' REDD+ program were actively promoting forest conservation and sustainable management (Bayrak & Marafa, 2016). The year 2018 falls within the period of the UN's Sustainable Development Goals adopted in 2015 which include specific targets for combating deforestation and promoting sustainable land use (Dugarova & Gülasan, 2017). The year 2023 provides the current state of forests offering a contemporary snapshot for analysis.

2.2.   Data

Table 1 below shows the datasets used in this study including maps, administrative boundaries and satellite imagery

Table 1: Data sources and usage

Data	Description of Data	Data Type	Data Usage	Source
Landsat 5/8	Multispectral satellite imagery	Raster	This was used to analyse and classify	United States Geological Survey. https://earthexplorer.usgs.gov/
Land use/Landcover map	Map of Gaza province depicting landuse and land cover patterns in Mozambique	Vector	This was used as a reference data to identify different landcover types in Gaza, Mozambique	Mozambique National and Remote Sensing Centre (CENACARTA)
Administrative Boundary	This is a shapefile of all the provinces in Mozambique	Vector	This was overlayed with other spatial data and ensure a good alignment of administrative areas	The humanitarian data exchange portal. https://data.humdata.org/dataset/geoboundaries-admin-boundaries-for-mozambique

2.2.1.   Preparation of remote sensing data

For this study, Landsat 5 and Landsat 8 satellite images were selected due to their long lifespan, continuous data collection, free and easy access to data. The years of interest are 1993, 1998, 2004, 2008, 2013, 2018, and 2023 covering 30 years in 5-year interval. These years were chosen to analyse long-term changes in vegetation over three decades in Gaza Province, Mozambique. A region that has experienced environmental changes due to various factors such as climate variability, land use changes and socio-economic developments (Mucova et al., 2018b).

Landsat 5 was used for data from 1993 to 2013. This satellite system offers one of the longest records of earth observation however, imagery data from 2003 had transmission errors (Trigg et al., 2006) hence the image for the year 2004 was used. Landsat 8 was selected to retrieve data from 2013 to 2023 because Landsat 5 does not provide coverage for this period. Both Landsat 5 and Landsat 8 Collection 2 Tier 1 data were used. Tier 1 scenes include Level-1 Precision Terrain (L1TP) processed data with well-characterized radiometry and inter-calibrated sensors which ensures consistent and high-quality data(Hermosilla et al., 2016). The geo- registration of these scenes is within prescribed tolerances (≤12 m root mean square error, RMSE) (Doucette et al., 2013).

The preparation of the satellite images acquired included the application of a Spatial Filter (SF) to restrict the image collection to those intersecting the Region of Interest (ROI). Temporal Filter (TF) was done to limit images to those captured within the specified date range of each selected year (1993, 1998, 2004, 2008, 2013, 2018, and 2023). A Metadata Filter (MF) was also done to include images with less than 5% cloud cover to ensure clearer and more usable data.

To create representative images for each year to be used for classification, the median composite method was employed. This involved:

Dry Season Selection: Images from the dry months (April 1st to November 30th) were used to focus on vegetation studies.

Monthly Median Calculation: The median value for each pixel across all images in a month was calculated. The median is effective in RS as it reduces the impact of outliers like clouds or shadows (Mastin, 1985).

Composite Construction: The median values for each month were combined to create a median composite for the dry season. This approach smooths out variability and highlights the central trend of the data.

By using the median rather than the mean, the final composite images more accurately represent the typical conditions during the dry season minimizing the effect of outliers and providing a clearer picture of vegetation change over the study period. Table 2 below details the band and corresponding wavelength ranges of Landsat 5 TM and Landsat 8 OLI. Both satellites have temporal resolutions of 16 days and spatial resolution of 30meters. The radiometric resolution differs with Landsat 5 (8bits) and Landsat 8 (12bits) as shown in Table 2.

Table 2 : Landsat 5/8 sensor characteristics.

Band	Description	Wavelength (µm) Landsat 5 TM	Wavelength (µm) Landsat 8 OLI	Temporal Resolution	Spatial Resolution (Landsat 5 TM)	Spatial Resolution (Landsat 8 OLI)
1	Blue	0.45-0.52	0.45-0.51	16 days	30 m	30 m
2	Green	0.52-0.60	0.53-0.59	16 days	30 m	30 m
3	Red	0.63-0.69	0.64-0.67	16 days	30 m	30 m
4	NIR	0.76-0.90	0.85-0.88	16 days	30 m	30 m
5	SWIR 1	1.55-1.75	1.57-1.65	16 days	30 m	30 m
6	TIR	10.40-12.50	-	16 days	30 m	-
7	SWIR 2	2.08-2.35	2.11-2.29	16 days	30 m	30 m
8	PAN (Panchromatic)	-	0.50-0.68	16 days	-	15 m
9	Coastal/Aerosol	-	0.43-0.45	16 days	-	30 m
10	Cirrus	-	1.36-1.38	16 days	-	30 m

Source: (USGS, 2024)

2.2.2.   Water masking

Prior to performing landcover classification using Random Forest (RF) waterbodies were excluded from the study area to prevent them from skewing the analysis of landcover changes. This exclusion process involved calculating the Normalized Difference Water Index (NDWI) and applying a threshold of zero (0) to detect water bodies(Gao, 1996a). Also, a similar study by McFeeters, (1996) established a threshold of zero for NDWI where NDWI >0 then the area is classified as water while <= 0, the area is classified as non-water. NDWI has been found effective for distinguishing waterbodies from non-water features (Gao, 1996b; McFeeters, 1996). This preprocessing step ensures that water bodies do not affect the accuracy of the landcover classification by isolating them from the dataset. Regarding healthy vegetation, while NDWI is primarily used for detecting water bodies it also captures water content in vegetation. Healthy vegetation contains water which can influence NDWI values. However, the NDWI threshold approach helps ensure that water bodies are distinguished from vegetated areas with high water content thereby maintaining the accuracy of the landcover classification. These preprocessing steps were crucial for preparing the satellite imagery for analysis.

2.2.3.   Landcover classes for classification

Landcover classifications are important for understanding how land is used and transformed by natural processes and human activities. Key categories such as Forest, Cropland, Built-up, and Other as shown in Table 3 are important in providing critical insights into these dynamics. Forest areas are predominantly characterized by tree cover whereas croplands are primarily devoted to the cultivation of crops including agroforestry systems. Built-up areas refer to land developed for human habitation and infrastructure while the Other category encompasses features such as sand, sand dunes, rocks and non-forest vegetated areas. These classes were selected to align with the specific objectives of the study. Nonetheless, the omission of categories such as Grass and Shrubs despite their importance in landscape analysis was necessitated by their limited visibility in historical high-resolution imagery of Google Earth Pro for 1993, 1998, and 2004, which posed challenge for reliable validation.

Table 3: Landcover classification scheme to produce maps

IPCC landcover classification scheme	
Landcover class	**Definition**
Forest	A land predominantly made up of trees
Cropland	This categorized as cultivated lands, agroforestry lands, that are not defined as forest
Built up	This category encompasses all developed lands that include Built up and transport infrastructure.
Other land	This is categorized into sand, sand dunes and rocks.

Source: (IPCC, 2006)

2.2.4.   Training and validation dataset preparation

To provide sufficient points for training the model, 1000 random points were generated within the study area. For each of these points, a buffer zone of 15 meters by 15 meters was created to capture a precise area around the point. Although this buffer size does not coincide directly with the 30-meter resolution of Landsat pixels, it was chosen to capture finer details and variations within each pixel. These buffered points were then overlayed onto high resolution imagery offered on the Google Earth Pro platform to assess the purity of the land cover classes. This assessment aimed to identify pure classes, defined as areas where the landcover within the entire pixel was homogeneous. Pixels containing more than one class within their boundaries were considered impure and were subsequently excluded from the analysis. In the end the study remained with 400 points of pure classes, 100 points per class. This meticulous process ensured that only pure class pixels were retained for further study enhancing the accuracy and reliability of the landcover classification model. Subsequently, the samples were randomly split 70% training and 30% validation sets(Q. H. Nguyen et al., 2021). This technique ensures an unbiased division of samples maintaining the original data distribution and preventing systematic bias that might occur with ordered data (Q. H. Nguyen et al., 2021). By ensuring both the training and validation sets are representative of the overall dataset, the model was trained and validated effectively. Figure 2 below illustrates the random points selection and buffer creation process.
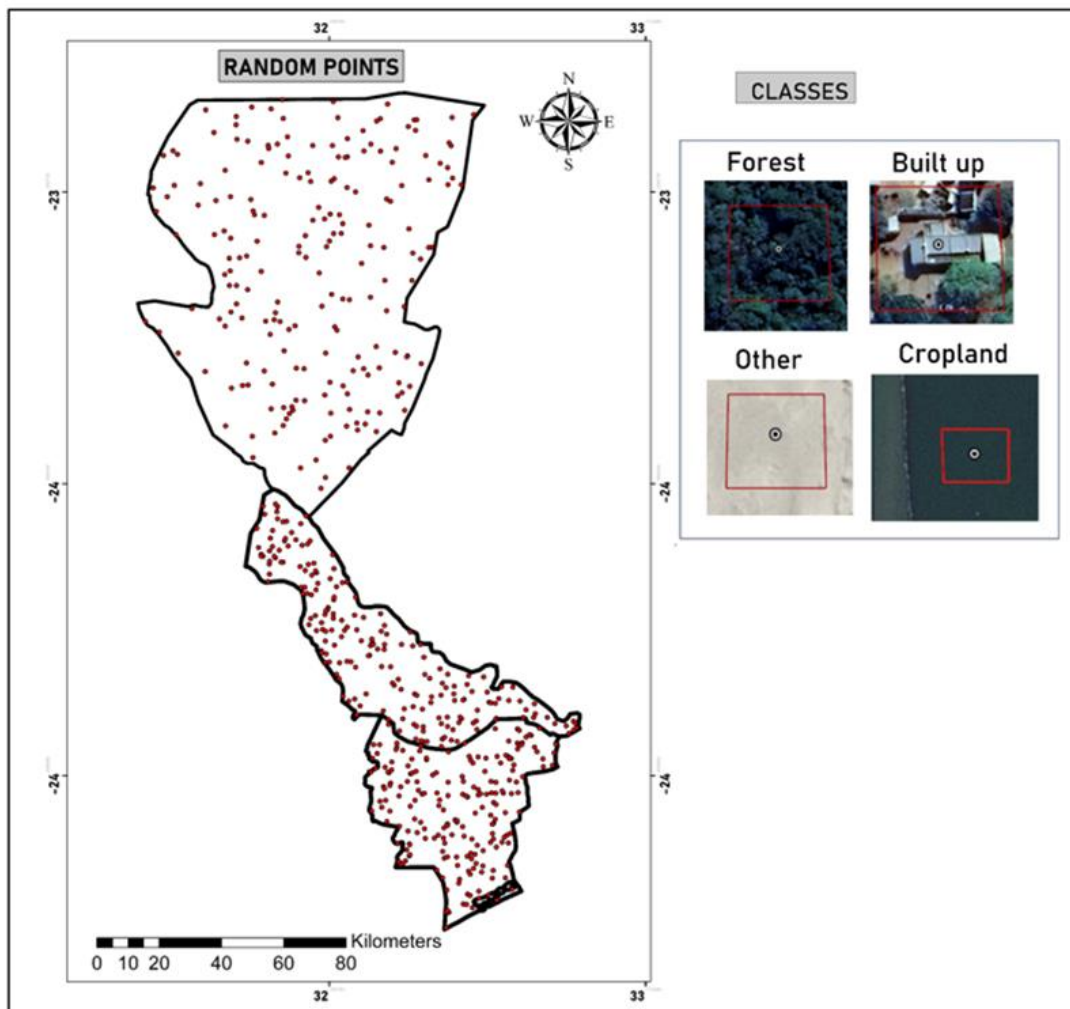

Figure 2 : Random point selection for landcover classification

2.2.5.   Preparation and implementation of random forest

The preparation of input variables for the RF model involved several important steps to ensure accuracy in differentiating forest cover types in the study area.

For this study, various spectral indices were derived including the Normalized Difference Vegetation Index (NDVI). NDVI is a widely used index for assessing vegetation health by comparing near-infrared (NIR) and red reflectance (Pamuji et al., 2023). NDVI values range from -1 to 1, with higher values indicating denser or healthier vegetation (Nugroho et al., 2021; Sihag & Sihag, 2021). Healthy vegetation reflects more NIR light and absorbs more red light due to chlorophyll content, resulting in higher NDVI values (Nugroho et al., 2021). The NDVI is calculated by subtracting red reflectance from NIR reflectance and dividing by their sum (Pamuji et al., 2023). This index has been used in various applications, including drought monitoring, crop production assessment, and environmental change detection (Kogan, 2020). NDVI data derived from satellite imagery provides a long-term, global perspective on vegetation health, enabling researchers to monitor and analyse changes in vegetation cover and condition over time (Kogan, 2020).

On the other hand, the Green Normalized Difference Vegetation Index (GNDVI) is gaining prominence in precision agriculture for assessing crop health and stress. It uses the green band instead of the red making it more sensitive to chlorophyll content and particularly useful in agriculture for assessing plant stress and water content (Croft et al., 2020). Studies have shown that GNDVI is among the strongest correlating indices for estimating crop chlorophyll content, particularly in maize farms (Peter et al., 2021). While traditional indices like NDVI remain useful, green-based indices like GNDVI demonstrate increased sensitivity to vegetation properties (de Lima et al., 2021). In corn monitoring, GNDVI has proven effective in detecting crop variability (Alvino et al., 2020). Additionally, NDVI sensors, which operate on similar principles as GNDVI, have shown strong correlations with commercial chlorophyll meters in estimating leaf chlorophyll content (Chen et al., 2021). These findings suggest that green-based indices, including GNDVI, are valuable tools for assessing plant health, water content, and stress levels in various agricultural applications, offering potential improvements over traditional red-based indices for precision agriculture and crop management.

The Soil Adjusted Vegetation Index (SAVI) was developed to minimize soil influences on canopy spectra by incorporating a soil adjustment factor L into the Normalized Difference Vegetation Index (NDVI) equation (Qi et al., 1994). Meanwhile, Normalized Difference Water Index (NDWI) is employed to monitor water content in vegetation and detect water bodies by comparing green and NIR reflectance (Huang et al., 2009; Sun et al., 2012). The Difference Vegetation Index (DVI) calculates the absolute difference between NIR and red bands but it lacks the normalization found in NDVI (Bacour et al., 2006). Lastly, the Enhanced Vegetation Index (EVI) offers improvements over NDVI by correcting for atmospheric conditions and soil background providing better sensitivity in areas with dense vegetation and high biomass (Fensholt et al., 2006; Zhen et al., 2023). These indices are important for assessing vegetation health, density, and related characteristics by analysing spectral data captured by satellites.

Additionally, topographic features such as slope, aspect, and elevation which describe the physical characteristics of the Earth's surface were extracted from Shuttle Radar Topography Mission (SRTM) data. Slope, a crucial topographic feature derived from SRTM data, represents terrain steepness and is essential for understanding water runoff, soil erosion, and land suitability (Manjare & Singh, 2022) .SRTM-derived slope accuracy depends on factors such as DEM resolution, orientation, and data precision(Saberi et al.,

2023). In low-relief terrain, SRTM DEMs provide more accurate slope estimates compared to interpolated DEMs, particularly the 30m SRTM-derived DEM-S(Kinsey-Henderson & Wilkinson, 2013). Aspect, the compass direction a slope faces influences microclimates, vegetation patterns and solar exposure(Lieffers & Larkin-Lieffers, 1987). Elevation or the height of the land above sea level affect climate, weather patterns and biodiversity. The high-resolution elevation data provided by SRTM is important for environmental monitoring.

Moreover, texture measures such as entropy, homogeneity, and contrast are key variables used to describe the texture of an image and they were calculated using the Grey Level Co-occurrence Matrix (GLCM) method. This statistical approach examines the spatial relationship between pixels by analysing how often pairs of pixels with specific gray level values occur in a defined spatial relationship within the image (Aouat et al., 2021). Entropy reflects the complexity or randomness of the texture with higher values indicating more varied textures(C. Xie et al., 2023). In image analysis, first-order intensity entropy quantifies the dynamic range and randomness of pixel intensities within a region (Carlson, 1989). In time series analysis, entropy measures the rate of information gain and degree of regularity, with higher scores indicating more random or chaotic series (Gan & Learmonth, 2015). Homogeneity measures the uniformity of pixel values where higher values suggest a smoother more consistent texture (C. Xie et al., 2023). Contrast represents the difference in intensity between neighbouring pixels with higher values indicating a more distinct and varied pattern(D. Das & Naskar, 2024). These texture measures provide valuable insights for differentiating and classifying various surface types or materials in the image.

On the other hand, the Tasselled Cap Transformation was used to derive brightness, greenness, and wetness components. It enhances interpretability and provides meaningful insights into the physical properties of the landscape (C. Chen et al., 2019). The stability of Tasselled Cap Transformation components varies across seasons and geographical locations, necessitating the use of appropriate parameters for specific contexts (Ivits et al., 2008). Tasselled Cap Transformation has proven useful in land cover discrimination ((Ivits et al., 2008) and water body extraction, where considering greenness and wetness can help minimize the influence of shadows and dense vegetation (Chen et al., 2022). Greenness highlights vegetation presence and health by emphasizing the difference between visible light which plants absorb and near-infrared light which they reflect with higher values indicating healthier vegetation(Baig et al., 2014a; Eniolorunda & Jibrillah, 2020). Wetness captures moisture content in soil and vegetation by assessing the contrast between shortwave infrared and near-infrared bands making it useful for identifying areas with higher water conten t(Eniolorunda & Jibrillah, 2020). Brightness represents the overall reflectance of the surface (Baig et al., 2014a; Eniolorunda & Jibrillah, 2020).

Remote sensing utilizes various electromagnetic spectral bands to analyse Earth's surface features. The blue band (0.45-0.52 µm) is effective for mapping water bodies and detecting atmospheric aerosols (Laake, 2022). The green band (0.52-0.60 µm) is sensitive to vegetation health, while the red band (0.63-0.69 µm) distinguishes vegetation from soil (Laake, 2022). Near-infrared (NIR) bands are crucial for vegetation analysis, with NIR:visible ratios effectively discriminating between soil and vegetation(Dormann et al., 2013). Combining blue, green, and red bands creates true-colour images resembling human perception (Laake, 2021). However, adding blue band information doesn't significantly improve discrimination between green and dry vegetation or soil (Pickup et al., 2000).

All these indices, spectral bands, topographic, texture and tasselled cap features were combined into a single image and normalized to ensure they were on comparable scales improving the performance of the RF classifier. This process prevents any single variable from disproportionately influencing the model due to its scale leading to more balanced and accurate predictions (Pelletier et al., 2016). Additionally, normalization improves the model's convergence speed during training making the process more efficient (Pelletier et al., 2016). This thorough preparation provided a strong and diverse set of input variables essential for accurate landcover classification and effective natural resource management in the region. A summary of the input variables is presented in Table 4 below.

Table 4: Input variables computed for this study

Variables	Formula	Purpose
NDVI	(NIR-Red)/(NIR+Red)	Identifies presence of vegetation
NDWI	(Green-NIR)/(Green+NIR)	Identifies presence of water in general.
DVI	Near Infrared-Red	Indicates the presence of vegetation
GNDVI	(Green-NIR)/NIR+Green	Enhances sensitivity to chlorophyll content in plants
SAVI	(NIR-Red)*(1+0.5)/(NIR+Red+0.5)	Minimizes soil brightness influence on the NDVI, especially where vegetation cover is sparse.
EVI	2.5*(NIR-Red)/(NIR+6*(Red-7.5*Blue+1)	Improve sensitivity in high biomass regions and reduces atmospheric influences
Slope	Derived from SRTM-DEM (Shuttle Radar Topography Mission) data	Indicates the steepness or inclination of terrain
Aspect	Derived from SRTM-DEM (Shuttle Radar Topography Mission) data	Indicates the compass direction that a terrain surface faces, affecting microclimate conditions
Elevation	Derived from SRTM-DEM (Shuttle Radar Topography Mission) data.	Essential for relief mapping.
Homogeneity	Texture analysis on image bands	Measures the closeness of the distribution of elements in a texture window
Entropy	Texture analysis on image bands	Measures the randomness in the textural image, indicating complexity or disorder.
Contrast	Texture analysis on image bands	Measures the local variations in an image, helpful in distinguishing features.
	(Blue * brightness [0]) + (Green *	

Brightness	brightness [1]) + (Red * brightness [2]) + (NIR * brightness [3]) + (SWIR1 * brightness [4]) + (SWIR2 * brightness [5])	Often used in algorithms to determine land cover features by their reflectivity.
Greenness	(Blue * greenness [0]) + (Green * greenness[1]) + (Red * greenness [2]) + (NIR * greenness [3]) + (SWIR1 * greenness [4]) + (SWIR2 * greenness [5])	Identifies healthy vegetation
Wetness	(Blue * wetness[0]) + (Green * wetness[1]) + (Red * wetness[2]) + (NIR * wetness[3]) + (SWIR1 * wetness [4]) + (SWIR2 * wetness [5])	Identifies areas with higher water content.

Source: (Baig et al., 2014b; Bofana et al., 2020; Cheng et al., 2023; Taye et al., 2013; C. Xie et al., 2023)

2.2.6. Variable selection

To determine the most important variables for distinguishing between different landcover classes (Forest, Cropland, Built up, Other), multiple statistical approaches were employed. First, a one-way ANOVA (Analysis of Variance) was conducted to assess whether there are statistically significant differences in the means of various variables across the landcover classes (McHugh, 2011). The rationale behind this is that if a variable shows significant differences between classes it implies that the variable can effectively differentiate between those classes (McHugh, 2011). For instance, a significant difference in a vegetation index between Forest and Cropland would suggest its usefulness in distinguishing these landcover types.

Following the ANOVA, Tukey's post-hoc test was performed. While ANOVA can indicate if there is a significant difference among group means it does not specify which groups differ from each other (McHugh, 2011). Tukey's test addresses this by providing detailed pairwise comparisons between groups (Voss, 2015) allowing to pinpoint exactly which landcover classes differ from each other for each significant variable identified by ANOVA. This step is crucial for understanding the differences between the landcover classes. After identifying the significant variables, Pearson correlation analysis was conducted to quantify the strength and direction of the linear relationship between pairs of variables (Chok, 2010; Coccia, 2020). This analysis aimed to understand the interaction and influence of the variables on each other. Highly correlated variables were removed to enhance the model's efficiency by eliminating redundant information. The Pearson correlation coefficient ranges from -1 to 1, with 1 indicating a perfect positive linear correlation, -1 indicating a perfect negative linear correlation, and 0 indicating no linear correlation (Coccia, 2020). To manage redundancy, variables with a high Pearson correlation coefficient (r)> 0.9 were further examined using a variable importance score (Asuero et al., 2006; Schober & Schwarte, 2018;).

The MDG metric (Mean Decrease in Gini) was used to evaluate the importance of each variable by measuring its contribution to the homogeneity of the nodes and leaves in the Random Forest model

(Nicodemus, 2011). This ensures that the selected variables are the most influential for accurately distinguishing between the landcover classes. By integrating one-way ANOVA, Tukey's post-hoc test, Pearson correlation, and MDG analysis, the study ensured identification and prioritization of the most effective variables for landcover classification.

2.2.7.   Evaluation of RF model

Evaluating the performance of the RF algorithm is important for understanding its effectiveness in making accurate predictions in landcover classification. This study used User Accuracy (UA), Producer Accuracy (PA), and the F1 Score for this evaluation. UA (Precision) measures the correctly predicted positive observations to the total predicted positive observations (Powers & Ailab, 2020) answering the question: "of all the observations that were predicted to be in a given class how many were correctly classified?" The equation for calculating UA is:

$$UA = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

PA measures the correctly predicted positive observations to all observations in the actual class (Powers & Ailab, 2020). This metric is important because it evaluates the model's ability to correctly identify instances of a specific class. The formula for calculating PA is:

$$PA = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

The F1 Score provides a single metric that balances the trade-off between precision and recall making it particularly useful when the class distribution is imbalanced (Powers & Ailab, 2020). The formula for calculating the F1 Score is:

$$\text{F1 Score} = 2 \text{ X } \frac{\text{User Accuracy X Producer Accuracy}}{\text{User Accuracy X Producer Accuracy}}$$

According to (Powers & Ailab, 2020) UA, PA and F-1 score are crucial for model validation especially in cases of imbalanced data where accuracy alone can be misleading. By using these metrics, we can evaluate the performance of the RF model in landcover classification ensuring a reliable assessment.

2.2.8   Change detection and reforestation detection

This study used an image differencing technique in ArcMap (ESRI, 2018) to calculate areas that have transitioned into different classes over time. This method involves subtracting one temporal image from another (Panuju et al., 2020) to identify changes in landcover between specific years (1998-1993, 2004-1998, 2008-2004, 2013-2008, 2018-2013, 2023-2018). Image differencing was performed using the raster calculator in ArcMap (ESRI, 2018) to highlight changes by subtracting pixel values of earlier images from later ones (Panuju et al., 2020). The entire process is straightforward and computationally efficient requiring basic mathematical computations and provides a quantitative measure of change, allowing for precise calculation of both the magnitude and direction of changes (Lu et al., 2004; Panuju et al., 2020). This makes it a valuable tool for analysing landcover changes over time.

To quantify the extent of forest cover loss over the period from 1993 to 2023 this formula was used: Forest Loss = Net Change in Forest Cover + Regeneration from Other Landcover Classes. Net Change in Forest Cover represents the difference in forested area between the initial and final years of the study period. This change reflects both the loss of forest due to deforestation or land conversion and any gain from natural or assisted regeneration. The term Regeneration from Other Land Cover Classes accounts

for areas that were not originally forested at the beginning of the period but have since been converted back to forest.

To detect forest regeneration, a function   was created in google earth engine to identify specific landcover changes over a three-year period. The function requires two inputs thus a classified collection of landcover data and a sequence of three years (start, middle, and end). For each year in the specified three-year period, the function retrieves the corresponding landcover image from the dataset which shows how landcover was classified in that particular year. The function then checks for two key conditions. Firstly, areas that were forested in the start year but became non-forested in the middle year and secondly, areas that were non-forested in the middle year but returned to being forested by the end year. These conditions help pinpoint regions that have undergone the transition pattern indicative of forest regeneration. The function creates a mask that highlights areas where the specific landcover change pattern is detected. This mask shows which regions have experienced forest regeneration. The function repeats this process for every possible three-year period within the dataset. Providing a detailed view of how forests have regenerated over time. A map is then generated from the output and areas that changed were calculated by converting the pixel count to square kilometres. The regenerated forest was further assessed using this formula: Regenerated = Regenerated Forest Area/Time period (years).

3. RESULTS

This chapter presents the findings of the study on analysis of variance using ANOVA, pairwise comparison among class means using Tukey's HSD test, evaluating correlation among variables, variable importance in landcover classification model, evaluation of the classification result, landcover change detection statistics, forest loss and forest regeneration statistics.

3.1. Analysis of variance among variables

From Table 5 below, higher F-values greater than the critical value imply significant differences between group means. The results of the ANOVA for various variables are summarized below. The ANOVA results indicate that there are significant differences (p-value <0.05) among the groups for all the variables tested except Aspect. The variables with F-values significantly greater than the F-critical value (2.627441 as shown in appendix, table 1) indicates that there are significant differences between group means. These variables are Homogeneity, DVI, Contrast, Elevation, Blue, Green, Red, EVI, SAVI, Wetness, Brightness, Entropy, Greenness, SWIR2, SWIR1, NDWI, NDVI, GNDVI, NIR and Slope. Aspect has an F- value (1.09386) less than the critical value indicating no significant difference between group means and was removed from further analysis.

Table 5: ANOVA results for various variables

Variables	P-value	F-score
Homogeneity	9.71E-56	121.3288
DVI	1.49E-67	158.731
Elevation	1.98E-34	65.54534
Blue	9.98E-80	203.036
Green	3.08E-70	167.9708
Red	1.39E-89	243.7862
NDWI	1.4E-87	235.1226
GNDVI	1.4E-87	235.1226
SWIR 1	1.6E-109	341.987
Contrast	1.28E-21	38.03352
NDVI	9.7E-129	460.794
SWIR 2	8.5E-112	354.7006
Entropy	1.95E-17	29.88421
Brightness	1.5E-84	222.3969
Greenness	4.73E-94	263.8601

Wetness	1.2E-122	420.3086
SAVI	9.26E-95	267.1363
EVI	9.26E-95	267.1363
Slope	0.005095	4.33011
NIR	2.16E-25	45.72288
GNDVI	1.4E-87	235.1226
Aspect	0.351534	1.09386

3.2. Pairwise comparison among class means using tukey's HSD test

The results of the Tukey test for homogeneity as presented in Table 6 indicate significant differences between several pairs of landcover categories. Specifically, statistically distinct homogeneity was found between Cropland and Built-up (p = 0.0005), Forest and Built-up (p = 0.0000), Forest and Cropland (p = 0.0000), Other and Cropland (p = 0.0001), and Other and Forest (p = 0.0000). However, no significant difference was observed between Other and Built-up (p = 0.9910) suggesting similar homogeneity between these categories. For slope, the Tukey test results indicate that most pairwise comparisons did not show statistically significant differences: Cropland vs. Built-up (p = 0.6447), Forest vs. Built-up (p = 1.0000), Forest vs. Cropland (p = 0.6289), and Other vs. Cropland (p = 0.2230). However, significant differences were found between Other and Built-up (p = 0.0117) and Other and Forest (p = 0.0107) suggesting that the slope values for Other landcover significantly differ from those of Built-up and Forest areas. The NDVI analysis showed significant differences across all pairwise comparisons of landcover categories: Cropland vs. Built-up (p = 0.0000), Forest vs. Built-up (p = 0.0000), Other vs. Built-up (p = 0.0000), Forest vs. Cropland (p = 0.0000), Other vs. Cropland (p = 0.0000), and Other vs. Forest (p = 0.0000). This indicates that the NDVI values are significantly different between all landcover categories suggesting distinct vegetation characteristics.

Table 6 : Tukey HSD test for variables

Index	Comparison	diff	lwr	upr	p adj
Homogeneity	Cropland-Built up	0.079002	0.027822	0.130183	0.0005
	Forest-Built up	0.322474	0.271294	0.373655	0.0000
	Other-Built up	-0.00586	-0.05704	0.045318	0.9910
	Forest-Cropland	0.243472	0.192421	0.294524	0.0000
	Other-Cropland	-0.08486	-0.13592	-0.03381	0.0001
	Other-Forest	-0.32834	-0.37939	-0.27728	0.0000

	Cropland-Built up	0.441717	-0.53035	1.413785	0.6447
	Forest-Built up	-0.00828	-0.98035	0.963785	1.0000
	Other-Built up	1.161717	0.189649	2.133785	0.0117
	Forest-Cropland	-0.45	-1.41962	0.519622	0.6289
Slope	Other-Cropland	0.72	-0.24962	1.689622	0.2230
	Other-Forest	1.17	0.200378	2.139622	0.0107
	Cropland-Built up	0.160867	0.129639	0.192096	0.0000
	Forest-Built up	0.346006	0.314778	0.377235	0.0000
	Other-Built up	-0.06353	-0.09476	-0.03231	0.0000
	Forest-Cropland	0.185139	0.153989	0.216289	0.0000
NDVI	Other-Cropland	-0.2244	-0.25555	-0.19325	0.0000
	Other-Forest	-0.40954	-0.44069	-0.37839	0.0000

3.3. Evaluating correlation among variables

From Figure 3, it is observed that variables namely homogeneity, contrast, slope, elevation, NIR, NDWI, entropy, and wetness have correlation coefficients less than 0.9, indicating they are not highly correlated and should be kept. These variable provides different information and do not share similar information.

On the other hand, highly correlated pairs with coefficients greater than 0.9 were found among the variables Red, Green, Blue, SWIR1, SWIR2, and brightness. These variables share similar information and therefore one of these variables should be selected. Also, NDVI, DVI, EVI, GNDVI, SAVI, and Greenness were also highly correlated among each other. While the r=0.9 is a cut-off used by Dormann et al., (2013), in this analysis the selection of variables was driven primarily by expert knowledge, given the study's objectives related to forest analysis. Variables like Red and NDVI were selected based on their known importance in forest monitoring and analysis, rather than solely on statistical criteria. Red was chosen from the group of highly correlated spectral variables (Red, Green, Blue, SWIR1, SWIR2, and brightness) due to its relevance in detecting vegetation health and stress. Similarly, NDVI was selected from the group of vegetation indices (NDVI, DVI, EVI, GNDVI, SAVI, and Greenness) because of its established role in assessing vegetation density and forest cover.
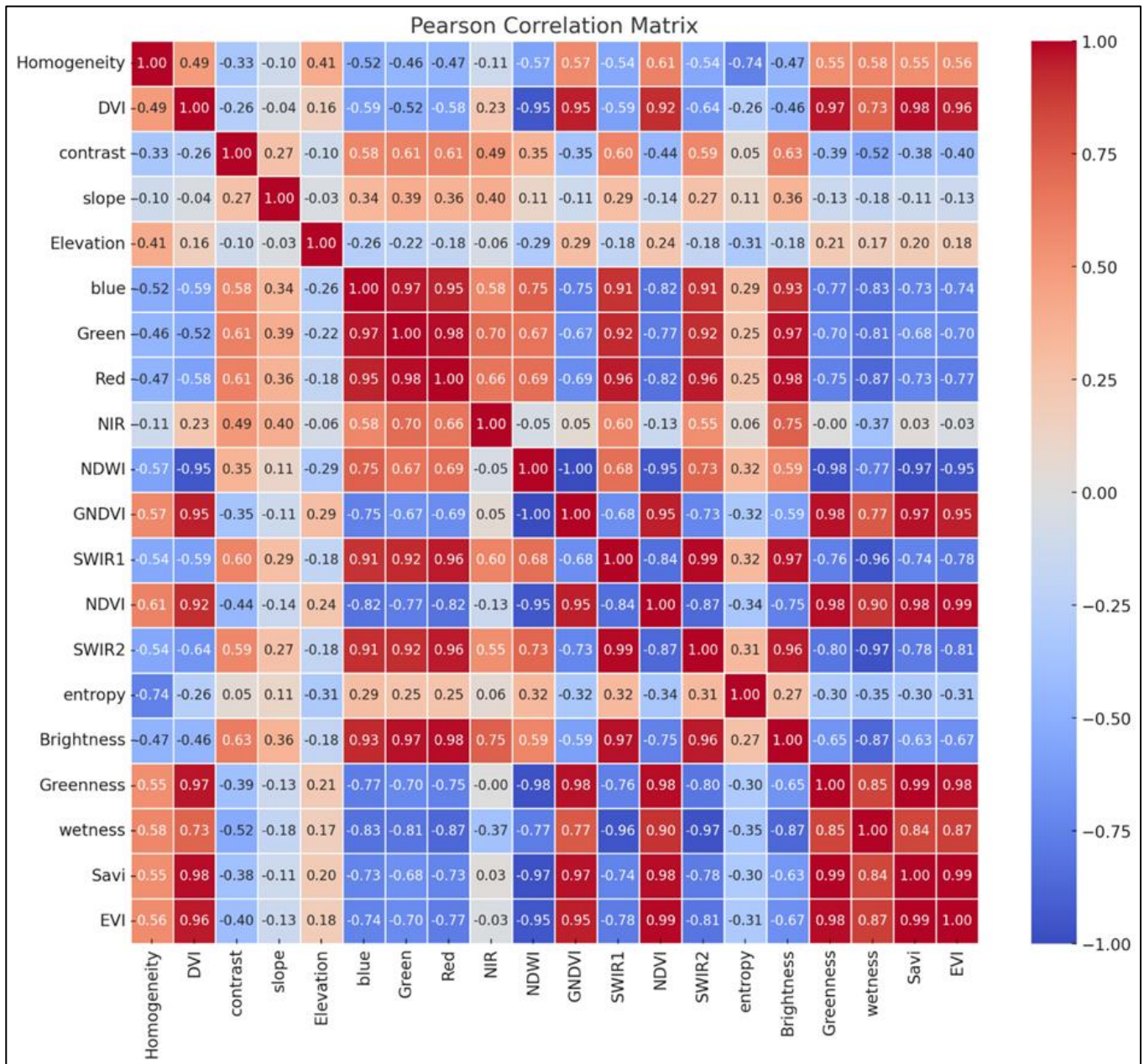
Figure 3: A Pearson correlation matrix showing the relationship among variables, high correlation among (Red, Green, Blue, SWIR1, SWIR2, brightness) and (NDVI, DVI, EVI, GNDVI, SAVI, and Greenness) and low correlation among (homogeneity, contrast, slope, elevation, NIR, NDWI, entropy, and wetness)

3.4.   Variable importance in landcover classification model

Further analysis was conducted to select the most important variable from each of these groups as shown in appendix, Figure 6. Red was chosen from the first group due to its highest variable importance score above 40 and NDVI being selected from the second group with an importance score above 20 as shown in Figure 4 below. From Figure 4, each bar represents the importance score of a variable with higher scores indicating greater importance in the landcover classification. NDVI, contrast, elevation, homogeneity, wetness and NDVI have scores above 15 indicating their importance in the classification process whereas slope has the least score below 10. This ranking underscores the diverse factors that collectively enhance the accuracy and reliability of the landcover classification model.
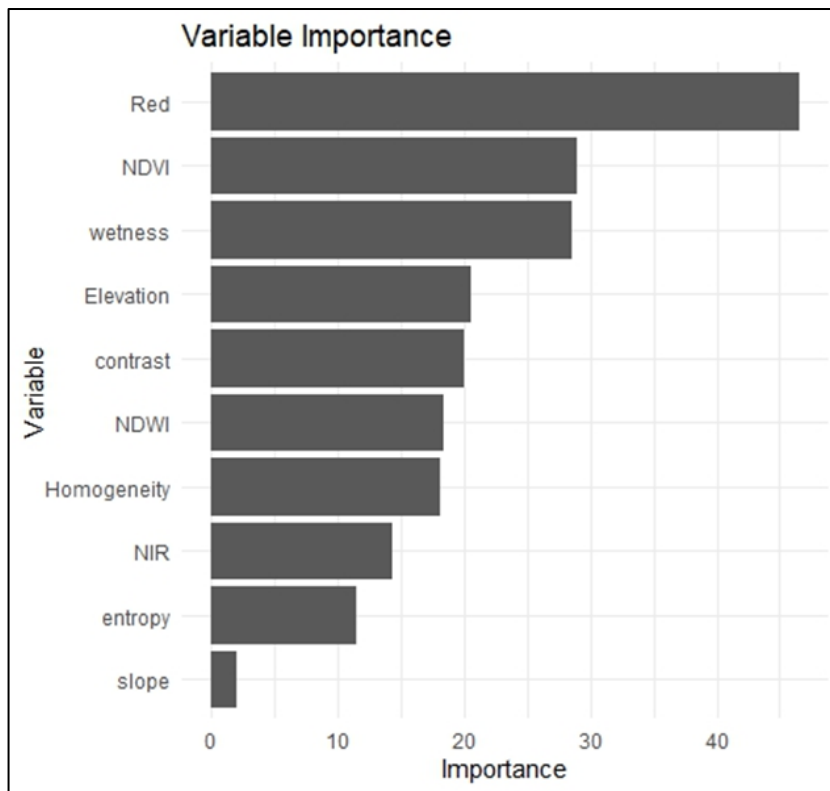
Figure 4:Ranking of different variables in landcover classification

Table 7 below shows the selected variables for classification based on the variable selection process that have been explained above. The final selection consists of the following variables, Red, NDVI, Wetness, Elevation, Contrast, NDWI, Homogeneity, NIR, Entropy, and Slope as shown in the table.

Table 7: Selected variables for classifying landcover

Red
NDVI
Wetness
Elevation
Contrast
NDWI
Homogeneity
NIR
Entropy
Slope

3.5.   Evaluation of the classification results

In this study landcover maps generated shown in figure 5 illustrate the landcover classification outputs for the years 1993 through 2023. The maps depict the landcover classifications for the years 1993, 1998, 2004, 2008, 2013, 2018, and 2023 displaying the categories Forest, Cropland, Built-Up, and Other. The study reported overall classification accuracies of 84%, 85%, 86%, 85%, 80%, 83%, and 80% for the years 1993, 1998, 2004, 2008, 2013, 2018, and 2023, respectively as detailed in Table 8. Additionally, the RF classification model demonstrated uncertainty levels of 16%, 15%, 14%, 20%, 17%, and 20% for the corresponding years reflecting the model's reliability and the degree of confidence in the classification outputs.

The performance of the RF classification model was evaluated for its effectiveness in categorizing landcover classes using selected variables. Table 8 below presents the model's performance metrics including UA, PA, uncertainty, and the F1 score which provide insights into the model's accuracy in classifying different landcover types. The RF classification model demonstrates high performance for forest areas with UA, PA, F1-score above 80%. On the other hand, the RF classification model was capable of classifying Cropland by achieving UA, PA, F1-score above 79%. Moreover, Built up and Other has UA, PA, F1-score above 56%. Forest class consistently exhibits strong classification performance with high accuracy and minimal confusion with other classes as seen in appendix, confusion matrix (figure 14 to figure 15). Similarly, the "Built up" class also demonstrates robust classification with only occasional misclassifications. The "Cropland" class while generally well-classified does show some minor misclassifications particularly with the "Other" class though these errors are relatively minimal. However, the "Other" category emerges as the most challenging frequently being confused with the "Built up" class (as seen in appendix, confusion matrix (figure 14 to figure 15). This pattern is particularly noticeable and may suggest that the features used to distinguish "Other" from "Built up" are not as distinct as those for other classes. This indicate that RF is a better classifier for Forest and Cropland as compared with Built up and Other based on study site.
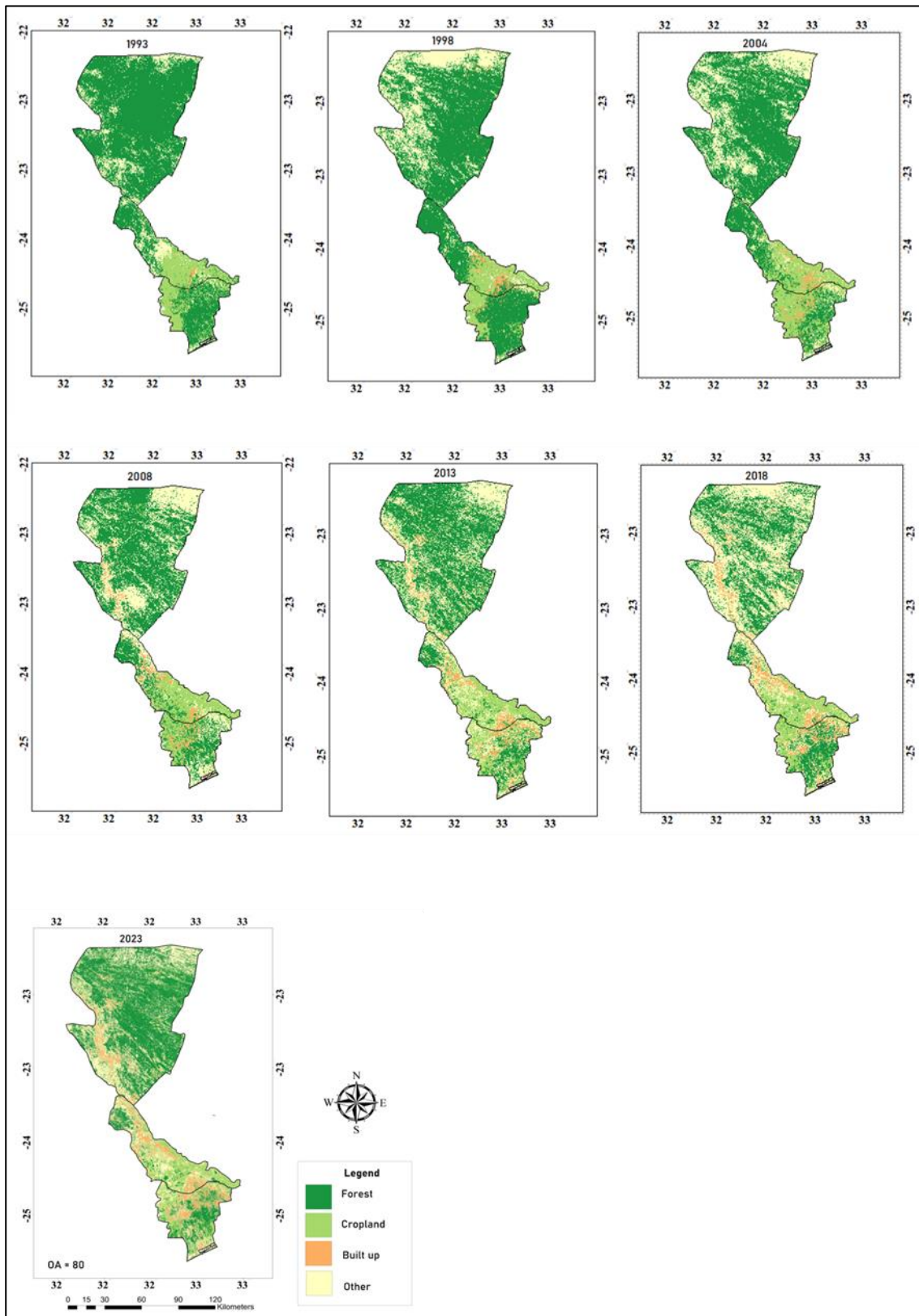
Figure 5: Landcover maps for the year 1993, 1998, 2004, 2008, 2013, 2018 and 202

Table 8: User accuracy, producer accuracy, uncertainty and F1-score by classes for the landcover maps

	1993			**1998**		
Class	User Accuracy (UA) (%)	Producer Accuracy (PA) (%)	F1-Score (%)	User Accuracy (UA) (%)	Producer Accuracy (PA) (%)	F1-Score (%)
Forest	92	92	92	95	88	83
Cropland	80	85	82	79	90	84
Other	82	76	78	85	78	81
Built up	80	84	82	80	84	81
Overall Accuracy (OA)	**84**			**85**		
Uncertainty	0.16			0.15		

	2004			**2008**		
Class	User Accuracy (UA) (%)	Producer Accuracy (PA)	F1-Score (%)	User Accuracy (UA) (%)	Producer Accuracy (PA) (%)	F1-Score (%)
Forest	95	92	93	95	84	89
Cropland	80	100	88	94	85	88
Other	82	63	71	82	73	77
Built up	88	94	90	72	100	83
Overall Accuracy (OA)	**86**			**85**		
Uncertainty	0.14			0.15		

	2013			**2018**		
Class	User Accuracy (%)	Producer Accuracy (%)	F1-Score (%)	User Accuracy (%)	Producer Accuracy (%)	F1-Score (%)
Forest	95	84	89	95	80	86
Cropland	100	90	94	95	90	92
Other	78	65	70	85	78	81

Built up	57	84	67	64	88	74
Overall Accuracy (OA)	**80**			**83**		
Uncertainty	0.20			0.17		

	2023		
Class	User Accuracy (%)	Producer Accuracy (%)	F1-Score (%)
Forest	100	80	88
Cropland	95	95	95
Other	86	56	67
Built up	56	94	70
Overall Accuracy (OA)	**80**		
Uncertainty	0.20		

3.6. Landcover change detection statistics for 1993, 1998, 2004, 2008, 2013, 2018 and 2023

Over the past 30 years, the extent of land coverage for various landcover types has changed as illustrated in Figure 6. Between 1993 and 1998, the forested area decreased by approximately 10% (i.e. from 10592km² to 9474km²). This decline continued from 1998 to 2004, where the forest cover was reduced by an additional 11% (i.e. from 9474km² to 8367km²). The rate of decrease slowed slightly between 2004 and 2008 with an 8% (i.e. from 8367km² to 7655km².) reduction in forested land. However, the downward trend persisted from 2008 to 2013, which saw a 9% (i.e. from 7655km² to 6917km²) decrease. The most significant reduction in forested land occurred between 2013 and 2018 where a dramatic 22% (i.e. from 6917km² to 5377km²) of the remaining forest cover was lost. Interestingly, the trend reversed between 2018 and 2023 where the data indicates a 22% (i.e. from 5377km² to 6613km²) increase in forested area contrary to the previous periods of decline. Overall, from 1993 to 2023, there was an approximate 37% decrease in forested land indicating a significant reduction over the 30-year study period.

In contrast, between 1993 and 1998, there was a substantial 32% (i.e. from 2130km² to 1443km²) decrease in cropland area. However, from 1998 to 2004, cropland expanded significantly by 75% (i.e. from 1443 to 2531km²). This was followed by an 18% (i.e. from **2531**km² to **2072**km²) decrease from 2004 to 2008 and then a 28% (i.e. from 2072km² to 2662m²) increase from 2008 to 2013. The period from 2013 to 2018 saw another large decrease of 32% (i.e. from 2662km² to 1809km²) while the last period from 2018 to 2023 showed a 35% (i.e. from 1809km² to 2442km²) increase.
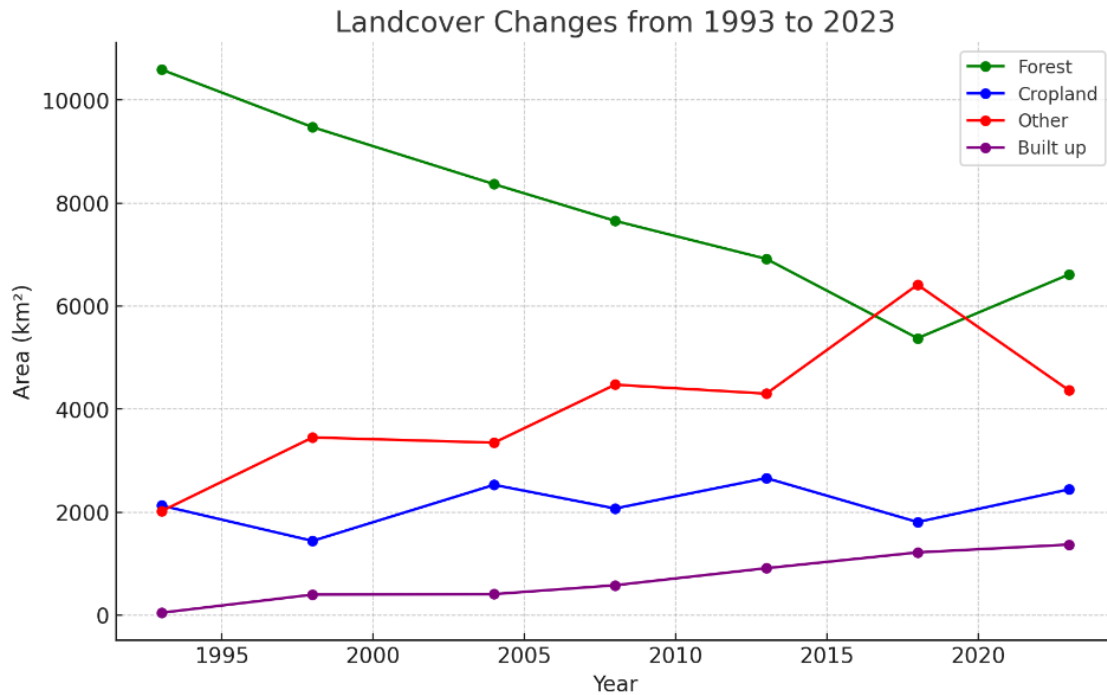
Figure 6: Statistics of landcover changes from 1993 to 2023 measured in km²

The built-up land category reflects the most dramatic changes highlighting rapid urbanization. Between 1993 and 1998, the built-up area increased by an astounding 733% (i.e. from 48km² to 400km²), followed by a modest 2% (i.e. from 400km² to 410km²) increase from 1998 to 2004. Urban expansion continued with a 41% (i.e. from 410km² to 582km²) increase from 2004 to 2008 and an increase of 56% (i.e. from 582km² to 913m²) from 2008 to 2013. The built-up area grew by 33% (i.e. from 913km² to 1219km²) between 2013 and 2018 and by 12% (i.e. from 6412km² to 4368km²) from 2018 to 2023. Overall, the built-up area expanded by a staggering 2746% from 1993 to 2023, underscoring the significant urban growth over the study period.

The "Other" land cover category has seen significant changes as well. From 1993 to 1998, this category increased by a substantial 71% (i.e. from 2018km² to 3451km²), followed by a slight decrease of 2% (i.e. from 3451km² to 3349km²) from 1998 to 2004. The area classified as "Other" grew by 33% (i.e. from 3349km² to 4475km²) from 2004 to 2008 before seeing a minor 3% (i.e. from 3349km² to 4475km²) decrease from 2008 to 2013. A notable increase of 49% (i.e. from 4301km² to **2072**km²) occurred between 2013 and 2018 with a subsequent 6% (i.e. from 2072km² to 2662km²) decrease from 2018 to 2023. Over the 30-year period the "Other" landcover category expanded by 116% (i.e. from 6415km² to 4368km²) indicating a significant shift in landcover.

3.7.   Regenerated forest cover and forest cover loss statistics

Over the 30-year period from 1993 to 2023 there has been a significant disparity between forest regeneration efforts and the rate of forest cover loss as shown in Figure 7. While a total of 2,456 km² of forest was regenerated, with an average annual rate of 81.9 km², the total forest cover loss amounted to 7,288.8 km² averaging 242.96 km² per year. Breaking down the trends by periods, from 1993 to 2004, regeneration accounted for 1,085 km² (10% of forest cover) compared to a loss of 2,323.6 km² (21%). In the subsequent periods, regeneration rates decreased significantly: 655 km² (6.9%) from 1998 to 2008, 282 km² (3.3%) from 2004 to 2013, and just 152 km² (2%) from 2013 to 2023. Meanwhile, forest cover loss fluctuated but remained substantial, peaking at 2,280.2 km² (29%) between 2008 and 2018.
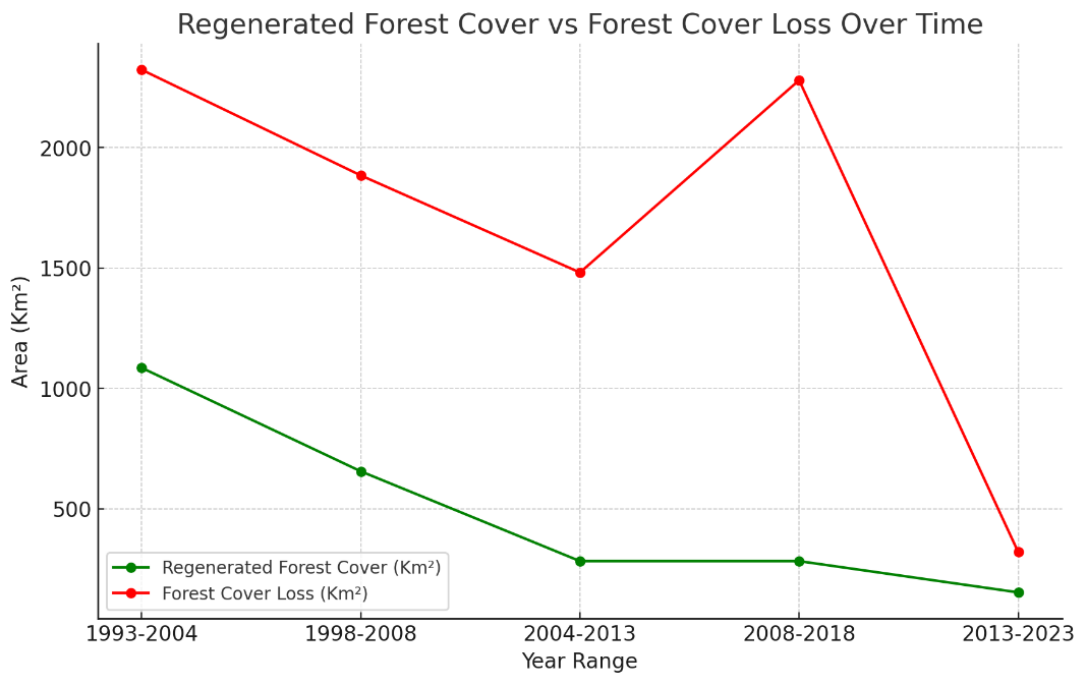
Figure 7:Statistics of regenerated forest cover and forest cover loss from 1993 to 2023 measured in km²

In Figure 8 it was observed that Mabalane experienced more forest regeneration compared to Chokwe and Bilene during the period from 1993 to 2004. Notably, from 1998 to 2008, Bilene witnessed a progress in forest cover regeneration. Moreover, from 2004 to 2013, Mabalane once again outpaced the other two districts in forest regeneration surpassing their combined totals. From 2013 to 2023, Mabalane witnessed a decline in forest regeneration as compared to the previous years.
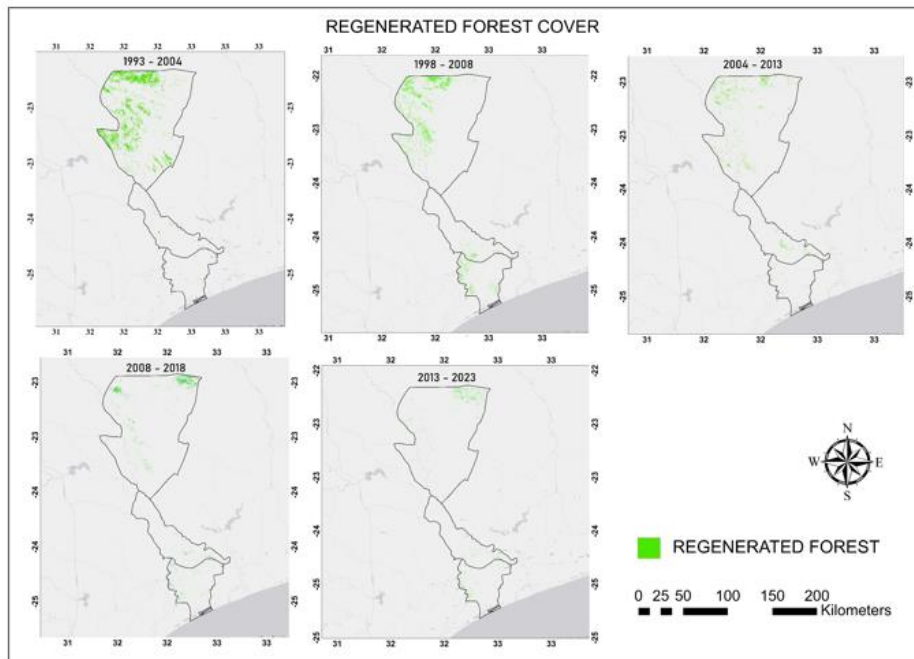
Figure 8: Regenerated forest cover map for the period 1993-2004, 1998-2008, 2004 – 2013, 2013-2018 and 2018 - 2023

Deforestation maps in Figure 9 illustrates the maps of forest cover loss from 1993 to 2023 in the study area. During the period from 1993 to 1998, Mabalane experienced a higher rate of forest cover loss compared to Bilene and Chokwe. In the subsequent period, 1998 to 2004, Bilene saw a greater loss of forest cover than Chokwe although Mabalane continued to have the most extensive forest cover loss during this time. Between 2004 and 2008 Chokwe experienced more forest cover loss than Bilene yet Mabalane again had the highest overall loss. In the period from 2008 to 2013 deforestation remained prevalent across all districts; however, it was observed that Mabalane began to see an increase in forest coverage between 2013 and 2018. Finally, from 2018 to 2023 the rate of forest cover loss declined across all three districts.
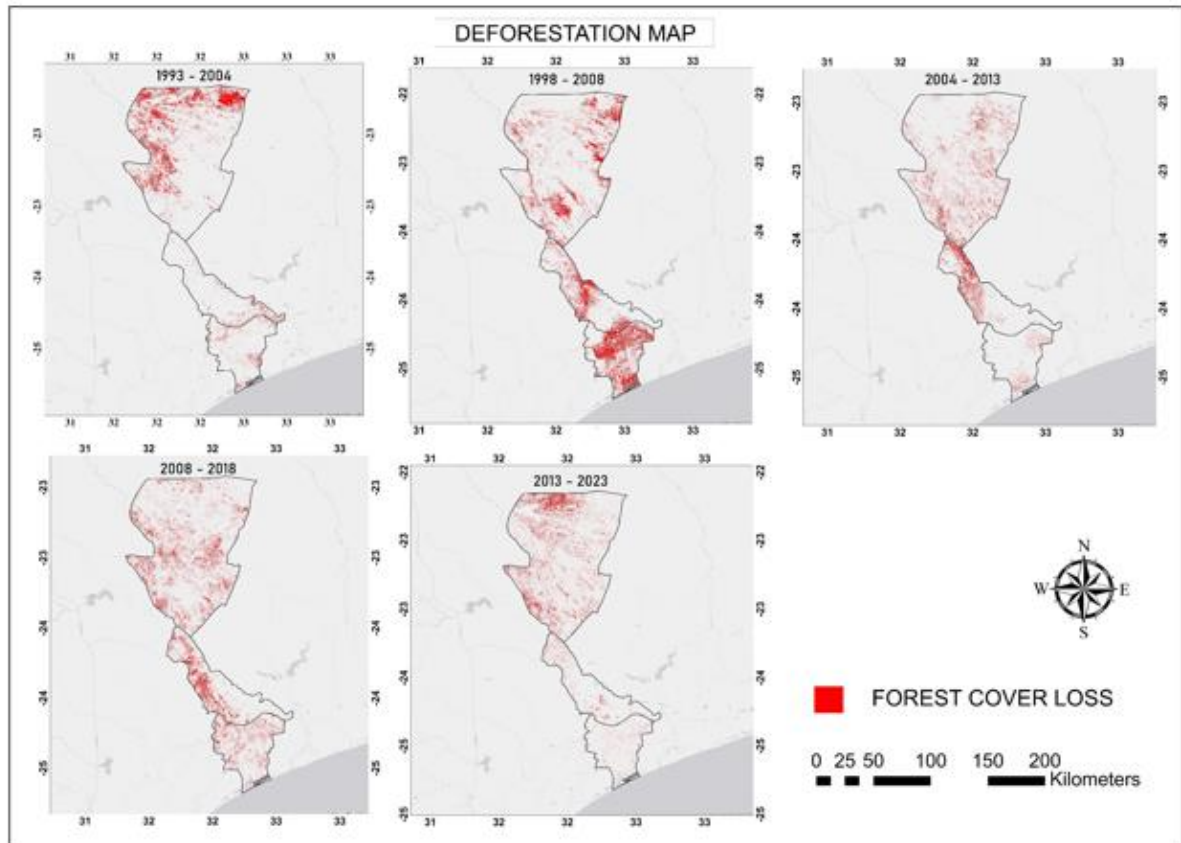
Figure 9: Deforestation maps for the period 1993 – 2004, 1998 - 2008, 2004 – 2013, 2008-2018, and 2013-2023

4. DISCUSSION

4.1. Analysis of variance among variables

The One-Way ANOVA analysis in this study revealed significant difference among group means for spectral bands, vegetation indices, topographic features, textural features, and tasselled cap features. These differences are attributed to the distinct characteristics of the landscape captured by these variables, such as vegetation health, surface reflectance, and spatial patterns which varies across different landcover types. For instance, spectral bands like NIR are sensitive to vegetation while indices like NDVI highlight differences in vegetation density(Matsushita et al., 2007). Topographic features such as elevation and slope influence environmental conditions, for instance, differences in slope lead to variations in air and soil temperature, moisture content, and evaporation rates, thereby creating distinct microclimatic zones that are closely associated with alterations in vegetation structure and composition (Singh, 2018). A study by Alvarez & Naughton-Treves (2003) in the Peruvian Amazon shows that roads and rivers serve as primary conduits for forest clearing, with roadside deforestation often at higher rates than riverside clearing. The difference among group means of the variables in the study represent well-defined categories and the One-Way ANOVA detects these differences. However, high p-value (>0.05) for "Aspect" is expected as its influence on vegetation and surface feature is nuanced compared to other factors. Other variables like elevation and spectral indices capture most of the variability thus, rendering aspect's contribution relatively low. For example, a study by Cheng et al. (2023) in northwestern Yunnan, China found that while elevation and slope had significant effect on vegetation structure and composition, aspect's influence was less pronounced suggesting how other variables such as elevation capture more variability in vegetation. On the other hand, a study by Stage & Salas (2007) conducted in Utah in the United State suggest that elevation interacts with aspect and slope to influence forest productivity and species composition.

4.2. Evaluating correlation among variables

The strong correlations observed among certain variables as outlined in section 3.3 (Figure 3) can be explained by their inherent characteristics. Spectral bands like Red, Green, Blue, SWIR1, and SWIR2 show high correlation because they capture similar reflectance properties particularly in non-vegetated areas such as urban regions and bare soil which tend to reflect light in a similar manner across these bands (Malamiri et al., 2023). For example, Meti et al. (2019) classified alkaline soils in Karnataka, India using Sentinel-2 and Landsat-8 bands and found that these spectral bands exhibited high correlation due to their ability to capture similar reflectance characteristics in non-vegetated surfaces. SWIR1 and SWIR2 operating in the shortwave infrared region are sensitive to moisture content in vegetation and soil. This makes them useful for detecting water stress, mapping soil properties, and distinguishing materials based on moisture content (Le et al., 2023; Swathandran & Aslam, 2019). Similarly, Madonsela (2018) classified tree species diversity in South Africa's savannah woodlands and found that Landsat-derived spectral variables especially those in the SWIR region were effective in explaining tree species diversity due to their sensitivity to moisture. Vegetation indices such as NDVI, DVI, EVI, GNDVI, and SAVI are primarily derived from the Red and Near-Infrared (NIR) regions of the electromagnetic spectrum. These indices often show high correlation because they are designed to measure closely related aspects of vegetation using similar spectral bands. Jopia et al. (2020) found that these indices while offering unique insights under specific conditions frequently exhibit high correlation due to their shared purpose. This is further supported by da Silva et al. (2020) who observed that indices like NDVI, EVI, and SAVI show high

correlation due to their similar spectral characteristics. The lower correlation coefficients (below 0.9) among variables like homogeneity, contrast, slope, elevation, NIR, NDWI, entropy, and wetness indicate that these variables are not highly correlated. Each variable reflects distinct environmental factors, such as vegetation, water content, topography, or image texture (Xie et al., 2023). This finding aligns with Lane et al. (2014) who used eight-band high-resolution satellite imagery to classify wetland areas and observed similarly low correlations among these variables. This suggests that each variable captures unique environmental aspects and does not overlap significantly with the others.

4.3.   Variable importance

The five most important variables derived after running the RF model with the selected variables will be discussed in relation to previous research, where these variables were important for land cover classification. They are in the following order: the red band, NDVI, wetness, elevation, contrast. The finding are in agreement with the research of Henareh Khalyani et al. (2012) on detecting Zagros forest in Iran using NDVI, elevation and red band as important variables for classifying and detecting Zagros forest. Also a study by Jin et al. (2018)shows that texture measure such as GLCM contrast proved important in landcover mapping in central Shandong. The red band of LANDSAT imagery was the most important variable. This is justified by its effectiveness in vegetation detection. Chlorophyll- in plants absorbs red light in the visible spectrum. The red band on the visible part of the electromagnetic spectrum is a strong indicator for differentiating forests and other vegetation because in dense vegetation like forests, red light will have a less reflection as compared to areas with sparse or no vegetation (Gutman et al., 2021). The red band is important in distinguishing between vegetation and other classes like built up areas, as concrete, asphalt and materials used for building rooftops have a high reflectance in the red band (XI et al., 2019). Similarly, NDVI values which quantify vegetation cover by using a mathematical equation involving both the red band and the NIR band. NDVI has the potential to distinguish dense vegetation and sparse vegetation, as well as non-vegetated areas like built up areas (Zheng et al., 2021). High NDVI values generally mean that the red band reflectance is low compared to the NIR reflectance, which corresponds to healthy vegetation that is actively photosynthesizing. The wetness variable is the next important feature and can capture the hydrological characteristics of different landcover classes. For example, built-up areas have lower wetness values mostly because surfaces like asphalt and concrete do not have high water retention capacity (Yesilnacar & Süzen, 2006). Conversely, even though affected by other human and environmental factors, forests and other vegetation, generally have a higher potential to retain water thereby increasing the wetness values (Yesilnacar & Süzen, 2006). Elevation provides additional context about the physical environment which influence the distribution of landcover types. Elevation affects the microclimates which influences the types of landcover in an area. An example of this can be seen in Table 6 of Naqvi et al., (2013). Elevation is important for understanding watershed and drainage systems which in turn influence built-up area expansion. However, In Mozambique, a study by Lisboa et al., (2024) demonstrated that deforestation levels decline with increasing elevation as areas of low elevation are suitable for smallholder agriculture and logging, and areas of high elevation are difficult to access. The contrast variable derived from GCLM was also important. This measure captures the variation in intensity levels between neighbouring pixels where higher contrast values indicate more pronounced texture.   information is crucial for distinguishing between landcover types with distinct textural characteristics such as differentiating urban areas from natural landscapes (Hall-Beyer, 2017). Furthermore, high contrast values often occur at the boundaries between different landcover types facilitating more precise delineation and classification of these boundaries. This capability is beneficial in

complex environments where features exhibit similar spectral signatures but differ in texture (Park & Guldmann, 2020). The results are valid as they agree with previous studies as demonstrated above.

4.4. Evaluation of classification

In the context of landcover classification for this study, the RF model's reported overall accuracies ranging between 80% and 86% across different years indicate a robust performance but also reveal some variability over time. For example, a study by Macarringue et al., (2023) performed landcover mapping using RF classification method in Mozambique and achieved comparable accuracies further validating the effectiveness of the RF model in this region. Specifically, their study, which also utilized moderate-resolution satellite data reported overall accuracies close to those observed in this study reinforcing the reliability of the RF classification model.

One of the strengths of the RF model is its ability to accurately classify forest areas as evidenced by the UA, PA, and F1-scores consistently exceeding 80%. This high accuracy is largely due to the distinctive spectral signatures of forests which are easier to differentiate from different landcover types. Forests typically exhibit a high NDVI and distinct textural features making them readily separable in classification tasks (Mucova et al., 2018). Observations from (Macarringue et al., 2023; Mucova et al., 2018) highlighting that the distinctiveness of forest spectral signatures contributes to the model's strong performance in this domain. Moreover, the temporal consistency in classification accuracies with values ranging from 84% to 86% across most years suggests that the features used by the RF classification model such as spectral bands and vegetation indices remain effective across different temporal datasets. Despite the high performance the RF model's accuracy shows a decline to an 80% in certain years such as 2013 and 2023. This decline is attributed to changes in landcover that introduce greater complexity into the classification task. For example, factors such as urban expansion, agricultural intensification, or natural disturbances (e.g., deforestation) result in more mixed pixels making accurate classification more challenging. For instance, Lisboa et al., (2024) highlights an increase in settlement and the intensification of smallholder agriculture in Mozambique, these activities not only contribute to landcover changes however, result in intermingling of urban, agriculture and natural areas making it difficult for remote sensing technologies to distinguish between different landcover types.

Another common challenge in landcover classification particularly highlighted in the context of landcover classification in this study is the confusion between certain classes, such as the "Other" and "Built-Up" categories. These classes often share similar spectral properties especially in areas with degraded landscapes leading to misclassifications. The difficulty in distinguishing mixed landcovers or transitional zones exhibit overlapping spectral characteristics is a well-documented issue in remote sensing (Chen et al., 2016). Another source of confusion is the similarity in spectral features between "Cropland" and "Other" classes where common elements like soil reflectance or barren land present in both categories (Nguyen et al., 2021). This overlap in spectral signature presents a challenge in remote sensing classification as the spectral resolution of sensors can limit the ability to distinguish between similar landcover types.

While the RF model was employed in this study, other methods like Support Vector Machines (SVM) and Convolutional Neural Networks (CNNs) have also shown significant promise in landcover classification tasks. For instance, Qian et al. (2014) utilized SVM for urban landcover classification and achieved a high accuracy exceeding 90%. Similarly, Fayaz et al. (2024) demonstrated the effectiveness of deep learning

models, specifically CNNs in landcover classification with accuracies surpassing 90%. These studies suggest that alternative methods such as SVM and neural networks have the potential to deliver even higher accuracies than RF.

4.5.   Change detection statistics for detecting landcover

This research highlights significant changes in landcover over the past three decades emphasizing the dynamic shifts in landcover between 1993 and 2023. One of the most prominent findings is the substantial reduction in forested areas accompanied by a consistent expansion of croplands and an increase in built-up areas. These trends are exemplified in regions like Chókwè where the expansion of cropland is largely driven by the growing demand for rice production. Studies by Ismael et al. (2021) and de Sousa et al. (2019) underscore the extensive landcover changes associated with the Chókwè Irrigation Scheme (CIS) which is Mozambique's largest irrigation project. The conversion of forested areas into agricultural fields to support rice cultivation has directly contributed to the observed decline in forest cover in this region as vast tracts of land have been cleared to meet agricultural needs (Temudo & Silva, 2012a, 2012b).

Urban expansion has contributed to the reduction of forested areas particularly in regions such as Bilene, Mozambique. The growth of urban areas often necessitates the conversion of forested land into agricultural, residential, and commercial zones. Similar studies by Tokura, Matimele, Smit, Timm Hoffman, et al. (2020) in the Licuati Forest Reserve, estimated deforestation due to charcoal production and settlement expansion. Sedano et al., (2016) conducted a study in Tete on charcoal production for urban energy consumption led to substantial forest degradation comparable in magnitude to deforestation. Allan et al. (2017) conducted a study in the Niassa National Reserve and found that, the loss of the forest resort was due to agricultural expansion around settlements and roads.

Woodlands serve as the major source of energy in Gaza with charcoal being the dominant fuel in urban centres as noted by Ugembe et al. (2022). Charcoal production not only provides employment opportunities but affects landcover. Luz et al. (2015) highlights that Gaza Province, particularly Mabalane District is a major supplier of charcoal to Maputo with Mabalane holding the highest number of licenses for charcoal production in the province. This activity further contributes to landcover changes particularly in the form of woodland depletion.

However, the period from 2013 to 2023 accounted for the highest regeneration. This surge in regeneration aligns with global trends observed in recent years where increased awareness of deforestation's impact has led to intensified reforestation efforts(Chazdon & Uriarte, 2016; Daigneault et al., 2022). For example, the United Nation's Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (UN-REDD+) have played a pivotal role during this period providing a framework and resources for large-scale regeneration initiatives (Bayrak & Marafa, 2016b). The data suggests that the recent decade has seen both enhanced policy implementation and improved techniques in reforestation leading to a significant recovery rate. This indicates a shift in priorities towards sustainable forest management and conservation. Forest regeneration plays a crucial role in sequestering carbon as it is indeed one of the most cost-effective strategies for mitigating climate change. For instance, according to Gilroy et al. (2014) regenerating forests can absorb carbon dioxide thus reducing the overall concentration of greenhouse gases in the atmosphere. Additionally, Strassburg et al. (2020) highlight that forest regeneration supports biodiversity by providing habitat for various species which is vital for maintaining

ecological balance. However, the findings underscore that forest regeneration remains slow and gradual process.

(Bayrak & Marafa, 2016b) However, the period from 2013 to 2023 accounted for the highest regeneration. This surge in regeneration aligns with global trends observed in recent years where increased awareness of deforestation's impact has led to intensified reforestation efforts. For example, the United Nation's Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (UN-REDD+) have played a pivotal role during this period providing a framework and resources for large-scale regeneration initiatives (Bayrak & Marafa, 2016b). The data suggests that the recent decade has seen both enhanced policy implementation and improved techniques in reforestation leading to a significant recovery rate. This indicates a shift in priorities towards sustainable forest management and conservation. Forest regeneration plays a crucial role in sequestering carbon as it is indeed one of the most cost-effective strategies for mitigating climate change. For instance, according to Gilroy et al. (2014) regenerating forests can absorb carbon dioxide thus reducing the overall concentration of greenhouse gases in the atmosphere. Additionally, Strassburg et al. (2020) highlight that forest regeneration supports biodiversity by providing habitat for various species which is vital for maintaining ecological balance. However, the findings underscore that forest regeneration remains slow and gradual process.

4.6.   Limitations and recommendations

This study faced several limitations. Among those was the 30-meter spatial resolution of the Landsat 5 and Landsat 8 satellites used. While effective for broad scale landcover mapping, this resolution presents challenges when applied to specific landcover classes such as forests, croplands, and built-up areas. It is generally suitable for mapping large, contiguous areas like dense forests, but struggles to capture smaller patches or regions with mixed landcover types. This limitation could have led to potential misclassification, particularly in areas with fragmented forests, interspersed land classes, or small patches of barren soil that may blend with adjacent landcover types. While large cropland areas can be mapped with high accuracies (UA and PA over 80%), the resolution is inadequate for smaller fields, especially in regions with mixed cropping systems.

Another limitation of this study lies in the number of classes used in the classification scheme. Although mixed cropping systems and other landcover types often have distinct spectral responses, the classification was likely too simplified to adequately capture this diversity and complexity, particularly within smaller plots. This oversimplification could have led to a lack of differentiation between various croplands or cropping systems, especially in areas characterized by heterogeneous land cover. As a result, the limited number of classes may have been insufficient to fully represent the variability within the study thereby reducing the overall accuracy of the landcover maps and hindering the ability to accurately distinguish between different landcover types within the study area.

Moreover, while the Landsat 5 and Landsat 8 satellites provided the necessary temporal coverage for this study, spanning the period from 1993 to 2023, the analysis was limited by their spectral resolution. The relatively broad spectral bands available on these satellites, while sufficient for broad scale landcover mapping, constrained the ability to distinguish between more nuanced landcover classes. This limitation led to the classification being restricted to four broad categories: forest, cropland, built-up areas, and an "other" category. The cropland class included a variety of crops, while the "other" category encompassed diverse landcover types such as sand dunes, barren land, and rocky areas. A finer spectral resolution, such as that offered by hyperspectral sensors with their numerous narrow bands, could have allowed for a more detailed classification, potentially improving the accuracy of the analysis and enabling the differentiation of these more complex and diverse landcover types. This limitation suggests that incorporating data from sensors with finer spectral resolution could have significantly enhanced the quality and precision of the study's outcomes.

To address some of the limitations identified in this study, one effective strategy would involve integrating ancillary data sources such as LiDAR or Digital Elevation Models (DEMs). This could also mitigate the limitations posed by the spectral resolution of Landsat satellites. LiDAR provides detailed 3D information on vegetation structure, significantly improving the detection and classification of small forest patches, built-up areas, and other complex landcover types. DEMs offer valuable topographical data, accounting for elevation and slope variations that influence vegetation patterns. By combining these datasets with traditional satellite imagery, future studies can achieve more precise classification outcomes, particularly in areas where spectral data alone may not suffice.

Also, expanding the classification scheme to include additional landcover classes could significantly improve the understanding of the landscape. The current classification may have oversimplified the landscape by restricting it to broad categories such as forest, cropland, built-up areas, and "other." By incorporating additional classes, such as shrubland, or distinguishing between different types of croplands, future studies could better capture the diversity and complexity of the study area, particularly in regions with heterogeneous landcover. This approach would refine the classification scheme, avoiding oversimplification and enhancing the ability to accurately differentiate between various landcover types, resulting in more detailed and informative landcover maps.

Incorporating hyperspectral data, where available, would further enhance the spectral resolution of the study. Hyperspectral sensors capture data across hundreds of narrow spectral bands, allowing for the

identification of subtle differences between landcover types that might be indistinguishable using multispectral data alone. This finer spectral resolution could greatly improve the accuracy of landcover classification and provide a deeper understanding of the ecological and environmental dynamics within the study area.

Moreover, exploring alternative machine learning algorithms or ensemble methods is another recommended strategy. While the Random Forest (RF) algorithm proved effective in this study, algorithms such as SVM, Neural Networks, and Gradient Boosting Machines (GBM) offer different strengths and could be tested in similar contexts to assess their effectiveness compared to RF. Additionally, ensemble approaches that combine the strengths of multiple algorithms could yield more robust and accurate classification results. Experimenting with these advanced techniques could help identify the most suitable algorithm for specific landcover classification challenges, ultimately improving the reliability and precision of the outcomes.

5. CONCLUSION

The study aimed to detect deforestation in the Gaza Province of Mozambique with a focus on Bilene, Mabalane and Chokwe using random forest machine learning algorithm. The main finding from the study was that it is possible to use RF to detect deforestation over the years with high accuracies greater than 80%. Consequently, there was regeneration of the forest from 1994 to 2023. Despite efforts to regenerate a total of 2456 km² of forest, the loss of 7,288.8 km² of forest cover significantly outpaced these efforts. The overall accuracy of the maps generated from the study can be improved by using other data sources like LiDAR and high-resolution satellites (Sentinel 2 and PlanetScope). It can be concluded that a mix of spectral bands, vegetation indices, and tasseled cap features particularly the Red band, NDVI, and wetness are important for accurate classification. The maps generated are valuable for illustrating forest regeneration in Gaza Province where many previous studies have focused on global scales highlighting the potential to apply these methods at a more localized level in this relatively unexplored region. Conclusively this study demonstrates the capability of using the random forest algorithm to detect deforestation and monitor forest regeneration in Mozambique with a high degree of accuracy. The results underscore the effectiveness of combining various remote sensing features for classification and highlight the significant potential for applying these methods to localized, under-mapped areas. This work contributes valuable insights into forest dynamics in Mozambique, providing a strong foundation for future environmental monitoring and conservation efforts in the region.

6. ETHICAL CONSIDERATION

This study was based entirely on secondary data meaning there were no direct interactions with individuals. The research followed the ethical guidelines established by the University of Twente's Research Ethics Policy. These guidelines emphasize the importance of obtaining consent when needed and enforce strict protocols to safeguard data privacy. To ensure confidentiality, all personal information was anonymized and securely stored with access to sensitive data limited to authorized personnel. These efforts underscore a strong dedication to upholding the highest ethical standards in managing data throughout the research process.

LIST OF REFERENCES

Abrefa Danquah, J., Appiah, M., Osman, A., Pappinen, A., & Christine I Nougbodé, O. A. (2019). Geographic Distribution of Global Economic Important Mahogany Complex: A Review. Annual Research & Review in Biology, 34(3), 1–22. https://doi.org/10.9734/ARRB/2019/V34I330154

Allan, J. R., Grossmann, F., Craig, R., Nelson, A., Maina, J., Flower, K., Bampton, J., Deffontaines, J. B., Miguel, C., Araquechande, B., & Watson, J. E. M. (2017). Patterns of forest loss in one of Africa's last remaining wilderness areas: Niassa National Reserve (Northern Mozambique). Parks, 23(2), 40–50. https://doi.org/10.2305/IUCN.CH.2017.PARKS-23-2JRA.EN

Alvarez, N. L., & Naughton-Treves, L. (2003). Linking National Agrarian Policy to Deforestation in the Peruvian Amazon: A Case Study of Tambopata, 1986–1997. Ambio, 32(4), 269–274. https://doi.org/10.1579/0044-7447-32.4.269

Alvino, F. C. G., Aleman, C. C., Filgueiras, R., Althoff, D., & da Cunha, F. F. (2020). Vegetation indices for irrigated corn monitoring. Engenharia Agrícola, 40(3), 322–333. https://doi.org/10.1590/1809-4430-ENG.AGRIC.V40N3P322-333/2020

Antonarakis, A. S., Pacca, L., & Antoniades, A. (2022). The effect of financial crises on deforestation: a global and regional panel data analysis. Sustainability Science, 17(3), 1037–1057. https://doi.org/10.1007/S11625-021-01086-8/TABLES/15

Aouat, S., Ait-hammi, I., & Hamouchene, I. (2021). A new approach for texture segmentation based on the Gray Level Co-occurrence Matrix. Multimedia Tools and Applications, 80(16), 24027–24052. https://doi.org/10.1007/S11042-021-10634-4/TABLES/5

Aquino, A., Lim, C., Kaechele, K., & Taquidir, M. (2018). Mozambique Country Forest Note.

Aria, M., Cuccurullo, C., & Gnasso, A. (2021). A comparison among interpretative proposals for Random Forests. Machine Learning with Applications, 6, 100094. https://doi.org/10.1016/J.MLWA.2021.100094

Arko, T., Mensah, A., Adomako, J., Denton, F., & Obani, P. (2024). The multifaceted socio-ecological impacts of charcoal production on the Afram Plains, Ghana. Trees, Forests and People, 16, 100586. https://doi.org/10.1016/J.TFP.2024.100586

Assede, E. S. P., Orou, H., Samadori, ·, Biaou, S. H., Coert, ·, Geldenhuys, J., Fiacre, ·, Ahononga, C., Paxie, ·, & Chirwa, W. (2023). Understanding Drivers of Land Use and Land Cover Change in Africa: A Review. Current Landscape Ecology Reports 2023 8:2, 8(2), 62–72. https://doi.org/10.1007/S40823-023-00087-W

Bacour, C., Bréon, F. M., & Maignan, F. (2006). Normalization of the directional effects in NOAA–AVHRR reflectance measurements for an improved monitoring of vegetation cycles. Remote Sensing of Environment, 102(3–4), 402–413. https://doi.org/10.1016/J.RSE.2006.03.006

Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014a). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. Remote Sensing Letters, 5(5), 423–431. https://doi.org/10.1080/2150704X.2014.915434

Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014b). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. Remote Sensing Letters, 5(5), 423–431. https://doi.org/10.1080/2150704X.2014.915434

Bayrak, M. M., & Marafa, L. M. (2016a). Ten Years of REDD+: A Critical Review of the Impact of REDD+ on Forest-Dependent Communities. Sustainability 2016, Vol. 8, Page 620, 8(7), 620. https://doi.org/10.3390/SU8070620

Bayrak, M. M., & Marafa, L. M. (2016b). Ten Years of REDD+: A Critical Review of the Impact of REDD+ on Forest-Dependent Communities. Sustainability 2016, Vol. 8, Page 620, 8(7), 620. https://doi.org/10.3390/SU8070620

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31. https://doi.org/10.1016/J.ISPRSJPRS.2016.01.011

Bofana, J., Zhang, M., Nabil, M., Wu, B., Tian, F., Liu, W., Zeng, H., Zhang, N., Nangombe, S. S., Cipriano, S. A., Phiri, E., Mushore, T. D., Kaluba, P., Mashonjowa, E., & Moyo, C. (2020). Comparison of Different Cropland Classification Methods under Diversified Agroecological Conditions in the Zambezi River Basin. Remote Sensing 2020, Vol. 12, Page 2096, 12(13), 2096. https://doi.org/10.3390/RS12132096

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2016). Feature selection for high-dimensional data. Progress in Artificial Intelligence, 5(2), 65–75. https://doi.org/10.1007/S13748-015-0080-Y/TABLES/3

Bullock, E. L., Woodcock, C. E., & Olofsson, P. (2020). Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis. Remote Sensing of Environment, 238, 110968. https://doi.org/10.1016/J.RSE.2018.11.011

Carlson, G. E. (1989). Characteristics of first order intensity entropy as a measure of image region texture. Digest - International Geoscience and Remote Sensing Symposium (IGARSS), 4, 2405–2408. https://doi.org/10.1109/IGARSS.1989.577881

Carrilho, J., Dgedge, G., Santos, P. M. P. dos, & Trindade, J. (2024). Sustainable land use: Policy implications of systematic land regularization in Mozambique. Land Use Policy, 138, 107046. https://doi.org/10.1016/J.LANDUSEPOL.2023.107046

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189–215. https://doi.org/10.1016/J.NEUCOM.2019.10.118

Chazdon, R. L., Brancalion, P. H. S., Laestadius, L., Bennett-Curry, A., Buckingham, K., Kumar, C., Moll-Rocek, J., Vieira, I. C. G., & Wilson, S. J. (2016). When is a forest a forest? Forest concepts and definitions in the era of forest and landscape restoration. Ambio, 45(5), 538–550. https://doi.org/10.1007/S13280-016-0772-Y/FIGURES/3

Chazdon, R. L., & Uriarte, M. (2016). Natural regeneration in the context of large-scale forest and landscape restoration in the tropics. Biotropica, 48(6), 709–715. https://doi.org/10.1111/BTP.12409

Chen, C., Bu, J., Zhang, Y., Zhuang, Y., Chu, Y., Hu, J., & Guo, B. (2019). The application of the tasseled cap transformation and feature knowledge for the extraction of coastline information from remote sensing images. Advances in Space Research, 64(9), 1780–1791. https://doi.org/10.1016/J.ASR.2019.07.032

Chen, C. ;, Chen, H. ;, Liang, J. ;, Huang, W. ;, Xu, W. ;, Li, B. ;, Wang, J., Wu, Q., Shen, X., Li, J., Zhang, C., Chen, C., Chen, H., Liang, J., Huang, W., Xu, W., Li, B., & Wang, J. (2022). Extraction of Water Body Information from Remote Sensing Imagery While Considering Greenness and Wetness Based on Tasseled Cap Transformation. Remote Sensing 2022, Vol. 14, Page 3001, 14(13), 3001. https://doi.org/10.3390/RS14133001

Chen, J. J., Zhen, S., & Sun, Y. (2021). Estimating leaf chlorophyll content of buffaloberry using normalized difference vegetation index sensors. HortTechnology, 31(3), 297–303. https://doi.org/10.21273/HORTTECH04808-21

Chen, Y., Wang, Q., Wang, Y., Duan, S. B., Xu, M., & Li, Z. L. (2016). A Spectral Signature Shape-Based Algorithm for Landsat Image Classification. ISPRS International Journal of Geo-Information 2016, Vol. 5, Page 154, 5(9), 154. https://doi.org/10.3390/IJGI5090154

Cheng, Z., Aakala, T., & Larjavaara, M. (2023). Elevation, aspect, and slope influence woody vegetation structure and composition but not species richness in a human-influenced landscape in northwestern Yunnan, China. Frontiers in Forests and Global Change, 6, 1187724. https://doi.org/10.3389/FFGC.2023.1187724/BIBTEX

Chirwa, P. W., & Adeyemi, O. (2020a). Deforestation in Africa: Implications on Food and Nutritional Security. https://doi.org/10.1007/978-3-319-95675-6

Chirwa, P. W., & Adeyemi, O. (2020b). Deforestation in Africa: Implications on Food and Nutritional Security. 197–211. https://doi.org/10.1007/978-3-319-95675-6_62

Costa, V. G., & Pedreira, C. E. (2022). Recent advances in decision trees: an updated survey. Artificial Intelligence Review 2022 56:5, 56(5), 4765–4800. https://doi.org/10.1007/S10462-022-10275-5

Croft, H., Arabian, J., Chen, J. M., Shang, J., & Liu, J. (2020). Mapping within-field leaf chlorophyll content in agricultural crops for nitrogen management using Landsat-8 imagery. Precision Agriculture, 21(4), 856–880. https://doi.org/10.1007/S11119-019-09698-Y/FIGURES/10

Daigneault, A., Baker, J. S., Guo, J., Lauri, P., Favero, A., Forsell, N., Johnston, C., Ohrel, S. B., & Sohngen, B. (2022). How the future of the global forest sink depends on timber demand, forest management, and carbon policies. Global Environmental Change, 76, 102582. https://doi.org/10.1016/J.GLOENVCHA.2022.102582

Das, D., & Naskar, R. (2024). Image splicing detection using low-dimensional feature vector of texture features and Haralick features based on Gray Level Co-occurrence Matrix. Signal Processing: Image Communication, 125, 117134. https://doi.org/10.1016/J.IMAGE.2024.117134

Das, T., & Saha, P. (n.d.). The Amazonia and its Biodiversity: Impact of World's Largest Rainforest on Biodiversity. 2(2), 51–57.

de Lima, I. P., Jorge, R. G., & de Lima, J. L. M. P. (2021). Remote Sensing Monitoring of Rice Fields: Towards Assessing Water Saving Irrigation Management Practices. Frontiers in Remote Sensing, 2, 762093. https://doi.org/10.3389/FRSEN.2021.762093/BIBTEX

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., Mcclean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36(1), 27–46. https://doi.org/10.1111/J.1600-0587.2012.07348.X

Doucette, P., Antonisse, J., Braun, A., Lenihan, M., & Brennan, M. (2013). Image georegistration methods: A framework for application guidelines. Proceedings - Applied Imagery Pattern Recognition Workshop. https://doi.org/10.1109/AIPR.2013.6749317

Dugarova, E., & Gülasan, N. (n.d.). Challenges and Opportunities in the Implementation of the Sustainable Development Goals 2 Lead Authors. Retrieved July 25, 2024, from www.unrisd.org

Edwards, D. P., Tobias, J. A., Sheil, D., Meijaard, E., & Laurance, W. F. (2014). Maintaining ecosystem function and services in logged tropical forests. Trends in Ecology & Evolution, 29(9), 511–520. https://doi.org/10.1016/J.TREE.2014.07.003

Eniolorunda, N. B., & Jibrillah, A. (2020). Application of Tasselled-Cap Transformation to Soil Textural Mapping of a Semi-Arid Environment: A Case of Usmanu. Article in Nigerian Journal of Environmental Sciences and Technology. https://doi.org/10.36263/nijest.2020.01.0158

Ezeokoli, O. T., Nuaila, V. N. A., Obieze, C. C., Muetanene, B. A., Fraga, I., Tesinde, M. N., Ndayiragije, A., Coutinho, J., Melo, A. M. P., Adeleke, R. A., Ribeiro-Barros, A. I., & Fangueiro, D. (2021). Assessing the impact of rice cultivation and off-season period on dynamics of soil enzyme activities and bacterial communities in two agro-ecological regions of mozambique. Agronomy, 11(4), 694. https://doi.org/10.3390/AGRONOMY11040694/S1

Fensholt, R., Sandholt, I., & Stisen, S. (2006). Evaluating MODIS, MERIS and vegetation indices using in situ measurements in a semiarid environment. IEEE Transactions on Geoscience and Remote Sensing, 44(7), 1774–1786. https://doi.org/10.1109/TGRS.2006.875940

Gan, C. C., & Learmonth, G. (2015). Comparing entropy with tests for randomness as a measure of complexity in time series. https://arxiv.org/abs/1512.00725v1

Gao, B. C. (1996a). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gao, B. C. (1996b). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gilroy, J. J., Woodcock, P., Edwards, F. A., Wheeler, C., Baptiste, B. L. G., Medina Uribe, C. A., Haugaasen, T., & Edwards, D. P. (2014). Cheap carbon and biodiversity co-benefits from forest regeneration in a hotspot of endemism. Nature Climate Change 2014 4:6, 4(6), 503–507. https://doi.org/10.1038/nclimate2200

Gou, Y., & Gou, Y. (2016). The Role of Maps in Capturing Distal Drivers of Deforestation and Degradation: A Case Study in Central Mozambique. Land Use Competition, 91–109. https://doi.org/10.1007/978-3-319-33628-2_6

Hall-Beyer, M. (2017). (PDF) GLCM Texture: A Tutorial v. 3.0 March 2017. https://www.researchgate.net/publication/315776784_GLCM_Texture_A_Tutorial_v_30_March_2017

Henareh Khalyani, A., Falkowski, M. J., & Mayer, A. L. (2012). Classification of Landsat images based on spectral and topographic variables for land-cover change detection in Zagros forests. International Journal of Remote Sensing, 33(21), 6956–6974. https://doi.org/10.1080/01431161.2012.695095

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., & Campbell, L. B. (2016). Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring.

International Journal of Digital Earth, 9(11), 1035–1054.
https://doi.org/10.1080/17538947.2016.1187673

Horn, D., Demircioğlu, A., Bischl, B., Glasmachers, T., & Weihs, C. (2018). A comparative study on large scale kernelized support vector machines. Advances in Data Analysis and Classification, 12(4), 867–883. https://doi.org/10.1007/S11634-016-0265-7/TABLES/4

Houborg, R., & McCabe, M. F. (2018). A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. Remote Sensing of Environment, 209, 211–226. https://doi.org/10.1016/J.RSE.2018.02.067

Huang, J., Chen, D., & Cosh, M. H. (2009). Sub-pixel reflectance unmixing in estimating vegetation water content and dry biomass of corn and soybeans cropland using normalized difference water index (NDWI) from satellites. International Journal of Remote Sensing, 30(8), 2075–2104. https://doi.org/10.1080/01431160802549245

Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., & Baesens, B. (2011). An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. Decision Support Systems, 51(1), 141–154. https://doi.org/10.1016/J.DSS.2010.12.003

Ivits, E., Lamb, A., Langar, F., Hemphill, S., & Koch, B. (2008). Orthogonal Transformation of Segmented SPOT5 Images. Photogrammetric Engineering and Remote Sensing, 74(11), 1351–1364. https://doi.org/10.14358/PERS.74.11.1351

Jin, Y., Liu, X., Chen, Y., & Liang, X. (2018). Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong. International Journal of Remote Sensing, 39(23), 8703–8723. https://doi.org/10.1080/01431161.2018.1490976

Jordanov, I., Petrov, N., & Petrozziello, A. (2018). Classifiers accuracy improvement based on missing data imputation. JAISCR, 8(1), 31. https://doi.org/10.1515/jaiscr-2018-0002

Kadhim, N., Mourshed, M., & Bray, M. (2016). Advances in remote sensing applications for urban sustainability. Euro-Mediterranean Journal for Environmental Integration, 1(1). https://doi.org/10.1007/S41207-016-0007-4

Kinsey-Henderson, A. E., & Wilkinson, S. N. (2013). Evaluating Shuttle radar and interpolated DEMs for slope gradient and soil erosion estimation in low relief terrain. Environmental Modelling & Software, 40, 128–139. https://doi.org/10.1016/J.ENVSOFT.2012.08.010

Kogan, F. (2020). New Satellite-Based Vegetation Health Technology. 103–134. https://doi.org/10.1007/978-3-030-46020-4_5

Laake, A. (2022). Electromagnetic Spectral Bands Used for Remote Sensing. 21–35. https://doi.org/10.1007/978-3-030-73319-3_2

Landeros, A., & Lange, K. (2023). Algorithms for Sparse Support Vector Machines. Journal of Computational and Graphical Statistics, 32(3), 1097–1108. https://doi.org/10.1080/10618600.2022.2146697

Lieffers, V. J., & Larkin-Lieffers, P. A. (1987). Slope, aspect, and slope position as factors controlling grassland communities in the coulees of the Oldman River, Alberta. Canadian Journal of Botany, 65(7), 1371–1378. https://doi.org/10.1139/B87-189/ASSET/IMAGES/B87-189C6.GIF

Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Change detection techniques. International Journal of Remote Sensing, 25(12), 2365–2401. https://doi.org/10.1080/0143116031000139863

Macarringue, L. S., Bolfe, É. L., Duverger, S. G., Sano, E. E., Caldas, M. M., Ferreira, M. C., Zullo Junior, J., & Matias, L. F. (2023). Land Use and Land Cover Classification in the Northern Region of Mozambique Based on Landsat Time Series and Machine Learning. ISPRS International Journal of Geo-Information 2023, Vol. 12, Page 342, 12(8), 342. https://doi.org/10.3390/IJGI12080342

Maishanu, S. M., Sambo, A. S., & Garba, M. M. (2019). Sustainable bioenergy development in Africa: issues, challenges, and the way forward. Sustainable Bioenergy: Advances and Impacts, 49–87. https://doi.org/10.1016/B978-0-12-817654-2.00003-4

Manjare, B. S., & Singh, V. (2022). Slope analysis from SRTM DEM data: A case study of some part of upper vena river basin,maharashtra, India.

Martino, R. (2024). The relationship between medium-scale farms and deforestation in Sub-Saharan Africa Suggested Citation. Retrieved August 17, 2024, from www.dai.com

Masolele, R. N., De Sy, V., Marcos, D., Verbesselt, J., Gieseke, F., Mulatu, K. A., Moges, Y., Sebrala, H., Martius, C., & Herold, M. (2022). Using high-resolution imagery and deep learning to classify land-

use following deforestation: a case study in Ethiopia. GIScience & Remote Sensing, 59(1), 1446–1472. https://doi.org/10.1080/15481603.2022.2115619

Massingue, A. O. (2019). Ecological Assessment and Biogeography of Coastal Vegetation and Flora in Southern Mozambique.

Matsushita, B., Yang, W., Chen, J., Onda, Y., & Qiu, G. (2007). Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest. Sensors 2007, Vol. 7, Pages 2636-2651, 7(11), 2636–2651. https://doi.org/10.3390/S7112636

Mavume, A. F., Banze, B. E., Macie, O. A., & Queface, A. J. (2021). Analysis of Climate Change Projections for Mozambique under the Representative Concentration Pathways. Atmosphere 2021, Vol. 12, Page 588, 12(5), 588. https://doi.org/10.3390/ATMOS12050588

McDermott, C. L., Cashore, B., & Kanowski, P. (2010). Global Environmental Forest Policies: An International Comparison. Global Environmental Forest Policies: An International Comparison, 1–360. https://doi.org/10.4324/9781849774925/Global-environmental-forest-policies-constance-Mcdermott-Benjamin-Cashore-Peter-Kanowski

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing, 17(7), 1425–1432. https://doi.org/10.1080/01431169608948714

Modu, B., the, B. H.-T. I. A. of, & 2014, undefined. (2014). sensed observations of Congo basin of East African high Land to drain water using gravity for sustainable management of low laying Chad basin of Central Africa. Isprs-Archives.Copernicus.Org, 40(1), 279–286. https://doi.org/10.5194/isprsarchives-XL-1-279-2014

Mozambique - Philip Briggs - Google Books. (n.d.). Retrieved July 25, 2024, from https://books.google.nl/books?hl=en&lr=&id=MSNQBAAAQBAJ&oi=fnd&pg=PP1&dq=Briggs,+P.+(2014).+Mozambique.+Bradt+Travel+Guides.+&ots=8zlL_1qCCc&sig=SjPTIrZBc0ofmofCyKObp7D1ngU&redir_esc=y#v=onepage&q=Briggs%2C%20P.%20(2014).%20Mozambique.%20Bradt%20Travel%20Guides.&f=false

Mucova, S. A. R., Filho, W. L., Azeiteiro, U. M., & Pereira, M. J. (2018a). Assessment of land use and land cover changes from 1979 to 2017 and biodiversity & land management approach in Quirimbas National Park, Northern Mozambique, Africa. Global Ecology and Conservation, 16. https://doi.org/10.1016/J.GECCO.2018.E00447

Mucova, S. A. R., Filho, W. L., Azeiteiro, U. M., & Pereira, M. J. (2018b). Assessment of land use and land cover changes from 1979 to 2017 and biodiversity & land management approach in Quirimbas National Park, Northern Mozambique, Africa. Global Ecology and Conservation, 16. https://doi.org/10.1016/J.GECCO.2018.E00447

Mucova, S. A. R., Filho, W. L., Azeiteiro, U. M., & Pereira, M. J. (2018c). Assessment of land use and land cover changes from 1979 to 2017 and biodiversity & land management approach in Quirimbas National Park, Northern Mozambique, Africa. Global Ecology and Conservation, 16, e00447. https://doi.org/10.1016/J.GECCO.2018.E00447

Naqvi, H. R., Mallick, J., Devi, L. M., & Siddiqui, M. A. (2013). Multi-temporal annual soil loss risk mapping employing Revised Universal Soil Loss Equation (RUSLE) model in Nun Nadi Watershed, Uttrakhand (India). Arabian Journal of Geosciences, 6(10), 4045–4056. https://doi.org/10.1007/S12517-012-0661-Z

Nguyen, C. T., Chidthaisong, A., Diem, P. K., & Huo, L. Z. (2021). A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8. Land 2021, Vol. 10, Page 231, 10(3), 231. https://doi.org/10.3390/LAND10030231

Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Van Le, H., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. Mathematical Problems in Engineering, 2021(1), 4832864. https://doi.org/10.1155/2021/4832864

Nogueira Lisboa, S., Grinand, C., Betbeder, J., Montfort, F., & Blanc, L. (2024a). Disentangling the drivers of deforestation and forest degradation in the Miombo landscape: A case study from Mozambique. International Journal of Applied Earth Observation and Geoinformation, 130. https://doi.org/10.1016/J.JAG.2024.103904

Nogueira Lisboa, S., Grinand, C., Betbeder, J., Montfort, F., & Blanc, L. (2024b). Disentangling the drivers of deforestation and forest degradation in the Miombo landscape: A case study from Mozambique. International Journal of Applied Earth Observation and Geoinformation, 130, 103904. https://doi.org/10.1016/J.JAG.2024.103904

Nugroho, K. H., Wahyudi, R., Baihaqi, W. M., & Nurfaizah. (2021). NDVI Method for Near-Infrared Image Analysis for Health Monitoring of Ornamental Plants Spathiphyllum Wallisii Mauna LOA. 2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 61–65. https://doi.org/10.1109/ICITISEE53823.2021.9655944

Olagunju, T. E. (n.d.). Impacts of Human-induced Deforestation, Forest Degradation and Fragmentation on Food Security. Retrieved August 17, 2024, from http://www.sciencepub.net/newyork

Pamuji, R., Mahardika, A. I., Wiranda, N., Saputra, N. A. B., Adini, M. H., & Pramatasari, D. (2023). Utilizing Electromagnetic Radiation in Remote Sensing for Vegetation Health Analysis Using NDVI Approach with Sentinel-2 Imagery. Kasuari Physics Education Journal (KPEJ), 6(2), 127–135. https://doi.org/10.37891/KPEJ.V6I2.486

Panuju, D. R., Paull, D. J., & Griffin, A. L. (2020). Change Detection Techniques Based on Multispectral Images for Investigating Land Cover Dynamics. Remote Sensing 2020, Vol. 12, Page 1781, 12(11), 1781. https://doi.org/10.3390/RS12111781

Park, Y., & Guldmann, J. M. (2020). Measuring continuous landscape patterns with Gray-Level Co-Occurrence Matrix (GLCM) indices: An alternative to patch metrics? Ecological Indicators, 109, 105802. https://doi.org/10.1016/J.ECOLIND.2019.105802

Parmar, A., Katariya, R., & Patel, V. (2019). A Review on Random Forest: An Ensemble Classifier. Lecture Notes on Data Engineering and Communications Technologies, 26, 758–763. https://doi.org/10.1007/978-3-030-03146-6_86/FIGURES/1

Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. Remote Sensing of Environment, 187, 156–168. https://doi.org/10.1016/J.RSE.2016.10.010

Pennington, R. T., Lehmann, C. E. R., & Rowland, L. M. (2018). Tropical savannas and dry forests. Current Biology, 28(9), R541–R545. https://doi.org/10.1016/J.CUB.2018.03.014

Peter, B. G., Messina, J. P., Carroll, J. W., & Chikowo, R. (2021). A case for green-based vegetation indices: plot-scale sUAS imagery related to crop chlorophyll content on smallholder maize farms in Malawi. Remote Sensing Letters, 12(8), 778–787. https://doi.org/10.1080/2150704X.2021.1938733

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. Remote Sensing 2020, Vol. 12, Page 2291, 12(14), 2291. https://doi.org/10.3390/RS12142291

Portillo-Quintero, C. A., & Sánchez-Azofeifa, G. A. (2010). Extent and conservation of tropical dry forests in the Americas. Biological Conservation, 143(1), 144–155. https://doi.org/10.1016/J.BIOCON.2009.09.020

Powers, D. M. W., & Ailab. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. https://arxiv.org/abs/2010.16061v1

Priyanka, & Kumar, D. (2020). Decision tree classifier: A detailed survey. International Journal of Information and Decision Sciences, 12(3), 246–269. https://doi.org/10.1504/IJIDS.2020.108141

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. Remote Sensing of Environment, 48(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Radeloff, V. C., Roy, D. P., Wulder, M. A., Anderson, M., Cook, B., Crawford, C. J., Friedl, M., Gao, F., Gorelick, N., Hansen, M., Healey, S., Hostert, P., Hulley, G., Huntington, J. L., Johnson, D. M., Neigh, C., Lyapustin, A., Lymburner, L., Pahlevan, N., … Zhu, Z. (2024). Need and vision for global medium-resolution Landsat and Sentinel-2 data products. Remote Sensing of Environment, 300, 113918. https://doi.org/10.1016/J.RSE.2023.113918

Raven, P. H., Gereau, R. E., Phillipson, P. B., Chatelain, C., Jenkins, C. N., & Ulloa, C. U. (2020). The distribution of biodiversity richness in the tropics. Science Advances, 6(37). https://doi.org/10.1126/SCIADV.ABC6228/ASSET/203E9C1F-5A74-40C6-BA2E-1CDB8ED3C05C/ASSETS/GRAPHIC/ABC6228-F2.JPEG

Roberts, P., Boivin, N., Lee-Thorp, J., Petraglia, M., & Stock, J. (2016). Tropical forests and the genus Homo. Evolutionary Anthropology: Issues, News, and Reviews, 25(6), 306–317. https://doi.org/10.1002/EVAN.21508

Roy, D. P., Wulder, M. A., Loveland, T. R., C.E., W., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R., Sheng, Y., Vermote, E. F., Belward, A. S., Bindschadler, R., Cohen, W. B., Gao, F., … Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. Remote Sensing of Environment, 145, 154–172. https://doi.org/10.1016/J.RSE.2014.02.001

Saberi, A., Kabolizadeh, M., Rangzan, K., & Abrehdary, M. (2023). Accuracy assessment and improvement of SRTM, ASTER, FABDEM, and MERIT DEMs by polynomial and optimization algorithm: A case study (Khuzestan Province, Iran). Open Geosciences, 15(1). https://doi.org/10.1515/GEO-2022-0455/MACHINEREADABLECITATION/RIS

Salite, D. (2019). Traditional prediction of drought under weather and climate uncertainty: analyzing the challenges and opportunities for small-scale farmers in Gaza province, southern region of Mozambique. Natural Hazards, 96(3), 1289–1309. https://doi.org/10.1007/S11069-019-03613-4/FIGURES/1

Sedano, F., Silva, J. A., Machoco, R., Meque, C. H., Sitoe, A., Ribeiro, N., Anderson, K., Ombe, Z. A., Baule, S. H., & Tucker, C. J. (2016). The impact of charcoal production on forest degradation: a case study in Tete, Mozambique. Environmental Research Letters, 11(9), 094020. https://doi.org/10.1088/1748-9326/11/9/094020

Sihag, M. S., & Sihag, M. S. (2021). Normalized Difference Vegetation Index (NDVI) based vegetation change detection in Som River Catchment. Research Review International Journal of Multidisciplinary, 6(6). https://doi.org/10.31305/RRIJM.2021.V06.I06.004

Singh, S. (n.d.). Understanding the role of slope aspect in shaping the vegetation attributes and soil properties in Montane ecosystems. Retrieved August 15, 2024, from www.tropecol.com

Sothe, C., de Almeida, C. M., Liesenberg, V., & Schimalski, M. B. (2017). Evaluating Sentinel-2 and Landsat-8 Data to Map Sucessional Forest Stages in a Subtropical Forest in Southern Brazil. Remote Sensing 2017, Vol. 9, Page 838, 9(8), 838. https://doi.org/10.3390/RS9080838

Stage, A. R., & Salas, C. (2007). Interactions of Elevation, Aspect, and Slope in Models of Forest Species Composition and Productivity. Forest Science, 53(4), 486–492. https://doi.org/10.1093/FORESTSCIENCE/53.4.486

Stamford, J. D., Vialet-Chabrand, S., Cameron, I., & Lawson, T. (2023). Development of an accurate low cost NDVI imaging system for assessing plant health. Plant Methods, 19(1), 1–19. https://doi.org/10.1186/S13007-023-00981-8/FIGURES/10

Stanturf, J. A., & Mansourian, S. (2020). Forest landscape restoration: state of play. Royal Society Open Science, 7(12). https://doi.org/10.1098/RSOS.201218

Strassburg, B. B. N., Beyer, H. L., Crouzeilles, R., Iribarrem, A., Barros, F., de Siqueira, M. F., Sánchez-Tapia, A., Balmford, A., Sansevero, J. B. B., Brancalion, P. H. S., Broadbent, E. N., Chazdon, R. L., Filho, A. O., Gardner, T. A., Gordon, A., Latawiec, A., Loyola, R., Metzger, J. P., Mills, M., … Uriarte, M. (2020). Author Correction: Strategic approaches to restoring ecosystems can triple conservation gains and halve costs (Nature Ecology & Evolution, (2019), 3, 1, (62-70), 10.1038/s41559-018-0743-8). Nature Ecology and Evolution, 4(5), 765. https://doi.org/10.1038/S41559-020-1211-9

Sun, F., Sun, W., Chen, J., & Gong, P. (2012). Comparison and improvement of methods for identifying waterbodies in remotely sensed imagery. International Journal of Remote Sensing, 33(21), 6854–6875. https://doi.org/10.1080/01431161.2012.692829

Sutradhar, P. (n.d.). Landsat 7 ETM+ Enhancing Earth Observation And Environmental Monitoring '. Retrieved July 25, 2024, from https://www.researchgate.net/publication/375592356

Tamiminia, H., Salehi, B., Mahdianpari, M., Beier, C. M., Johnson, L., Phoenix, D. B., & Mahoney, M. (2022). Decision tree-based machine learning models for above-ground biomass estimation using multi-source remote sensing data and object-based image analysis. Geocarto International, 37(26), 12763–12791. https://doi.org/10.1080/10106049.2022.2071475

Taye, G., Poesen, J., Wesemael, B. Van, Vanmaercke, M., Teka, D., Deckers, J., Goosse, T., Maetens, W., Nyssen, J., Hallet, V., & Haregeweyn, N. (2013). Effects of land use, slope gradient, and soil and

water conservation structures on runoff and soil loss in semi-arid Northern Ethiopia. Physical Geography, 34(3), 236–259. https://doi.org/10.1080/02723646.2013.832098

Temudo, M. P., & Silva, J. M. N. (2012a). Agriculture and forest cover changes in post-war Mozambique. Journal of Land Use Science, 7(4), 425–442. https://doi.org/10.1080/1747423X.2011.595834

Temudo, M. P., & Silva, J. M. N. (2012b). Agriculture and forest cover changes in post-war Mozambique. Journal of Land Use Science, 7(4), 425–442. https://doi.org/10.1080/1747423X.2011.595834

Thomas, T., P. Vijayaraghavan, A., & Emmanuel, S. (2020a). Applications of Decision Trees. Machine Learning Approaches in Cyber Security Analytics, 157–184. https://doi.org/10.1007/978-981-15-1706-8_9

Thomas, T., P. Vijayaraghavan, A., & Emmanuel, S. (2020b). Applications of Decision Trees. Machine Learning Approaches in Cyber Security Analytics, 157–184. https://doi.org/10.1007/978-981-15-1706-8_9

Tokura, W., Matimele, H., Smit, J., & Hoffman, M. T. (2020). Long-term changes in forest cover in a global biodiversity hotspot in southern Mozambique. Bothalia, African Biodiversity & Conservation, 50(1). https://doi.org/10.38201/BTHA.ABC.V50.I1.1

Tokura, W., Matimele, H., Smit, J., Timm Hoffman, M., & Timm Hoffman, M. (2020). Long-term changes in forest cover in a global biodiversity hotspot in southern Mozambique. Bothalia, African Biodiversity & Conservation, 50(1). https://doi.org/10.38201/BTHA.ABC.V50.I1.1

Trigg, S. N., Curran, L. M., & Mcdonald, A. K. (2006). Utility of Landsat 7 satellite data for continued monitoring of forest cover change in protected areas in Southeast Asia. Singapore Journal of Tropical Geography, 27(1), 49–66. https://doi.org/10.1111/J.1467-9493.2006.00239.X

Twumasi, Y. A., Merem, E. C., Namwamba, J. B., Asare-Ansah, A. B., Annan, J. B., Ning, Z. H., Armah, R. N. D., Apraku, C. Y., Yeboah, H. B., Atayi, J., Anokye, M., Frimpong, D. B., Okwemba, R., Mwakimi, O. S., Oppong, J., Petja, B. M., Mjema, J., Loh, P. M., Kangwana, L. A., … McClendon-Peralta, J. (2022). Flood Mapping in Mozambique Using Copernicus Sentinel-2 Satellite Data. Advances in Remote Sensing, 11(03), 80–105. https://doi.org/10.4236/ARS.2022.113006

Valbuena, R., Mauro, F., Rodriguez-Solano, R., & Manzanera, J. A. (2010). Exactitud y precisión de receptores GPS bajo cubiertas forestales en ambientes montañosos. Spanish Journal of Agricultural Research, 8(4), 1047–1057. https://doi.org/10.5424/SJAR/2010084-1242

Waring, B., Neumann, M., Prentice, I. C., Adams, M., Smith, P., & Siegert, M. (2020). Forests and Decarbonization – Roles of Natural and Planted Forests. Frontiers in Forests and Global Change, 3. https://doi.org/10.3389/FFGC.2020.00058

Wei, J., Peng, Y., Mahmood, R., Sun, L., & Guo, J. (2019). Intercomparison in spatial distributions and temporal trends derived from multi-source satellite aerosol products. Atmospheric Chemistry and Physics, 19(10), 7183–7207. https://doi.org/10.5194/ACP-19-7183-2019

Williams, P. H., Burgess, N. D., & Rahbek, C. (2000). Flagship species, ecological complementarity and conserving the diversity of mammals and birds in sub-Saharan Africa. Animal Conservation Forum, 3(3), 249–260. https://doi.org/10.1111/J.1469-1795.2000.TB00110.X

Woollen, E., Ryan, C. M., Baumert, S., Vollmer, F., Grundy, I., Fisher, J., Fernando, J., Luz, A., Ribeiro, N., & Lisboa, S. N. (2016). Charcoal production in the Mopane woodlands of Mozambique: what are the trade-offs with other ecosystem services? Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1703). https://doi.org/10.1098/RSTB.2015.0315

Wulder, M. A., Roy, D. P., Radeloff, V. C., Loveland, T. R., Anderson, M. C., Johnson, D. M., Healey, S., Zhu, Z., Scambos, T. A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C. J., Masek, J. G., Hermosilla, T., White, J. C., Belward, A. S., Schaaf, C., Woodcock, C. E., … Cook, B. D. (2022). Fifty years of Landsat science and impacts. Remote Sensing of Environment, 280, 113195. https://doi.org/10.1016/J.RSE.2022.113195

Xie, C., Wang, J., Haase, D., Wellmann, T., & Lausch, A. (2023). Measuring spatio-temporal heterogeneity and interior characteristics of green spaces in urban neighborhoods: A new approach using gray level co-occurrence matrix. Science of The Total Environment, 855, 158608. https://doi.org/10.1016/J.SCITOTENV.2022.158608

Yesilnacar, E., & Süzen, M. L. (2006). A land-cover classification for landslide susceptibility mapping by using feature components. International Journal of Remote Sensing, 27(2), 253–275. https://doi.org/10.1080/0143116050030042

Yosef, B. A. (2014). The Role of Forest and Soil Carbon Sequestrations on Climate Change Mitigation. https://www.researchgate.net/publication/324844908

Zhan, X., Sohlberg, R. A., Townshend, J. R. G., DiMiceli, C., Carroll, M. L., Eastman, J. C., Hansen, M. C., & DeFries, R. S. (2002). Detection of land cover changes using MODIS 250 m data. Remote Sensing of Environment, 83(1–2), 336–350. https://doi.org/10.1016/S0034-4257(02)00081-0

Zhen, Z., Chen, S., Yin, T., & Gastellu-Etchegorry, J. P. (2023). Globally quantitative analysis of the impact of atmosphere and spectral response function on 2-band enhanced vegetation index (EVI2) over Sentinel-2 and Landsat-8. ISPRS Journal of Photogrammetry and Remote Sensing, 205, 206–226. https://doi.org/10.1016/J.ISPRSJPRS.2023.09.024

Zheng, Y., Tang, L., & Wang, H. (2021). An improved approach for monitoring urban built-up areas by combining NPP-VIIRS nighttime light, NDVI, NDWI, and NDBI. Journal of Cleaner Production, 328. https://doi.org/10.1016/J.JCLEPRO.2021.129488

Zhou, Y., Dong, J., Liu, J., Metternicht, G., Shen, W., You, N., Zhao, G., & Xiao, X. (2019). Are There Sufficient Landsat Observations for Retrospective and Continuous Monitoring of Land Cover Changes in China? Remote Sensing 2019, Vol. 11, Page 1808, 11(15), 1808. https://doi.org/10.3390/RS11151808

APPENDIXES

Variability analysis

The Table 1 provides detail insight into variables according to their significant differences in relation to the P-value, F-value and F-critical

Table 1: Analysis of variance among variables

EVI

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.614298	3	0.871433	267.1363	9.26E-95	2.627441
Within Groups	1.291803	396	0.003262			
Total	3.906101	399				

Savi

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.614298	3	0.871433	267.1363	9.26E-95	2.627441
Within Groups	1.291803	396	0.003262			
Total	3.906101	399				

Wetness

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.145679	3	0.715226	420.3086	1.2E-122	2.627441
Within Groups	0.673861	396	0.001702			
Total	2.819541	399				

Greenness

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.632818	3	0.210939	263.8601	4.73E-94	2.627441
Within Groups	0.316577	396	0.000799			
Total	0.949395	399				

SWIR1

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	3.195783	3	1.065261	341.987	1.6E-109	2.627441
Within Groups	1.233507	396	0.003115			
Total	4.42929	399				

NDWI

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.939629	3	0.979876	235.1226	1.4E-87	2.627441
Within Groups	1.650335	396	0.004168			
Total	4.589964	399				

NDVI

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	10.06356	3	3.35452	460.794	9.7E-129	2.627441
Within Groups	2.882828	396	0.00728			
Total	12.94639	399				

GNDVI

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.939629	3	0.979876	235.1226	1.4E-87	2.627441
Within Groups	1.650335	396	0.004168			
Total	4.589964	399				

Brightness

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	5.27085	3	1.75695	222.3969	1.5E-84	2.627441
Within Groups	3.128425	396	0.0079			
Total	8.399276	399				

Entropy

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	12.69696	3	4.232319	29.88421	1.95E-17	2.627441
Within Groups	56.08306	396	0.141624			
Total	68.78002	399				

SWIR2

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	3.489544	3	1.163181	354.7006	8.5E-112	2.627441
Within Groups	1.298616	396	0.003279			
Total	4.78816	399				

NIR

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.32193	3	0.10731	45.72288	2.16E-25	2.627441
Within Groups	0.929397	396	0.002347			
Total	1.251327	399				

Table 2: Tukey's HSD test for evaluating variability among landcover classes

Index	Comparison	diff	lwr	upr	p adj
	Cropland-Built up	0.079002	0.027822	0.130183	0.0005
	Forest-Built up	0.322474	0.271294	0.373655	0.0000
	Other-Built up	-0.00586	-0.05704	0.045318	0.9910
	Forest-Cropland	0.243472	0.192421	0.294524	0.0000
	Other-Cropland	-0.08486	-0.13592	-0.03381	0.0001
Homogeneity	Other-Forest	-0.32834	-0.37939	-0.27728	0.0000
	Cropland-Built up	0.040057	0.027255	0.052858	0.0000
	Forest-Built up	0.092688	0.079887	0.10549	0.0000
DVI	Other-Built up	-0.00028	-0.01308	0.012525	0.9999

	Forest-Cropland	0.052631	0.039862	0.0654	0.0000
	Other-Cropland	-0.04033	-0.0531	-0.02756	0.0000
	Other-Forest	-0.09296	-0.10573	-0.0802	0.0000
	Cropland-Built up	-188702	-325913	-51490.8	0.0024
	Forest-Built up	-269645	-406857	-132434	0.0000
	Other-Built up	250490.1	113278.8	387701.4	0.0000
	Forest-Cropland	-80943.2	-217809	55922.94	0.4230
	Other-Cropland	439192.1	302326	576058.3	0.0000
Contrast	Other-Forest	520135.3	383269.2	657001.5	0.0000
	Cropland-Built up	0.441717	-0.53035	1.413785	0.6447
	Forest-Built up	-0.00828	-0.98035	0.963785	1.0000
	Other-Built up	1.161717	0.189649	2.133785	0.0117
	Forest-Cropland	-0.45	-1.41962	0.519622	0.6289
	Other-Cropland	0.72	-0.24962	1.689622	0.2230
Slope	Other-Forest	1.17	0.200378	2.139622	0.0107
	Cropland-Built up	-16.3332	-26.3855	-6.28094	0.0002
	Forest-Built up	36.15677	26.10447	46.20906	0.0000
	Other-Built up	13.63677	3.584475	23.68906	0.0029
	Forest-Cropland	52.49	42.463	62.517	0.0000
	Other-Cropland	29.97	19.943	39.997	0.0000
Elevation	Other-Forest	-22.52	-32.547	-12.493	0.0000
	Cropland-Built up	-0.02421	-0.03048	-0.01795	0.0000
	Forest-Built up	-0.03879	-0.04505	-0.03253	0.0000
Blue	Other-Built up	0.015739	0.009476	0.022002	0.0000

	Forest-Cropland	-0.01458	-0.02082	-0.00833	0.0000
	Other-Cropland	0.039952	0.033705	0.046199	0.0000
	Other-Forest	0.054528	0.048281	0.060775	0.0000
	Cropland-Built up	-0.02608	-0.0362	-0.01595	0.0000
	Forest-Built up	-0.04568	-0.05581	-0.03556	0.0000
	Other-Built up	0.037183	0.027058	0.047308	0.0000
	Forest-Cropland	-0.0196	-0.0297	-0.0095	0.0000
	Other-Cropland	0.063262	0.053162	0.073362	0.0000
Green	Other-Forest	0.082867	0.072767	0.092966	0.0000
	Cropland-Built up	-0.04653	-0.06109	-0.03196	0.0000
	Forest-Built up	-0.07634	-0.0909	-0.06177	0.0000
	Other-Built up	0.066165	0.0516	0.080731	0.0000
	Forest-Cropland	-0.02981	-0.04434	-0.01528	0.0000
	Other-Cropland	0.112694	0.098165	0.127223	0.0000
Red	Other-Forest	0.142502	0.127973	0.157031	0.0000
	Cropland-Built up	-0.00647	-0.02421	0.01127	0.7827
	Forest-Built up	0.016351	-0.00139	0.034093	0.0831
	Other-Built up	0.065889	0.048147	0.083631	0.0000
	Forest-Cropland	0.022823	0.005126	0.040521	0.0053
	Other-Cropland	0.072361	0.054664	0.090059	0.0000
NIR	Other-Forest	0.049538	0.03184	0.067235	0.0000
	Cropland-Built up	-0.079	-0.10264	-0.05537	0.0000
	Forest-Built up	-0.20843	-0.23207	-0.1848	0.0000
NDWI	Other-Built up	0.00218	-0.02146	0.025815	0.9953

	Forest-Cropland	-0.12943	-0.15301	-0.10586	0.0000
	Other-Cropland	0.081181	0.057605	0.104757	0.0000
	Other-Forest	0.210614	0.187038	0.23419	0.0000
	Cropland-Built up	0.079001	0.055366	0.102637	0.0000
	Forest-Built up	0.208435	0.184799	0.23207	0.0000
	Other-Built up	-0.00218	-0.02582	0.021456	0.9953
	Forest-Cropland	0.129433	0.105857	0.15301	0.0000
	Other-Cropland	-0.08118	-0.10476	-0.0576	0.0000
GNDVI	Other-Forest	-0.21061	-0.23419	-0.18704	0.0000
	Cropland-Built up	-0.07708	-0.09752	-0.05664	0.0000
	Forest-Built up	-0.14051	-0.16095	-0.12007	0.0000
	Other-Built up	0.098943	0.078503	0.119384	0.0000
	Forest-Cropland	-0.06343	-0.08382	-0.04304	0.0000
	Other-Cropland	0.176021	0.155632	0.19641	0.0000
SWIR1	Other-Forest	0.239455	0.219065	0.259844	0.0000
	Cropland-Built up	0.160867	0.129639	0.192096	0.0000
	Forest-Built up	0.346006	0.314778	0.377235	0.0000
	Other-Built up	-0.06353	-0.09476	-0.03231	0.0000
	Forest-Cropland	0.185139	0.153989	0.216289	0.0000
	Other-Cropland	-0.2244	-0.25555	-0.19325	0.0000
NDVI	Other-Forest	-0.40954	-0.44069	-0.37839	0.0000
	Cropland-Built up	-0.09868	-0.11966	-0.07771	0.0000
	Forest-Built up	-0.15814	-0.17912	-0.13717	0.0000
SWIR2	Other-Built up	0.086164	0.065191	0.107138	0.0000

	Forest-Cropland	-0.05946	-0.08038	-0.03854	0.0000
	Other-Cropland	0.184846	0.163926	0.205767	0.0000
	Other-Forest	0.244309	0.223388	0.265229	0.0000
	Cropland-Built up	-0.05292	-0.19075	0.084901	0.7548
	Forest-Built up	-0.43339	-0.57122	-0.29557	0.0000
	Other-Built up	-0.01895	-0.15678	0.118871	0.9847
	Forest-Cropland	-0.38047	-0.51795	-0.24299	0.0000
	Other-Cropland	0.033969	-0.10351	0.171446	0.9198
Entropy	Other-Forest	0.41444	0.276964	0.551917	0.0000
	Cropland-Built up	-0.09787	-0.13043	-0.06532	0.0000
	Forest-Built up	-0.15244	-0.18499	-0.11988	0.0000
	Other-Built up	0.149727	0.117174	0.18228	0.0000
	Forest-Cropland	-0.05456	-0.08703	-0.02209	0.0001
	Other-Cropland	0.247601	0.21513	0.280072	0.0000
Brightness	Other-Forest	0.302163	0.269692	0.334634	0.0000
	Cropland-Built up	0.044359	0.034009	0.05471	0.0000
	Forest-Built up	0.091223	0.080872	0.101573	0.0000
	Other-Built up	-0.00841	-0.01876	0.001939	0.1559
	Forest-Cropland	0.046864	0.036539	0.057188	0.0000
	Other-Cropland	-0.05277	-0.0631	-0.04245	0.0000
Greenness	Other-Forest	-0.09963	-0.10996	-0.08931	0.0000
	Cropland-Built up	0.073561	0.058453	0.088669	0.0000
	Forest-Built up	0.137735	0.122627	0.152843	0.0000
Wetness	Other-Built up	-0.05582	-0.07092	-0.04071	0.0000

	Forest-Cropland	0.064174	0.049104	0.079244	0.0000
	Other-Cropland	-0.12938	-0.14445	-0.11431	0.0000
	Other-Forest	-0.19355	-0.20862	-0.17848	0.0000
Savi	Cropland-Built up	0.081777	0.06087	0.102683	0.0000
	Forest-Built up	0.183418	0.162512	0.204325	0.0000
	Other-Built up	-0.02242	-0.04332	-0.00151	0.0301
	Forest-Cropland	0.101642	0.080788	0.122496	0.0000
	Other-Cropland	-0.10419	-0.12505	-0.08334	0.0000
	Other-Forest	-0.20583	-0.22669	-0.18498	0.0000
EVI	Cropland-Built up	0.138075	0.102576	0.173575	0.0000
	Forest-Built up	0.319137	0.283638	0.354637	0.0000
	Other-Built up	-0.06303	-0.09853	-0.02753	0.0000
	Forest-Cropland	0.181062	0.145652	0.216472	0.0000
	Other-Cropland	-0.20111	-0.23652	-0.16569	0.0000
	Other-Forest	-0.38217	-0.41758	-0.34676	0.0000

Boxplot showing variability among landcover classes

This section illustrates the use of boxplots to visualize the distribution of spectral bands, vegetation indices, topographic, texture features, and tasselled cap features across landcover classes: Forest, Built up, Cropland, and Other using Landsat data. This aided in evaluating if there were variability among classes by visually inspecting the means as well as based on statistical findings as shown in section Table 2.
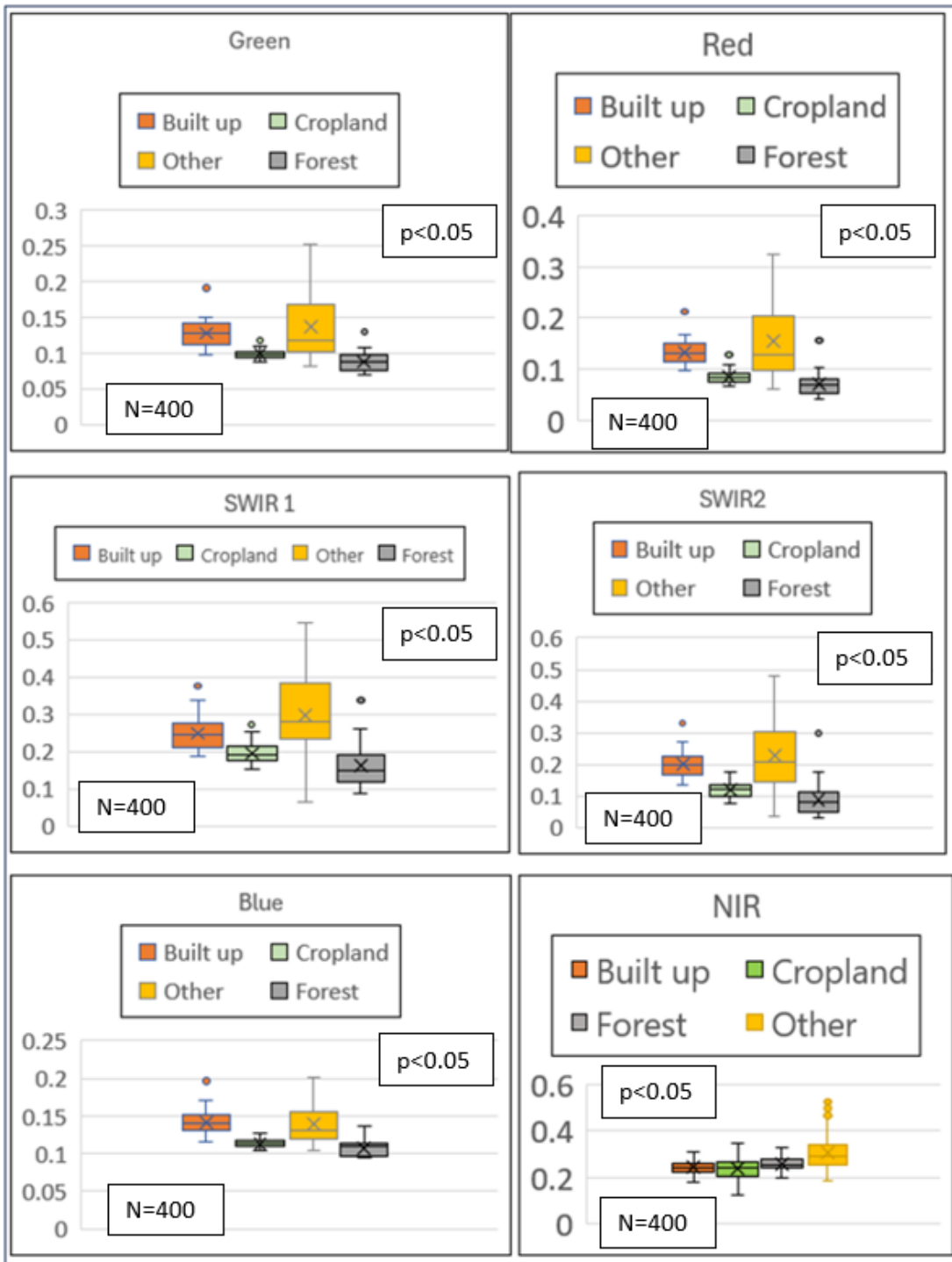
Figure 1: Boxplot showing the distribution of spectral bands (Green, Red, Blue, SWIR1, SWIR2, NIR) across landcover classes: Forest, Cropland, Built up and Other using Landsat data.
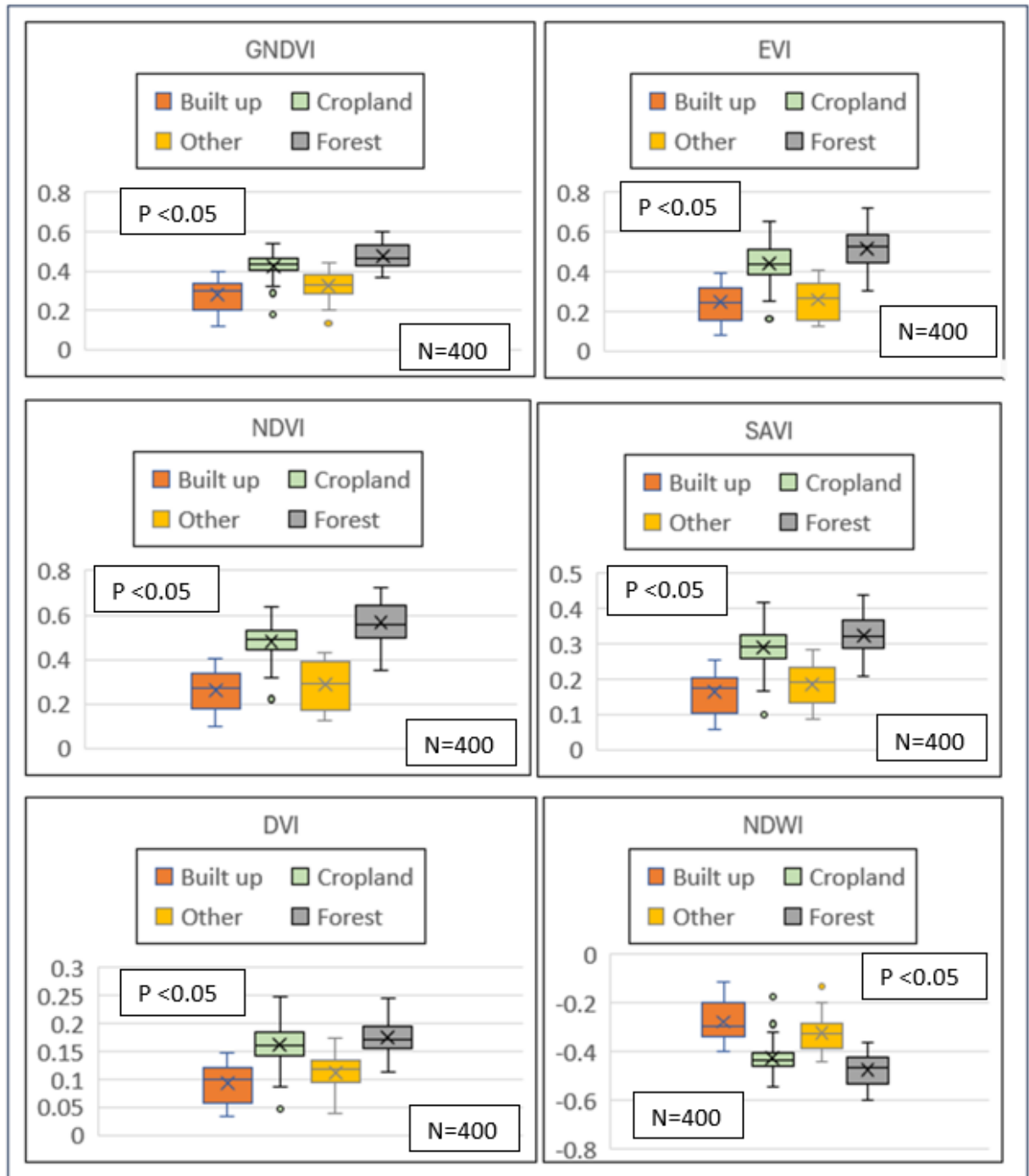
Figure 2: Boxplot showing the distribution of vegetation indices (NDVI, EVI, GNDVI, SAVI, NDWI, DVI) across landcover classes: Forest, Built up, Other and Cropland using Landsat data.
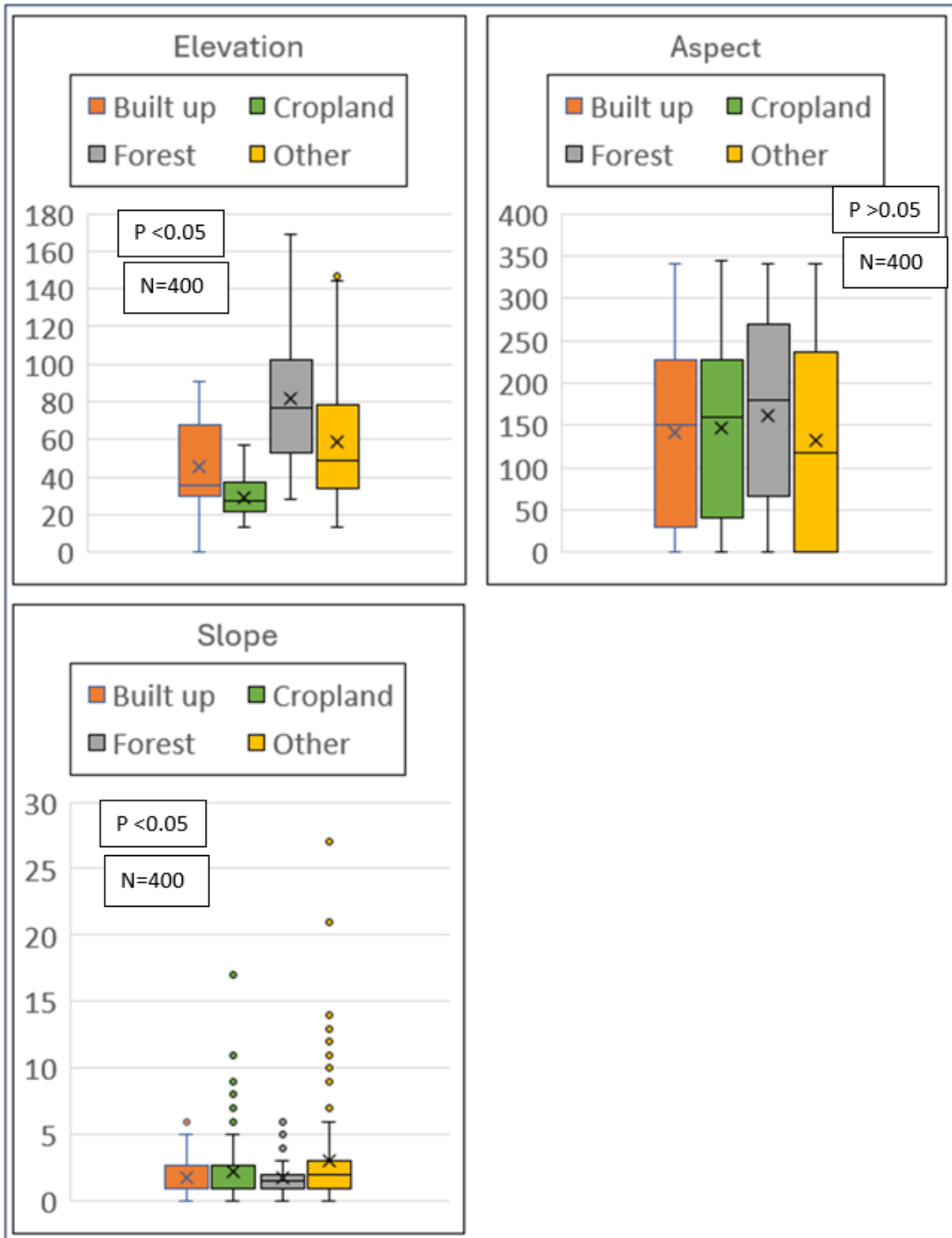
Figure 3: Boxplot showing the distribution of topographic features (Elevation, Slope, Aspect) across landcover classes: Forest, and Cropland, Built up and Other using Landsat data.
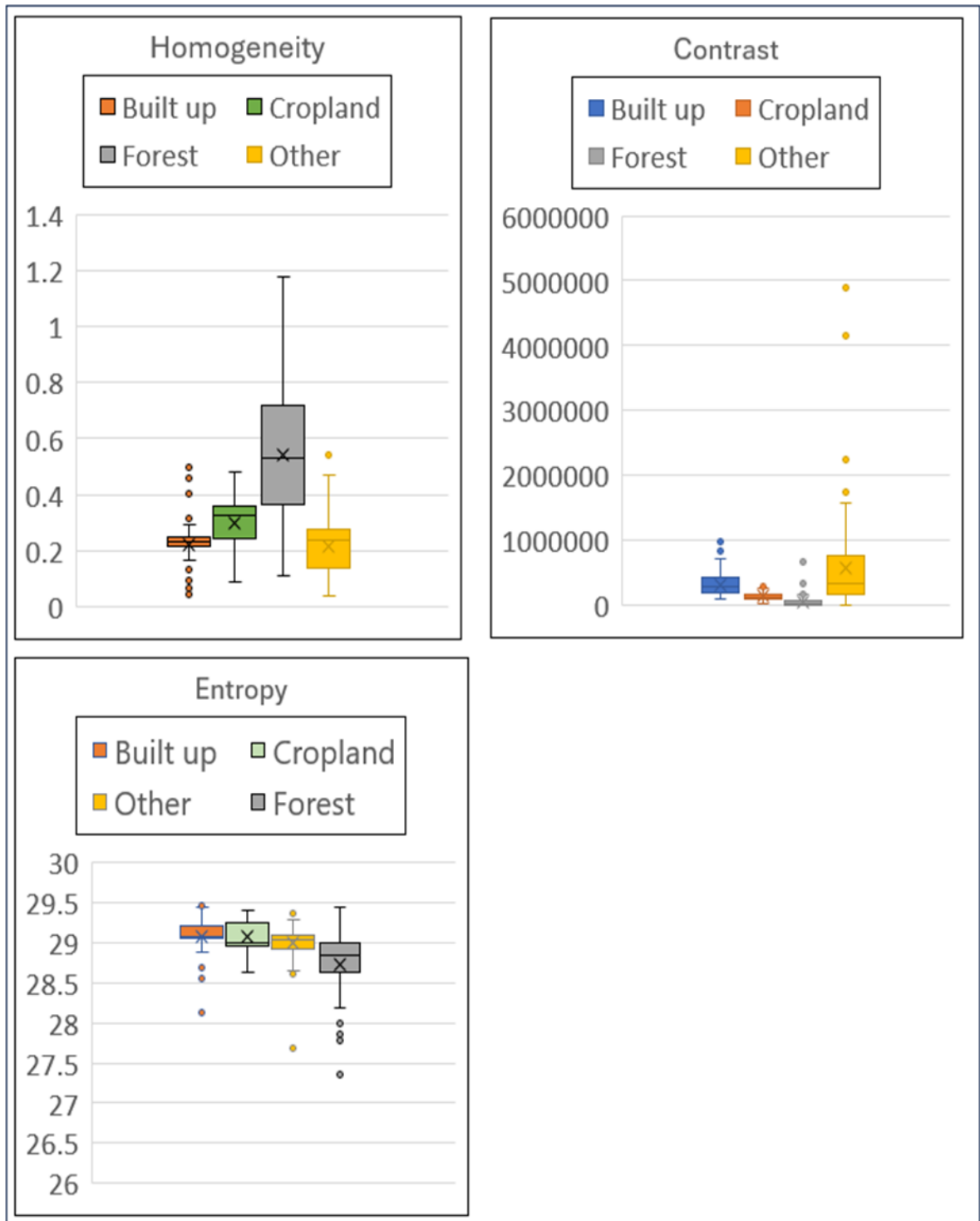
Figure 4: Boxplot showing the distribution of texture features (Homogeneity, Contrast, Entropy) across landcover classes: Forest, Cropland, Built up and Other using Landsat data.
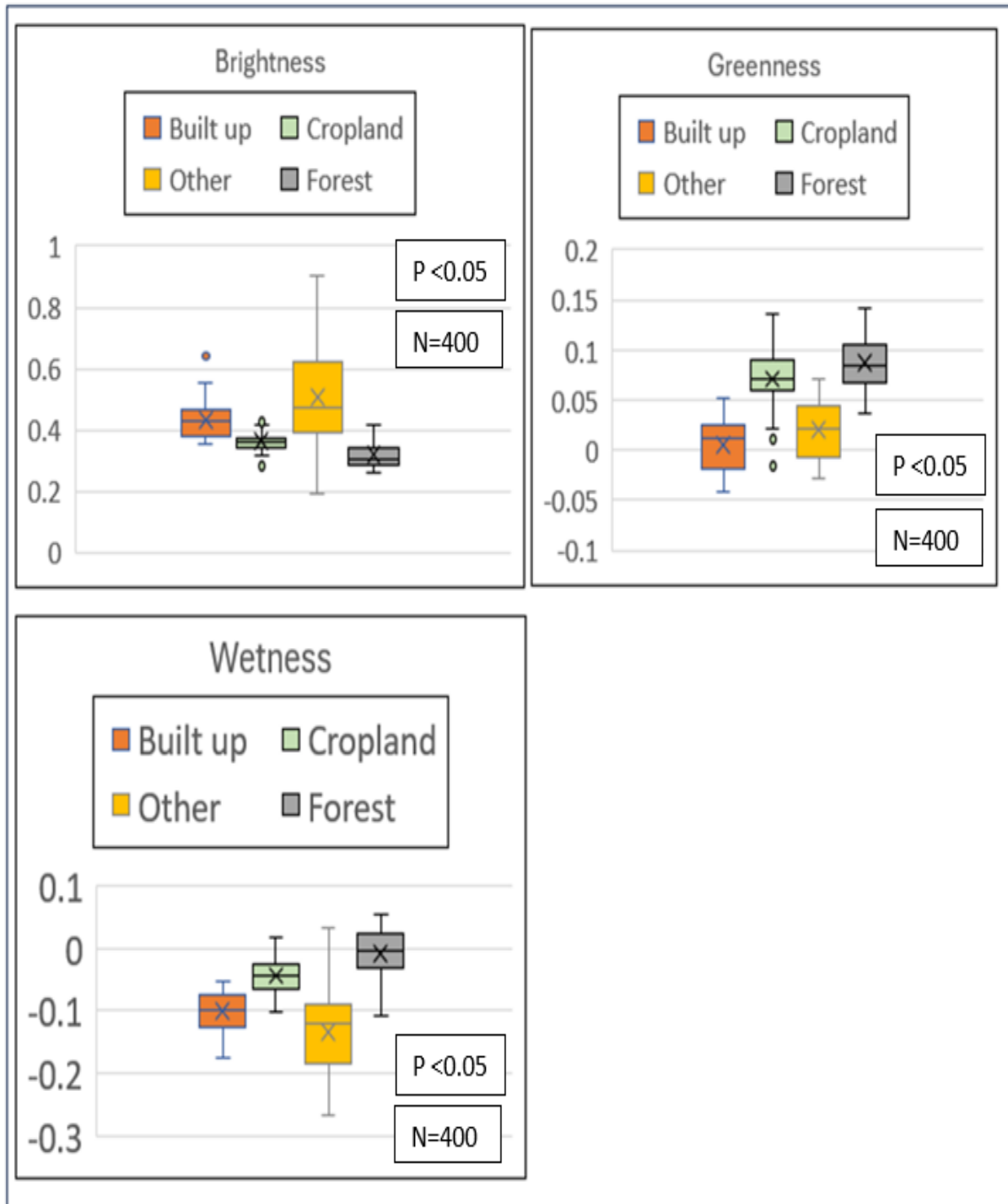
Figure 5: Boxplot showing the distribution of tasselled cap features (Brightness, Greenness, Wetness) across landcover classes: Forest, Cropland, Built up and Other using Landsat data.

Feature importance score for assessing variables

The Figure 6 below shows ranking for all variables that were identified as significant different and was further assessed for similarities.
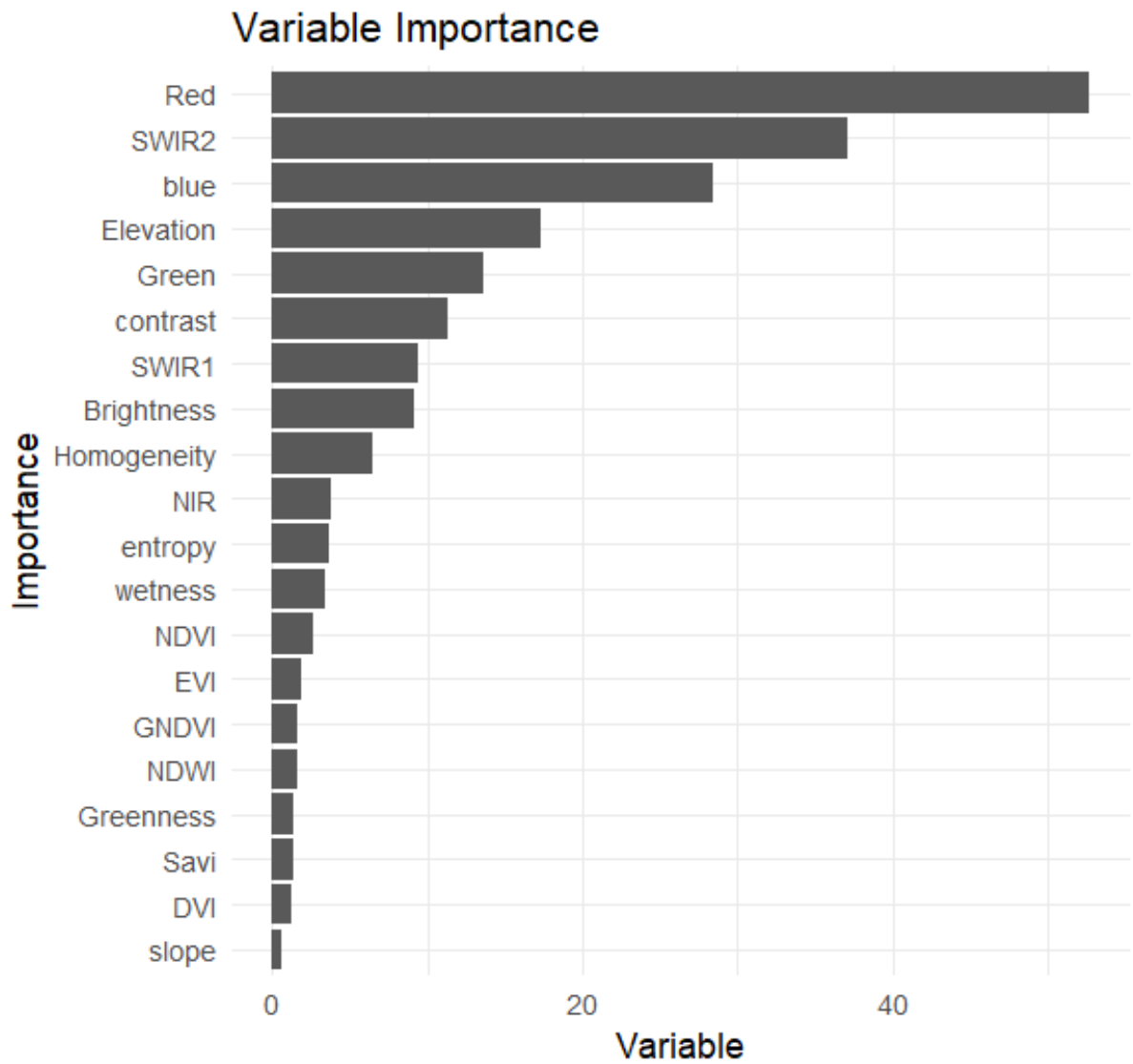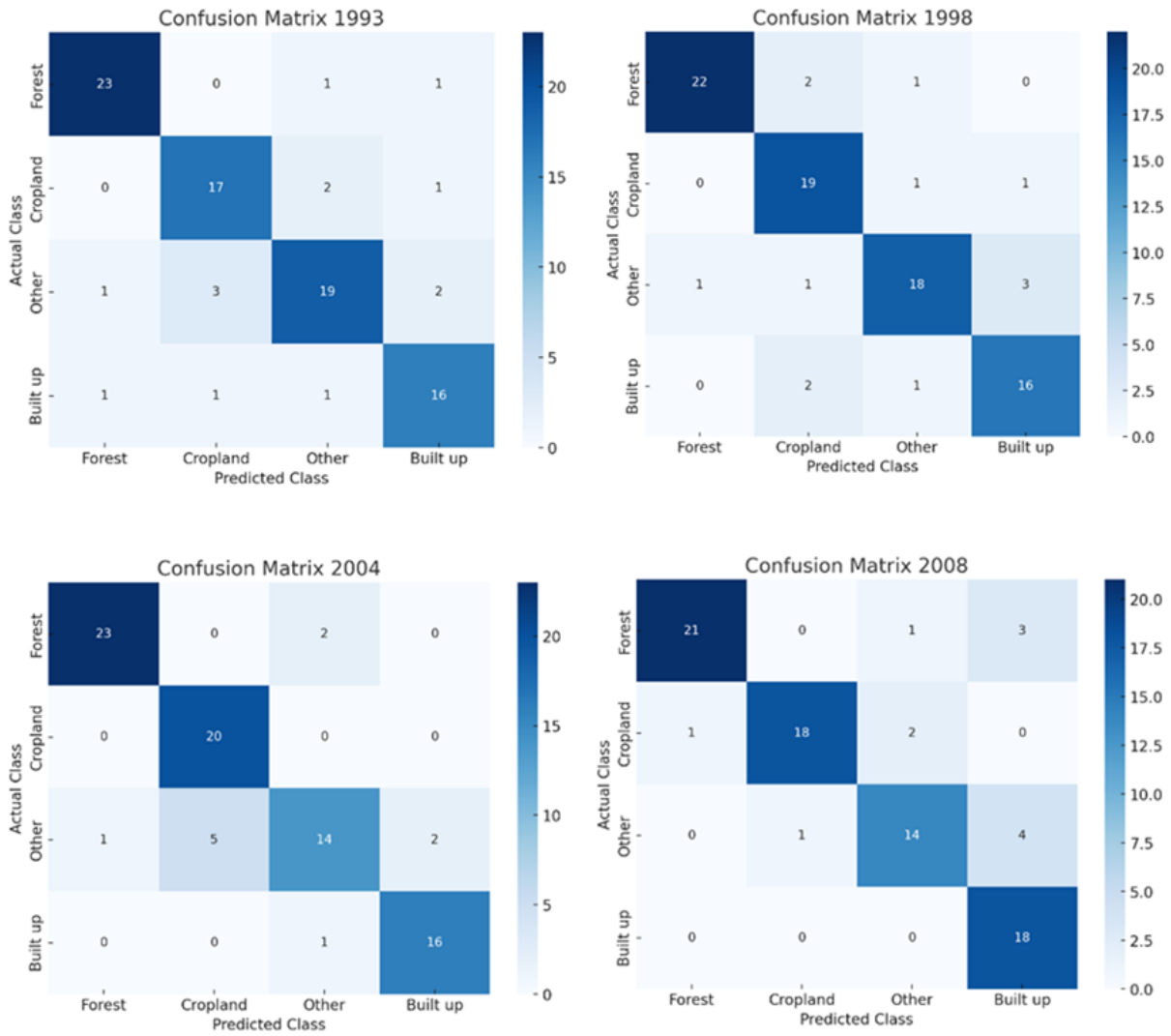

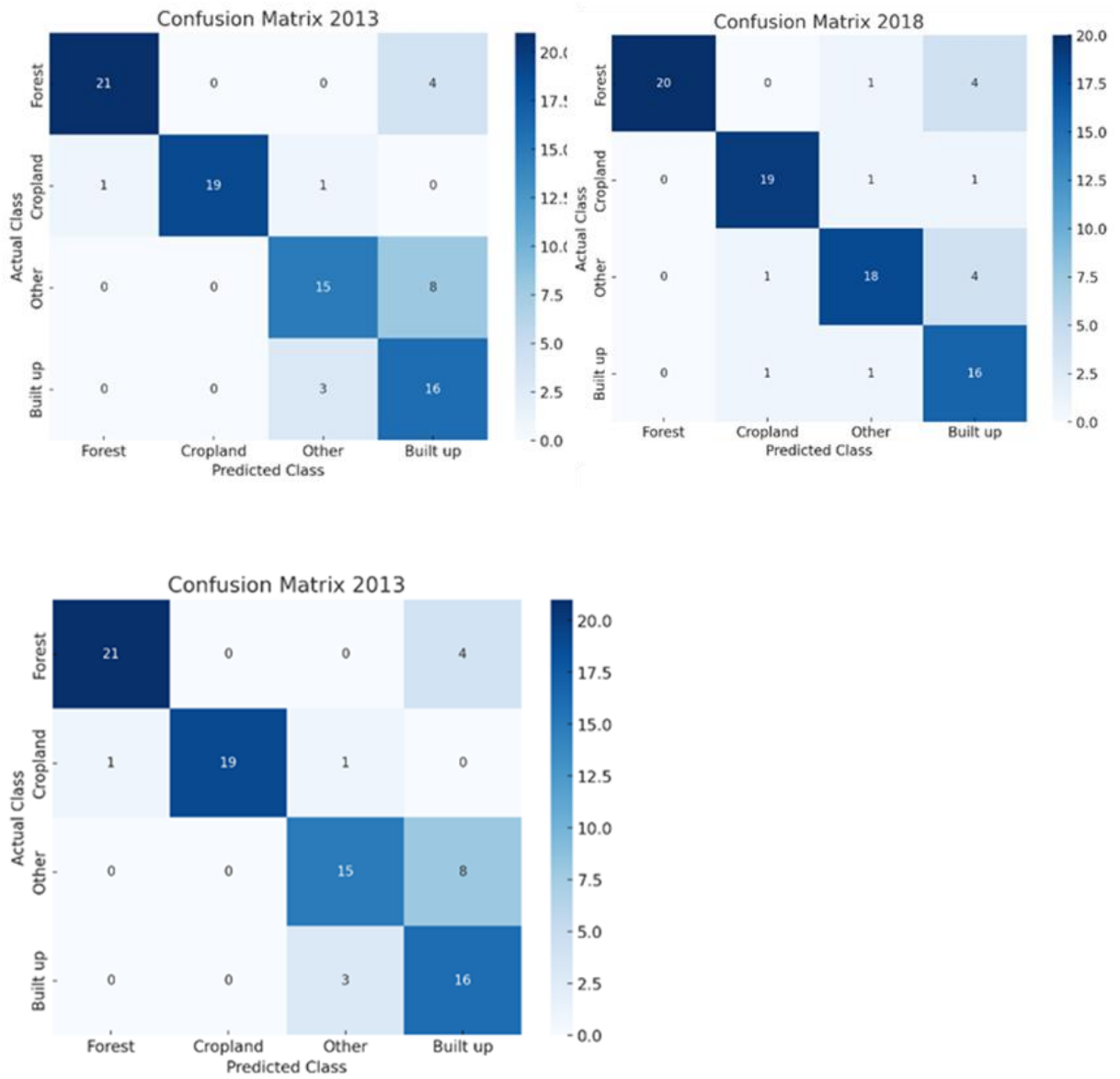Figure 6: Ranking of different variables in landcover classification

Confusion matix

Figure 15: Confusion matrix for the year 2013, 2018, and 2023