Discover Sustainability

Research

Grassland biome fragmentation analysis using sentinal-2 images and support vector machine learning model in South Africa

Andisani Netsianda[1] · **Paidamwoyo Mhangara**[1] · **Eskinder Gidey**[1,2]

Received: 7 June 2024 / Accepted: 14 November 2024
Published online: 02 December 2024

Abstract
The grassland biome, which is classified as a terrestrial ecosystem, contributes significantly to carbon sequestration. About one third of this ecosystem covers the land surface in south Africa and faces the danger of being eradicated. Much of the pressure is attributed to by synthetic initiatives that seeks to expand the economy of the country thereby meeting the demands of cumulative population. Mining, agriculture, and human settlement are the main characters. Given the paucity of research on the threatened ecosystem, the support vector machine learning algorithm (SVM) is employed to investigate fragmentation from 2016 to 2023. We used Sentinel-2A/B satellite images to learn more about spatial and temporal patterns, as well as the distribution of fragmentation in the grassland biome, using the Google Earth Engine platform. The findings revealed that grassland occupied 66% of the area in 2016, decreased to 52% in 2019, and then increased to 59% by 2023. The inconsistency in the pattern or trend of the grassland class is likely attributable to the expansion of the other classes. The SVM model indicated an overall classification accuracy of 97.62%, 97.66% and 98.58% in 2016, 2019, and 2023, respectively. In contrast, the models developed to relate LAI to NDVI, MSAVI2, OSAVI, and NDRE produced R^2 values of 0.6396, 0.6325, 0.6387, and 0.6344, respectively. An in-depth understanding of the fragmentation patterns observed in grasslands yields valuable information for the formulation of conservation strategies, sustainable land-use planning, and climate-resilient management approaches aimed at safeguarding South Africa's distinctive grassland ecosystems.

Keywords Grassland biome · Support vector machines · Fragmentation analysis · Vegetation indices · Change detection

1 Introduction

A large portion of the central and eastern regions of South Africa, where summer precipitation predominates, are covered with grasslands. The land surface coverage is approximately 30%. The grassland biome is composed of five groups: dry highveld grassland, Drakensberg grassland, mesic highveld grassland, Indian ocean costal belt and sub-escarpment grasslands. Grasslands in South Africa host some rare and endangered species. Grasslands play a crucial role in the production of water. The topography and soil richness make grasslands ideal for human settlement and food production. Owing to its underlying geological formations, this biome is home to significant coal and gold reserves [1]. Approximately 45 percent of grassland communities have undergone conversion or deterioration, resulting in highly fragmented natural areas. According to Fourie et al. [2], anthropogenic activities have extensively converted large portions of the concerned biome into mining, agricultural,

✉ Paidamwoyo Mhangara, paida.mhangara@wits.ac.za | [1]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg 2000, South Africa. [2]Department of Land Resources Management and Environmental Protection (LaRMEP), College of Dryland Agriculture and Natural Resources, Mekelle University, P.O Box: 231, Mekelle, Tigray, Ethiopia.

and plantation zones, significantly impacting terrestrial ecosystems. A strong correlation exists between changes in land use and fragmentation, which refers to the conversion of natural vegetation (e.g., grassland) into different land uses (e.g., infrastructure) [3]. On a global scale, statistics show that about 16% of the grassland biome is impacted. This degradation is primarily caused by anthropogenic activities, which are a direct result of population growth. Sibanda et al. [4] assert that mining activities, urban development, climate change, and overgrazing pose a threat to the grassland biome.

Anthropogenic activities, such as the combustion of fossil fuels like coal and oil, emit greenhouse gases like carbon dioxide (CO_2) into the atmosphere, thereby contributing to climate change. The carbon cycle is a widely discussed topic worldwide and alterations in land use have a substantial impact. Grassland ecosystems are among the largest ecosystems on Earth and have a significant impact on the global carbon cycle. The term "grasslands" refers to a type of ecosystem characterized by vast areas covered by grass. Grasslands not only sequester much of their carbon in the soil but also contribute significantly to the global carbon balance [5]. A carbon sink is a designated area with the capacity to absorb a greater amount of carbon than is emitted into the atmosphere. Terrestrial ecosystems sequester approximately one-third of the carbon in the atmosphere. Carbon sequestration capacity of terrestrial ecosystems varies spatially. Grasslands, which constitute 20 percent of the terrestrial ecosystem, also account for 10–30 percent of the global carbon storage. Climate change-induced vegetation expansion in grassland biomes has the potential to influence the carbon cycle in terrestrial ecosystems [6]. To mitigate climate change, Ward et al. [7] recommends capturing carbon dioxide and preserving grassland carbon (C) stocks. Thus, natural resource sustainability is necessary for reducing effects [8].

It is critical to gain a more comprehensive understanding of fluctuations in carbon storage within land-based ecosystems to assess the impact of increasing atmospheric CO_2 levels and shifting temperatures [5]. Pang et al. [9] research defines plant biomass as the amount of organic matter per unit area, which serves as a significant indicator of a plant's health status. Previous studies using remote sensing (RS) satellite sensors, fieldwork data, and regression models have demonstrated the feasibility of estimating the extent of usable grassland areas. Human activity frequently results in changes in land use and cover (LULC). This hinders the functioning of terrestrial ecosystems and the provision of benefits, such as the conservation of plant and animal biodiversity, as well as food and water production [7]. Ongoing changes in land use pose a threat to the grassland biome and significantly impact carbon cycling mechanisms. Fourie et al. [2] stated that the grassland biome in South Africa is highly vulnerable to extinction. Mining operations, urban expansion, changes in climate patterns, and agricultural practices are the main drivers.

We conducted a study on the fragmentation of grassland biomes, utilizing SVM algorithm and Sentinel-2A/B imagery, as it is crucial for safeguarding biodiversity, storing carbon, managing agriculture, and making informed policy decisions through close observation of the environment. There is no classification technique which is immune from limitations, however Support Vector Machines (SVM) has proven to perform better when compared to its rivals Random Forest (RF) and Maximum Likelihood. In previous research both Support Vector Machines (SVM) and Random Forests (RF) exhibited a high level of accuracy in identifying changes in vegetation [10]. Conversely, the RF algorithms can require significant computational resources and are susceptible to overfitting, resulting in outcomes that are frequently challenging to interpret. Likewise, MLC depends on the assumption of data normality, a rare occurrence in real-world scenarios, and is highly sensitive to the quality of training data, often resulting in lower accuracy compared to more advanced algorithms. Notwithstanding these difficulties, each approach possesses unique advantages and can be valuable, contingent upon the demands of the analysis. In the process of classifying earth observation datasets, SVM is extremely effective, and Sentinel-2A/B provides high-resolution, multispectral images that are ideal for detailed environmental monitoring. The primary objective of this study was to examine the changes in the portion of the grassland biome in the Mpumalanga province of South Africa between 2016, 2019, and 2023. Additionally, it aimed to determine the correlation between vegetation indices derived from Sentinel-2-A/B data and field measurements of the Leaf Area Index (LAI). This study is crucial to enhance our understanding of the complex interconnections between anthropogenic activities, ecosystem degradation, and conservation endeavours by conducting an analysis of grassland biome fragmentation. This initiative will enhance the administration and conservation of the country's unique grassland ecosystems.

2 Study area and methodology

This research was conducted in the farming community of Leandra, which is situated to the northwest of Secunda town in the Govan Mbeki municipality in the Mpumalanga province of South Africa. The precise coordinates of the location are −26.268 and −26.467 south and 28.656 and 29.253 east, respectively (Fig. 1). The area under investigation encompasses a total area of 3695 square kilometers, and it is anticipated that the human population will exceed

370,000. The grassland biome, and more specifically, the mesic high-veld grasslands, are one of its constituent parts. The vegetation of this region can be divided into 17 distinct groups, and the deep and rich soils play a role in the diversity of the vegetation found here. Shale and sandstones are examples of sedimentary rocks sliced by gabbro and dolerite dykes [1]. The same authors reported that mesic high-field grassland is home to several activities that have an impact on the economy of the nation. These activities include agriculture, mining, commercial and urban settlement. The threat faced by this terrestrial ecosystem can be attributed to economic drivers [2]. Because of the presence of a few wetland areas, rivers, and pans, this terrestrial ecosystem offers potential solutions to problems associated with water scarcity. The mesic highveld grassland provides a safe haven for a number of endangered species, including the Rudd's Lark, the giant girdled lizard, and African Grass Owl [1]. The capacity of terrestrial ecosystems to produce benefits diminishes because of the ongoing state of degradation and fragmentation [7]. It is well known for its elevation of approximately 1600 m above sea level and flat surface terrain [1]. The region is distinguished by hot and rainy summers and dry and cold winters. Based on the findings of research carried out in 2019, Naidoo et al. estimated that the average annual rainfall in this region ranges from 750 to 900 mm. Both of these characteristics were used to describe the climate of the region. The grassland biome of this region presents an opportunity to grow high-grassland vegetation [1].

2.1 Sentinel-2A/B image acquisition

A combination of free high-resolution imagery produced by satellites (like Sentinel-2, IKONOS, GeoEye, and RapidEye) and machine learning algorithms has contributed significantly to tracking spatio-temporal changes in the environment. These datasets are convenient to developing countries as they are often updated [11]. Sentinel-2 has emerged as one of the most effective sensors in the history of earth observation. Compared to other optical sensors, such as Landsat, this sensor provides remarkably high resolution. Table 1 demonstrates the characteristics of Sentinel-2, a multispectral sensor. The Sentinel-2 sensor was able to disperse the 13 observation bands across the electromagnetic spectrum. It has spatial resolution of 10, 20, and 60 m. Sentinel-2 has a total of three bands in the short-wave infrared

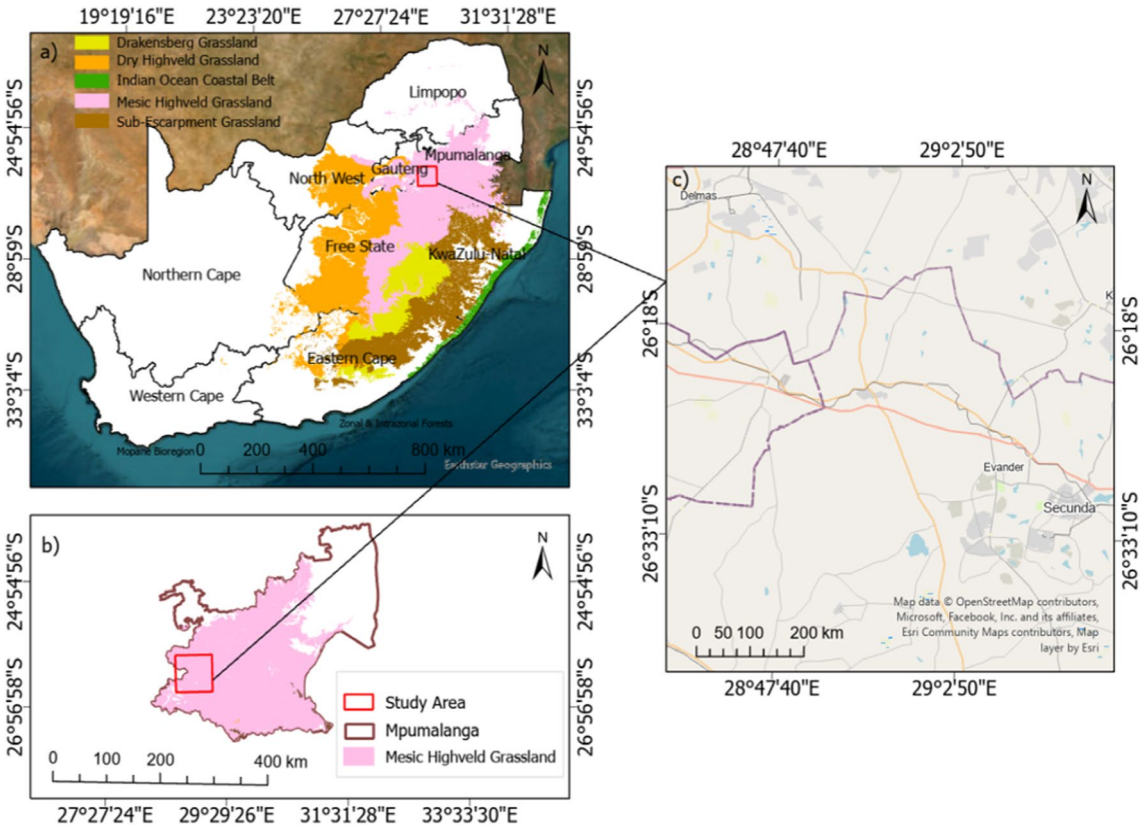

Fig. 1 Map illustrating the study areas at three different levels: national (**a**), provincial (**b**), and local (**c**)

Table 1 Sentinel-2 image characteristics for 2019 and 2023

Band (s)	Spatial resolution (m)	Wavelength (nm)	Description
B1	60	443	Aerosols
B2	10	490	Blue
B3	10	560	Green
B4	10	665	Red
B5	20	705	Red-edge A
B6	20	740	Red-edge B
B7	20	783	Red-edge C
B8	10	842	NIR
B8a	20	865	Red-edge D
B9	60	940	Water vapor
B10	60	1375	Cirrus
B11	20	1610	SWIR
B12	20	2190	SWIR

Table 2 Sentinel-2 satellite image collection for 2016, 2019 and 2023

Number of image collection	Description
2016	
1	S2A_MSIL2A_20161113T075152_N0500_R135_T35JPL_20231027T120152.SAFE
2	S2A_MSIL2A_20161113T075152_N0500_R135_T35JPM_20231027T120152.SAFE
3	S2A_MSIL2A_20161113T075152_N0500_R135_T35JQL_20231027T120152.SAFE
4	S2A_MSIL2A_20161113T075152_N0500_R135_T35JQM_20231027T120152.SAFE
2019	
1	COPERNICUS/S2_SR_HARMONIZED/20190226T074859_20190226T081205_T35JPL
2	COPERNICUS/S2_SR_HARMONIZED/20190226T074859_20190226T081205_T35JPM
3	COPERNICUS/S2_SR_HARMONIZED/20190226T074859_20190226T081205_T35JQL
4	COPERNICUS/S2_SR_HARMONIZED/20190226T074859_20190226T081205_T35JQM
2023	
5	COPERNICUS/S2_SR_HARMONIZED/20230111T075301_20230111T081233_T35JPL
6	COPERNICUS/S2_SR_HARMONIZED /20230111T075301_20230111T081233_T35JPM
7	COPERNICUS/S2_SR_HARMONIZED /20230111T075301_20230111T081233_T35JQL
8	COPERNICUS/S2_SR_HARMONIZED /20230111T075301_20230111T081233_T35JQM

(SWIR) portion of the spectrum, four bands in the visual near-infrared (VNIR) portion of the spectrum, and five bands in the near-infrared (NIR) portion of the spectrum. When combined Sentinel-2A/B has five-day revisit in addition to sharing the same orbit. Nguyen et al. [12] conducted a study that highlighted the distinctive characteristics of the Sentinel-2A/B sensor, which has greatly facilitated its extensive application in vegetation monitoring, land cover detection, and natural resource management. The current study utilized 12 distinct images Table 2. The presence of atmospheric, geometric, and radiometric corrections in the accessed sensor images eliminate the need for further corrections. However, for this investigation, we accessed data using the Google Earth Engine (GEE) platform [13]. The United States Geological Survey Earth Explorer and Copernicus Open Access Hub of the European Space Agency both provide unrestricted access to Sentinel-2 data. For the purpose of this study the data is accessed through the GEE platform.

2.2 Climate data

We acquired monthly precipitation data from the Climate Hazards Center InfraRed Precipitation with Station Data (CHIRPS) website (https://www.chc.ucsb.edu/data) for five years ranging from 2019 to 2023. This timing is selected to ensure an adequate distribution of comprehensive data on seasonal variations. The spatial resolution of the data is 0.05°, which is

roughly equivalent to a distance of 5 km within the study area [14]. This data is used to evaluate the precipitation patterns in the study region on a quarterly basis, representing the four seasons.

2.3 Field data collection for model validation

Prior studies have demonstrated a notable correlation between the data contained in the red, red-edge, green, and near-infrared bands and the LAI. The LI-COR LAI-2200C plant canopy analyzer, developed by Biosciences in the United States, was designed to measure LAI [16]. Data collection was conducted in May 2023, a month characterized by low precipitation. To ensure precise measurements at every location, we strategically positioned the LI-COR LAI-2200C plant canopy analyzer both above and below the canopy. The mean LAI was calculated using three measurements (Table 3). To obtain precise measurements using the LI-COR LAI-2200C plant canopy analyzer, we ensured a uniform level of natural light and prevented direct sunlight from reaching the view of the sensor [15]. To obtain a precise representation of the entire plot, we employed a random stratified sampling technique for each of the five plots established within the study area [17]. We successfully obtained geographical coordinates of 30 reference data points using a Garmin handheld global positioning system (GPS), which has a horizontal accuracy of + 5 m.

2.4 Data processing and analysis

- Sentinel-2 image classification

In this study, we used Sentinel-2A/B sensor imagery to generate classification maps for the periods 2016, 2019, and 2023. We collected images during the summer season, specifically in January and February, when a substantial amount of rainfall is anticipated. This study specifically focused on the growing season, the period during which plants experience growth [18]. We utilized ee.ImageCollection ('COPERNICUS/S2_SR_HARMONIZED') to access Sentinel-2 images from the Google Earth Engine (GEE) platform. To generate cloud-free images, code was developed to apply a mask to the clouds. This code employs the Sentinel-2 QA60 band, which captures clouds and cirrus in 10 and 11 bits, respectively. We employed a filter to manipulate the image and accurately align it with designated dates and cloud cover pixels. During the scene download process, the image resolution was modified to 10 m/pixel [19]. We applied the mosaic function to all scenes that were filtered according to their respective dates. A mosaic technique is employed to create a median composite image. A clipping function is utilized to cut the composite image to represent the study area precisely. We performed image classification on a set of five land use and land cover (LULC) categories. The five categories included the "other" class, which included all the developed areas and mining facilities in the study region, as well as water, agricultural land, grassland, and forest classes. We employed a random selection process to choose the training dataset to ensure an equitable representation of all classes within the study area. The training and validation datasets were handpicked using high-resolution Google Earth satellite images [19].

The classification process uses spectral variations among different land-use types [11]. We acknowledge that per-pixel classification is a highly effective classification method. This technique involves isolating individual pixels in an image and categorizing them as a single LULC type. Among the various classifiers used for analyzing Sentinel-2 imagery, the SVM is frequently regarded as one of the most effective due to its precision and resilience. SVM excels at processing multidimensional data and can effectively handle the intricate spectral information. In terms of overall accuracy, it delivers exceptional classification results, frequently surpassing other techniques like random forest and maximum likelihood. Moreover, SVM is less susceptible to overfitting and demonstrates excellent generalization capabilities, even with limited training datasets. These characteristics establish SVM as a formidable tool for categorizing land cover and utilizing Sentinel-2 imagery. The per-pixel classification methods can be divided into two distinct groups: unsupervised and supervised. Unsupervised classification is a type of classification that operates without the need for human intervention or prior knowledge about the study area. Computer software utilizes automated algorithms to perform this classification, efficiently dividing the image into separate classes by grouping digital values linked to each pixel. A supervised classification method works by having an individual select training sample within the spatial extent of the study area that effectively represents all classes. Supervised

Table 3 An overview of LAI ground samples

Measured variable	Sample quantity	Minimum sample	Maximum sample
LAI	30	3.44 m^2 m^{-2}	^{2} m^{-2}

classification is used in this study. We selected the training dataset using additional resources such as high-resolution Google Earth imagery [20].

- Accuracy assessment

It is not unusual for people to use the terms confusion matrix and error matrix interchangeably, often because they have identical definitions. According to Wei et al. [21], image analysts commonly employ a confusion matrix, also known as an error matrix, to verify the classification accuracy. This study utilizes confusion matrix methods to validate classifications. To guarantee the accuracy of any classification results, it is imperative to first carry out an accuracy assessment procedure. To assess classification accuracy, Nguyen et al. [12] obtained a confusion matrix from a classified map. Nguyen et al. [12] employed the techniques of overall accuracy (OA), kappa coefficient (K), producer's accuracy (PA), and consumer accuracy (CA) to assess classification accuracy. The overall accuracy (OA) is a measure that quantifies the proportion of correctly classified samples out of the total number of samples. We computed the kappa coefficient by dividing the discrepancy between the estimated probability of agreement and observed agreement by 1. Consequently, we subtracted the estimated probability of concurrence from the disparity. When evaluating accuracy at the class level, we utilized both PA and CA. Alternatively, we determined by dividing the number of accurately classified samples in each category by the total number of samples obtained from the reference data in each category. Nasiri et al. (2022) proposes a method to determine the accuracy of each class. We achieve this by using the formula (CA + PA) / 2 to calculate the overall accuracy scores for both the consumer and producer. To assess the accuracy of the classified maps, we allocated 30 percent of the validation dataset for testing purposes and utilized the remaining 70 percent to train the model. To guarantee a sufficient representation of every class, we employed a random selection process to select points from the validation dataset throughout the entire domain [19].

- Change detection analysis

Remote sensing imagery is highly valuable because of its capacity to identify and observe alterations. This method retrieves data from sensors orbiting the Earth. This process generates global sensor images and replicates them at the same location. Various fields, including disaster management, vegetation monitoring, tracking seasonal variations in crop production, and identifying stressed crops can use change detection. This investigation focused on identifying modifications in the post-image classification of images categorized as land cover. The classification change detection method compares two or more independently classified images collected from the same geographical location at different times [22]. This enabled us to monitor the state of the object as it evolved over time. One major advantage of using post-classification as a method for change detection is that it allows for separate classification of satellite images from two different dates. This makes it easier to adapt to the changes in the atmosphere and sensors that occur between the two dates. Furthermore, this method resolves the problem of reliable image registration for multiple dates, which is an additional benefit [22].

- Computing spectral indices

Four distinct vegetation indices (VIs) were calculated using GEE scripts. Shi et al. [23] stated that information obtained from the visible and near infrared portions of the electromagnetic spectrum is useful in highlighting vegetation. According to Yu et al. [18] NDVI is considered as the most common ratio. This index is highly beneficial; however, its inability to eliminate background soil noise poses a significant challenge. To take this into account, scientists have developed better versions of the Vegetation Index (VI), such as the Modified Soil-Adjusted Vegetation Index (MSAVI) and Optimized Soil-Adjusted Vegetation Index (OSAVI). The former has incorporated soil-adjustment factor and have the ability to eliminate unwanted soil interference while enhancing vegetation [24]. The values of all VIs ranged from -1 to 1. Low VIs values signify the existence of rocky outcrops and bare soil. Average values signify the existence of grasslands and shrubs, whereas high values indicate the existence of dense vegetation or forests (Table 4) [25].In this study, we examined the Normalized Difference Vegetation Index (NDVI) using the equation (Eq. 1):

$$\mathrm{NDVI} = \mathrm{NIR} - \mathrm{red} \,/\, \mathrm{NIR} + \mathrm{Red} \tag{1}$$

The underlying principle of OSAVI and MSAVI is the same as that of NDVI. They both use near-infrared and red bands of the electromagnetic spectrum. However, they are designed to eliminate the interference caused by soil while enhancing vegetation. The value of soil adjustment factor depends on the nature of biome in question. For example, arid regions, such as the grassland biome, tend to use a soil adjustment factor of 0.16, considering the sparse vegetation cover.

$$\text{OSAVI} = (1 + \text{L})\,\text{NIR} - \text{red}/\,\text{NIR} + \text{red} \tag{2}$$

$$\text{MSAVI} = \text{L} * ((2 * \text{NIR} + 1) \cdot -\sqrt{(2 * \text{NIR} + 1)^2} - 8 * (\text{NIR - Red}) \tag{3}$$

where L stands for the soil-adjusted factor.

We computed the normalized difference red-edge index (NDRE) using the near-infrared (NIR) and red-edge bands within the electromagnetic spectrum. NDRE substitutes for the red band in the conventional NDVI formula. The only distinguishing features between NDVI and NDRE are red and red-edge bands [26]. In the electromagnetic spectrum, the Sentinel-2 sensor's red-edge position falls between 690 and 750 nm (bands 5, 6, and 7). According to Chen et al. [29], the red edge of the spectrum contains important data on the chlorophyll content of the vegetation, which is a critical measure of vegetation production.

$$\text{NDRE} = \text{NIR} - \text{Red}\ /\ \text{NIR} + \text{Red} \tag{4}$$

The output of the four indices were used to compute linear regression models. VIs as predictor variables, and LAI as the predicted variable. In this case, Microsoft Excel is a data analysis tool used for constructing linear regression models [17]. We partitioned the thirty (30) data points obtained from the five plots into two separate sets: the training and testing datasets. We used a random selection process to allocate 80 percent of the training data for model training and reserved the remaining 20 percent for subsequent model testing. In their study, [26] employed metrics such as the root mean square error (RMSE) and the coefficient of determination to assess the effectiveness of the models they developed. The current study adopts this approach.

- Grassland biome fragmentation analysis using machine learning

This study utilized SVM learning algorithm to analyze the fragmentation of grassland biome. We conducted an analysis using the Google Earth Engine (GEE) platform. We interconnected a sequence of codes starting from the initial stages of collecting the Sentinel-2 images. Clouds were concealed using the Sentinel-2 QA60 band, which covers clouds and cirrus at 10 and 11 bits, respectively. A supplementary filter was used to eliminate the remaining cloud-covered pixels by defining the percentage of cloudiness. We resampled Sentinel-2 sensor data to a scale of 10 m. The subsequent stages of the code include image mosaicking and creation of a median composite image. The median composite was cropped using an ROI shaped file. We used five manually created samples for supervised SVM classification. The main aim of SVM is to employ dividers to construct a classification hyperplane that appropriately categorizes each sample while also isolating the samples. The method uses a non-parametric classification method, meaning the data is not assumed to follow a normal distribution. While SVM is a powerful classifier, it is complex because it requires the user to select the kernel function and other parameters. Prior studies have praised this method for its outstanding efficacy compared with alternative classification techniques [27]. The investigation employed a SVM model with a 70% training dataset and 30% validation dataset ratio. To evaluate the efficacy of the various models, we calculated the R^2 value, coefficient of determination, and root-mean-square error [28].

Natural or anthropogenic activities alter ecosystems and habitats, leading to the creation of smaller and separate patches within the original continuous area, a process known as fragmentation. Fragmentation in a terrestrial ecosystem is defined as the reduction in native vegetation, division of the remaining vegetation into smaller fragments or patches,

Table 4 The NDVI values correspond to the object on the surface

Land cover type	NDVI values
Waterbodies	–1.0
Bare soil	0.0–0.2
Shrubs and grassland	0.2–0.5
Dense vegetation and forest	< 0.5

and conversion of the landscape from native vegetation to different land uses [3]. To examine the division of grasslands, we employed the Landscape Ecology Statistics (LecoS). LecoS was installed as a plugin within the QGIS window interface. This extension is highly valuable for computing landscape matrices that provide comprehensive data on the fragmentation status of study areas such as grasslands. We imported previously classified LULC maps into the LecoS plugin. The classified maps consist of five distinct categories. The categories included "water," "agricultural land," "grassland," "forest," and "other" [29]. Nine distinct landscape matrices are used in this study. The matrices encompassing the land cover, landscape proportion, edge length, number of patches, greatest patch area, smallest patch area, overall core area, patch cohesion index, and landscape division Table 5 [30] provide a thorough depiction of landscape matrices.

3 Results

3.1 Evaluation of LULC from 2019 to 2023

Figure 2 illustrates the spatial distribution of various land cover categories within the designated study area. The findings suggest a gradual transition from grassland to agriculture and other land use categories. According to Table 6, grassland accounted for 66%, 52%, and 59% of the land in 2016, 2019, and 2023, respectively. Overall, grasslands experienced 7% growth. The expansion of other and agricultural classes may have contributed to the decrease in land occupation by grassland. The 'other' category also witnessed a 13% increase from 2016 to 2023. We anticipate the expansion of the 'other' category due to the inclusion of developed regions, mines, and uninhabited land. Anticipated population growth will lead to an increased need for residential areas and a corresponding expansion of mining areas to meet heightened demand for resources. The proportion of agricultural classes has decreased from 20% in 2016 to 15% in 2023. This signifies a 5% decrease in agricultural activity. The unanticipated decline in agricultural activities contrasts with our projections considering the growing population and heightened food demand. The water and forest categories lost 1% of their land between 2016 and 2023. In 2016, the agriculture class had the highest number of patches, and the water class had the lowest. The grassland category had the highest number of patches in 2019, whereas the water category had the lowest number. The grassland class displayed the most pronounced fragmentation in 2019, characterized by a high degree of disconnection and isolation.

During that particular year, the water class demonstrated the lowest amount of fragmentation, indicating a greater level of connectivity. In 2023, the other classes achieved the highest number of patches, indicating a notable degree of fragmentation. During that year, the water category exhibited the lowest number of patches. Consequently, the water category demonstrated the least amount of fragmentation during both years of investigation. In both 2019 and 2023, grasslands exhibited the highest level of patchiness, whereas forests displayed the lowest. As expected, grassland occupied the largest expanse of land, whereas forests occupied the smallest. Between 2016, 2019, and 2023, the landscape divisions of all classes consistently surpassed 0.9, approaching 1. This means that a landscape division with a value approaching 1 indicates a low level of connectivity [30]. Consequently, the study's classes exhibited significant fragmentation and patchiness. In 2016 and 2023, the grassland had a landscape fragmentation index of 0.74 and 0.75,

Table 5 Landscape matrix description

Landscape matrices	Description
Land cover	The total area covered by each landscape class
Landscape proportion	Proportion of area that each class takes up in the total landscape
Edge length	Total length of all the edge patches
Number of patches	Number of discrete patches identified for each class
Greatest patch area	Greatest patch area in a particular landscape for a class
Smallest patch area	Smallest patch area in a particular landscape for a class
Overall core area	The function reduces the class boundary by one cell value and returns its total area
Patch cohesion index	Numbers closer to zero indicate a class's fragmentation and physical disconnection, while higher numbers indicate cohesiveness
Landscape division	Total patch area in one class divided by total landscape area (values closer to 1 indicate a very patchy landscape, while values closer to zero indicate a single environment)

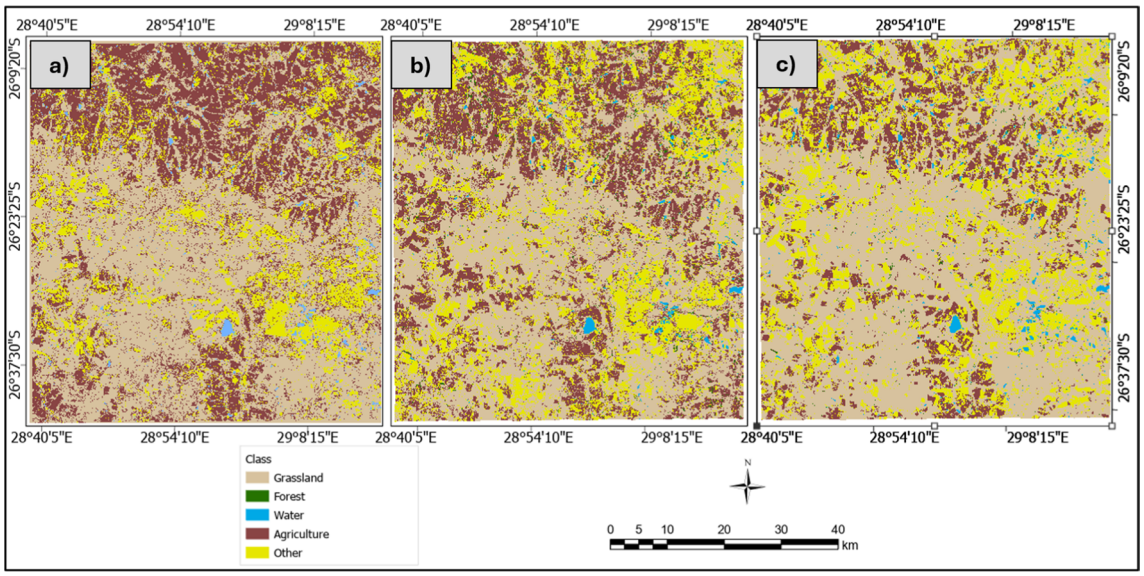

Fig. 2 Land cover types of the study area in 2016 (**a**), 2019 (**b**) and 2023 (**c**)

Table 6 The 2016, 2019 and 2023 classification maps' landscape matrices

Class name	Area in km^2	Landscape proportion	Edge length	Number of patches	Greatest patch area in km^2	Overall Core area	Landscape division
2016							
Grassland	2680.522	66%	63.6	110442	2041.399	2116.364	0.74
Forest	101.338	2%	16.1	153604	0.7798	19.399	1.00
Water	67.0412	2%	2.9	11511	4.6414	44.3661	0.99
Agriculture	798.2946	20%	59.4	229245	21.4782	394.3393	0.99
Other	422.6298	10%	35.9	164507	92.917	179.9477	0.99
2019							
Grassland	1915.25	52%	60.1	146277	837.54	1466.10	0.94
Forest	52.33	1%	9.09	86394	0.19	7.34	1.00
Water	31.32	1%	0.77	1592	4.11	25.00	1.00
Agriculture	936.28	25%	42.1	111618	12.43	618.34	1.00
Other	759.85	21%	41.5	125459	296.29	463.97	0.99
2023							
Grassland	2198.08	59%	39.16	76325.00	1845.99	1885.07	0.75
Forest	21.39	1%	3.80	36826.00	0.20	3.10	1.00
Water	46.21	1%	1.19	2687.00	3.94	36.53	1.00
Agriculture	572.01	15%	16.07	43129.00	7.82	448.86	1.00
Other	857.43	23%	33.82	85387.00	109.64	606.35	1.00

respectively, indicating a slightly lower degree of fragmentation than the other categories. The given value suggests that the grassland categories in 2016 and 2023 had less fragmentation than the same category in 2019, although the value is still close to 1 and represents a state of disconnectedness. Based on the results of this study and supporting literature, it is evident that all classes exhibit a notable degree of fragmentation and restricted connectivity. However, there were signs of recuperation in the grassland category.

The current study reveals changes in land use across various categories, with both increases and decreases, during the five-year period under examination, despite not being easily observable. Between 2016 and 2023, an observed change occurred in grasslands and other categories, with corresponding increases of 7% and 13%, respectively. Between 2016

and 2023, there was a decrease of 5% in the amount of agricultural land. These findings indicate that the ecosystem experienced continuous and persistent alterations.

3.2 LULC accuracy assessment

Table 7 shows the confusion matrix results, whereas Table 8 presents the accuracy evaluation outcomes for the classified maps for 2016, 2019, and 2023. In this context, accuracy evaluation relies on a confusion matrix. The overall accuracy (OA) in 2016 and 2019 was 98.27% and 97.66%, and the kappa statistics were 97.62 and 97.07%, respectively. In 2023, OA achieved a score of 98.58%, along with a kappa statistic of 98.14%. Table 8 presents the OA and kappa coefficients.

The findings indicated that the OA and kappa statistics for both years exceeded 97%, and the values were relatively close. Nevertheless, the 2023 outperformed the 2016 and 2019 classifications in terms of OA. In both years, the CA and PA for each class were higher than 80%, indicating satisfactory outcomes. SVM method has demonstrated high performance achieving the high levels of accuracy [12].

3.3 Analysis of the relationships between vegetation indices and biophysical parameter LAI

The vegetation indices NDVI, MSAVI2, OSAVI, and NDRE were significantly correlated with LAI, with R^2 values of 0.6396, 0.6325, 0.6387, and 0.6344, respectively (Fig. 3). The results indicated that all models exhibited satisfactory performance, as evidenced by an R^2 value exceeding 60% and an RMSE value approaching zero. The NDVI vs. LAI model performed better than the other models, because the independent variable NDVI explained 63.96% of the data variation. The OSAVI model was the second most effective, explaining 63.87% of the variability in the data. NDRE ranked third in terms of performance, explaining 63.44% of the variation in data. MSAVI2 had the lowest performance, accounting for 63.25% of data variation. Models with lower RMSE values were considered to have excellent fits [18]. The models in this study demonstrated a satisfactory fit, as indicated by their low root-mean-square error (RMSE) values. The models exhibited statistical significance, indicating a robust correlation between all vegetation indices (VIs) and the LAI.

3.4 3.4 Rainfall vs NDVI analysis

Figure 4 shows the average temporal changes in NDVI over a span of 5 years in the study area. As anticipated, NDVI values exhibited a notable increase during the wet season (January), signifying robust vegetation. However, these values gradually declined as the weather transitioned to drier conditions during subsequent months. During the

Table 7 Confusion matrix for the periods 2016–2023

Class name	Grassland	Forest	Water	Agriculture	Other
2016					
Grassland	56	0	0	0	0
Forest	2	10	0	0	0
Water	0	0	38	0	0
Agriculture	0	0	0	41	0
Other	0	0	1	0	64
2019					
Grassland	70	0	0	1	0
Forest	0	33	0	0	0
Water	0	0	41	0	0
Agriculture	0	0	0	10	14
Other	1	0	0	0	17
2023					
Grassland	41	0	0	0	0
Forest	1	37	0	0	2
Water	0	0	45	0	0
Agriculture	1	0	0	41	0
Other	1	0	0	0	45

Table 8 Results of the accuracy assessment 2016–2023

Class name	Consumer accuracy (%)	Producer accuracy (%)	Kappa statistics (%)	Overall accuracy (%)
2016				
Grassland	98	99		
Forest	100	100		
Water	100	100	97.62	98.27
Agriculture	90	91		
Other	94	94		
2019				
Grassland	93	100		
Forest	100	92		
Water	100	100	97.07	97.66
Agriculture	100	97		
Other	95	97		
2023				
Grassland	96	100		
Forest	100	83		
Water	97	100	98.14	98.58
Agriculture	100	100		
Other	100	98		

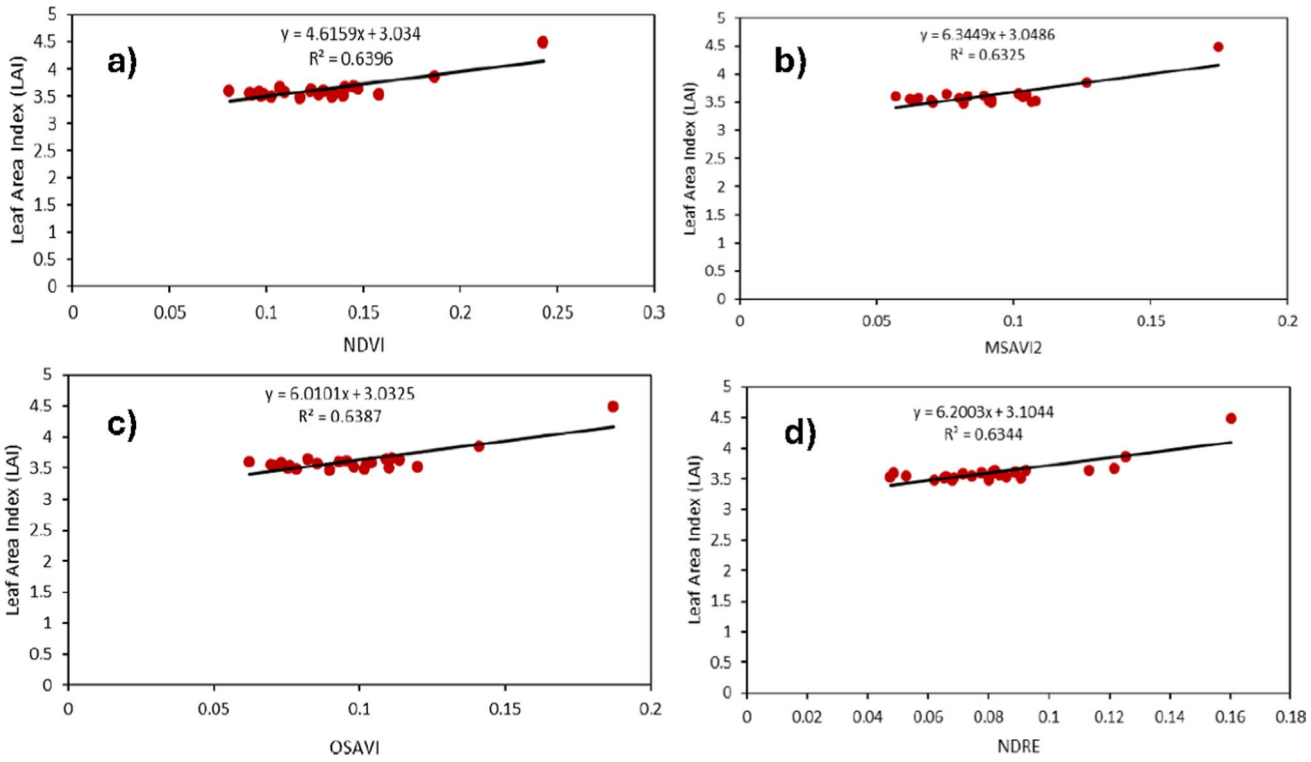

Fig. 3 Statistical relationship between the LAI and other vegetation indices

dry season (May to October), low NDVI values indicated the presence of barren soil or dry vegetation. Urban areas and landfills had lower NDVI values. Figure 5 shows the chronological sequence of the mean monthly precipitation

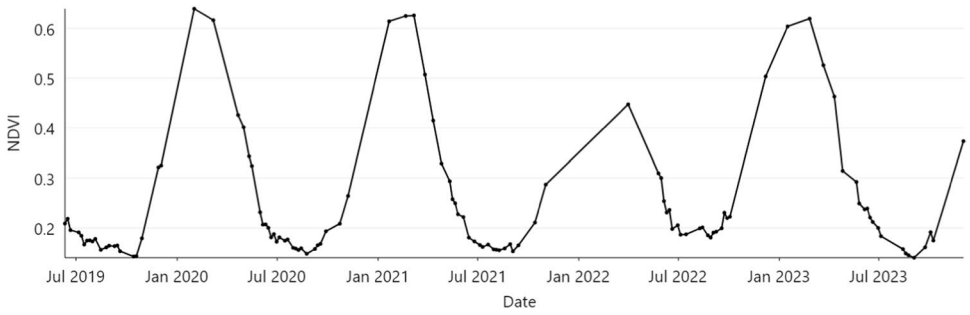

Fig. 4 Temporal changes in NDVI

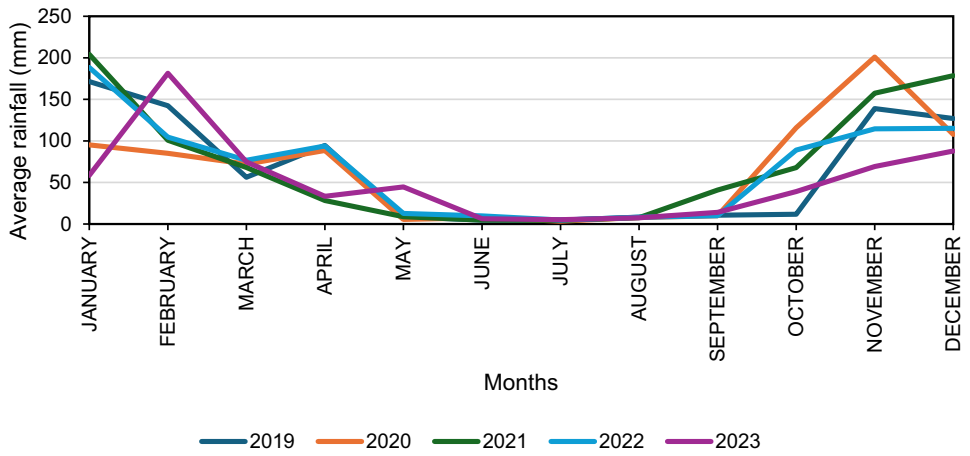

Fig. 5 Average monthly rainfall over a five-year period

from 2019 to 2023. A consistent pattern of low rainfall is observed from May to August. Although the onset of the wet season was uncertain, it was established in November. Historically, December and January have consistently experienced the greatest amount of rainfall. NDVI exhibited a positive correlation with the average monthly rainfall, as shown in Figs. 4 and 5. The reduction in precipitation from May to September was associated with a decrease in NDVI during the same time frame. There was a strong correlation between rainfall and the NDVI.

4 Discussion

The fragmentation results presented in this study provide evidence that the grassland biome is highly fragmented. Fourie et al. [2] study, indicating the biome's global vulnerability, supports these findings. As expected, the grassland class covered most of the land in both years; however, the high number of patches and the landscape division, which is closer to 1, indicate that the grassland environment is extremely patchy and highly disconnected. The information presented suggests the presence of activities that contribute to this fragmentation. As in the study conducted by Cadman et al. [1], in the current study anthropogenic activity are the major contributors to the landscape divisions. These activities are attracted by the rich geological formations, topographical landscape, as well as the deep and nutritious soil. The classification results highlighted the exceptional performance of the SVM classifier. In both years, the overall accuracy, which is a measure of performance, was greater than 90%. Just like any other technique, SVM possesses its own flaws. In this regard, we noticed the tendency of misclassification between the grassland and the agriculture class. Although efforts were put in place to avoid this, it was not possible to completely eliminate the issue. Some pixels in the agriculture class were identical to those of the grassland class, especially fields that were matured and close to harvest turning brown in color resembling the latter. This study acknowledges the uncertainty this might have introduced in the outcome. Furthermore, the study investigated the use of vegetation indices to assess the vegetation conditions. The anticipated behavior of the vegetation indices showed that seasonal variations significantly influenced vegetation growth. We observed this phenomenon when wet seasons experienced higher values of vegetation indices than dry seasons. Additionally, our findings showed that vegetation produces more biomass

during the wet season than during the dry season. During the wet season, we anticipated a significant amount of rainfall, whereas during the dry season, we anticipated very little to no rain. Rainfall data provided a valuable supplement to the vegetation index data, confirming our expectations. According to Gebregergs et al. [31], a decline in vegetation cover may have a negative impact on the environment and community livelihoods, necessitating regular evaluation. To understand the impact of these factors, it is critical to investigate the patterns of rainfall and vegetation biomass. Sentinel-2 derived VIs (NDVI, MSAVI2, OSAVI, and NDRE) were used to identify possible links with the biophysical parameter LAI. All VIs achieved satisfactory results, NDVI yielded the best results. However, these findings contradict literature, we expected NDVI to yield the least results due to its inability to eliminate soil background noise. The high R^2 and low RMSE values indicated that the results were statistically significant. Yu et al. [18] research literature draws this conclusion, stating that high R^2 and low RMSE values are indicative of a well-performing model. The results show that vegetation indices are valuable tools for quantifying biomass. Despite financial constraints, scarcity of validation datasets, restricted access to certain survey areas, and adverse weather conditions, this study successfully achieved its aims. The findings revealed that the VI vs. LAI models were statistically significant, with an R^2 value greater than 0.6 for all models.

5 Conclusion

This study analyzed grassland biome changes from 2016 to 2023 using the SVM model and established a relationship between vegetation indices and LAI measurements. The results indicated a gradual transition from grassland to agriculture and other land use categories within the designated study area. This is likely due to the expansion of developed regions, mines, and agricultural land. Despite not being easily observable, the study reveals changes in land use across various categories, with both increases and decreases. The observed statistics concerning the accuracy of the SVM classification was significant. No technique is exempted from flaws, this is including SVM. Even with its shortcoming the technique yielded good results. This study shows that SVM is a reliable and effective classification method. There was a significant correlation between LAI and NDVI, MSAVI2, OSAVI, and NDRE. This conclusion is based on the R^2 values that are high and above 60% and RMSE values nearing zero. Overall, the abundance of patches within the grassland class signifies the biome's fragmentation and isolation. This study also demonstrated the influence of rainfall on biomass, higher rainfall results in a high biomass yield. This will help improve grassland ecosystem management and conservation in the study area. Future research should consider using more than one classification method for complementary purposes and multiple field parameters to improve the assessment of grassland biome fragmentation.

Author contributions Conceptualisation: A.N., P.M. E.G., Methodology: A.N., P.M. E.G., Validation: A.N., P.M. E.G., Formal analysis: A.N., P.M. E.G., Investigation: A.N., P.M. E.G., Writing-Original Draft: A.N., P.M. E.G., Writing-Review & Editing: A.N., P.M. E.G.,

Funding No funding was provided for this research.

Data availability The data used in the study is freely available on the Copernicus Open Access Hub.

Declarations

Ethical approval and consent to participate. The University of the Witwatersrand granted the ethical approval for the research. No participants were involved in this research.

Consent to publications No consent to publish is required for this research.

Competing interests The authors declare no competing interests.

References

1. Cadman M, De Villiers C, Lechmere-Oertel R, McCulloch D. Grassland ecosystem guidelines: landscape interpretation for planners and managers. South African National Biodiversity Institute, Pretoria. 2013
2. Fourie L, Rouget M, Lötter M. Landscape connectivity of the grassland biome in M pumalanga S outh A frica. Austral Ecol. 2015;40(1):67–76. https://doi.org/10.1111/aec.12169.
3. Tao B, Yang Y, Yang J, Smith R, Fox J, Ruane AC, Ren W. Recent shrinkage and fragmentation of bluegrass landscape in Kentucky. Remote Sens. 2020;12(11):1815. https://doi.org/10.3390/rs12111815.
4. Sibanda M, Mutanga O, Rouget M, Kumar L. Estimating biomass of native grass grown under complex management treatments using worldview-3 spectral derivatives. Remote Sens. 2017;9(1):55. https://doi.org/10.3390/rs9010055.
5. Wang LK, Lian PY, Fu YJ. Impact of land-use change on grassland carbon stocks: an overview of the literature. Appl Mech Mater. 2014;448:948–51. https://doi.org/10.4028/www.scientific.net/AMM.448-453.948.
6. Du L, Gong F, Zeng Y, Ma L, Qiao C, Wu H. Carbon use efficiency of terrestrial ecosystems in desert/grassland biome transition zone: a case in Ningxia province, northwest China. Ecol Ind. 2021;120: 106971. https://doi.org/10.1016/j.ecolind.2020.106971.
7. Ward A, Yin KS, Dargusch P, Fulton EA, Aziz AA. The impact of land use change on carbon stored in mountain grasslands and shrublands. Ecol Econ. 2017;135:114–24. https://doi.org/10.1016/j.ecolecon.2016.12.023.
8. Gidey E, Gitet S, Mhangara P, Dikinya O, Hishe S, Girma A, Birhane E. Impact of urban and peri-urban growth on arable land (1976–2029) in a medium sized city of Shire Indaselassie, North Western Tigray Ethiopia. SN Appl Sci. 2023;5(4):102. https://doi.org/10.1007/s42452-023-05322-x.
9. Pang H, Zhang A, Kang X, He N, Dong G. Estimation of the grassland aboveground biomass of the inner Mongolia plateau using the simulated spectra of sentinel-2 images. Remote Sens. 2020;12(24):4155. https://doi.org/10.3390/rs12244155.
10. Makobe B, Mhangara P, Gidey E, Kganyago M. Monitoring the invasion of campuloclinium macrocephalum (less) DC plants using a novel MaxEnt and machine learning ensemble in the cradle nature reserve. South Africa Environ Syst Res. 2024;13(1):24. https://doi.org/10.1186/s40068-024-00351-w.
11. Gidey E, Mhangara P. An application of machine-learning model for analyzing the impact of land-use change on surface water resources in Gauteng Province South Africa. Remote Sens. 2023;15(16):4092. https://doi.org/10.3390/rs15164092.
12. Nguyen HTT, Doan TM, Tomppo E, McRoberts RE. Land Use/land cover mapping using multitemporal Sentinel-2 imagery and four classification methods—a case study from Dak Nong Vietnam. Remote Sens. 2020;12(9):1367. https://doi.org/10.3390/rs12091367.
13. Vajsová B, Fasbender D, Wirnhardt C, Lemajic S, Devos W. Assessing spatial limits of Sentinel-2 data on arable crops in the context of checks by monitoring. Remote Sens. 2020;12(14):2195. https://doi.org/10.3390/rs12142195.
14. Murungweni FM, Mutanga O, Odiyo JO. Rainfall trend and its relationship with normalized difference vegetation index in a restored semi-arid wetland of South Africa. Sustainability. 2020;12(21):8919. https://doi.org/10.3390/su12218919.
15. Masemola C, Cho MA, Ramoelo A. Comparison of landsat 8 OLI and Landsat 7 ETM+ for estimating grassland LAI using model inversion and spectral indices: case study of Mpumalanga, South Africa. Int J Remote Sens. 2016;37(18):4401–19. https://doi.org/10.1080/01431161.2016.1212421.
16. Danner M, Locherer M, Hank T, Richter K 2015. Measuring Leaf Area Index (LAI) with the LI-Cor LAI 2200C or LAI-2200 (+2200Clear Kit) – theory, measurement, problems, interpretation, EnMAP Flight Campaigns Technical Report; GFZ Data Services. https://doi.org/10.2312/ENMAP.2015.009
17. Shafian S, Rajan N, Schnell R, Bagavathiannan M, Valasek J, Shi Y, Olsenholler J. Unmanned aerial systems-based remote sensing for monitoring sorghum growth and development. PLoS ONE. 2018;13(5): e0196605. https://doi.org/10.1371/journal.pone.0196605.
18. Yu R, Yao Y, Wang Q, Wan H, Xie Z, Tang W, Bei X. Satellite-derived estimation of grassland aboveground biomass in the three-river headwaters region of China during 1982–2018. Remote Sens. 2021;13(15):2993. https://doi.org/10.3390/rs13152993.
19. De Luca G, Silva MN, Di Fazio S, Modica G. Integrated use of Sentinel-1 and Sentinel-2 data and open-source machine learning algorithms for land cover mapping in a Mediterranean region. Euro J Remote Sens. 2022;55(1):52–70. https://doi.org/10.1080/22797254.2021.2018667.
20. Li M, Zang S, Zhang B, Li S, Wu C. A review of remote sensing image classification techniques: the role of spatio-contextual information. Euro J Remote Sens. 2014;47(1):389–411. https://doi.org/10.5721/EuJRS20144723.
21. Wei P, Zhu W, Zhao Y, Fang P, Zhang X, Yan N, Zhao H. Extraction of Kenyan grassland information using PROBA-V based on RFE-RF algorithm. Remote Sens. 2021;13(23):4762. https://doi.org/10.3390/rs13234762.
22. Singh A. Review article digital change detection techniques using remotely-sensed data. Int J Remote Sens. 1989;10(6):989–1003. https://doi.org/10.1080/01431168908903939.
23. Shi Y, Gao J, Li X, Li J, Torre DMG, Brierley GJ. Improved estimation of aboveground biomass of disturbed grassland through including bare ground and grazing intensity. Remote Sens. 2021;13(11):2105. https://doi.org/10.3390/rs13112105.
24. Wang Z, Ma Y, Zhang Y, Shang J. Review of remote sensing applications in grassland monitoring. Remote Sens. 2022;14(12):2903. https://doi.org/10.3390/rs14122903.
25. Muavhi N. A simple approach for monitoring vegetation change using time series remote sensing analysis: a case study from the Thathe Vondo Area in Limpopo Province, South Africa. South African J Sci. 2021;117:1–9. https://doi.org/10.17159/sajs.2021/8226.
26. Zhang Y, Yang W, Sun Y, Chang C, Yu J, Zhang W. Fusion of multispectral aerial imagery and vegetation indices for machine learning-based ground classification. Remote Sensing. 2021;13(8):1411. https://doi.org/10.3390/rs13081411.
27. Naidoo L, van Deventer H, Ramoelo A, Mathieu R, Nondlazi B, Gangat R. Estimating above ground biomass as an indicator of carbon storage in vegetated wetlands of the grassland biome of South Africa. Int J Appl Earth Obs Geoinformation. 2019;78:118–29. https://doi.org/10.1016/j.jag.2019.01.021.
28. Nasiri V, Deljouei A, Moradi F, Sadeghi SMM, Borz SA. Land use and land cover mapping using sentinel-2, landsat-8 satellite images, and google earth engine: a comparison of two composition methods. Remote Sens. 2022;14:1977. https://doi.org/10.3390/rs14091977.

29. Chen Y, Guerschman JP, Cheng Z, Guo L. Remote sensing for vegetation monitoring in carbon capture storage regions: a review. Appl Energy. 2019;240:312–26. https://doi.org/10.1016/j.apenergy.2019.02.027.
30. Abalo M, Badabate D, Fousseni F, Kpérkouma W, Koffi A. Landscape-based analysis of wetlands patterns in the Ogou River basin in Togo (West Africa). Environ Challe. 2021;2: 100013. https://doi.org/10.1016/j.envc.2020.100013.
31. Gebregergs T, Teka K, Taye G, Gidey E, Dikinya O. Impacts of phased–out land restoration programs on vegetation cover change in Eastern Tigray Ethiopia. Environ Syst Res. 2021;10(1):27. https://doi.org/10.1186/s40068-021-00231-7.