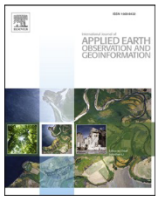
International Journal of Applied Earth Observation and Geoinformation 119 (2023) 103307

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Spectral discrimination of invasive *Lantana camara* L. From co-occurring species

Julius Maina Waititu [a,b,*], Charles Ndegwa Mundia [b], Arthur W. Sichangi [b]

[a] Department of Spatial and Environmental Planning, Kenyatta University, Nairobi, Kenya
[b] Institute of Geomatics, GIS and Remote Sensing, Dedan Kimathi University of Technology, Nyeri, Kenya

ARTICLE INFO

Keywords:
Lantana camara
Invasive species
Species discrimination
Hyperspectral
Sentinel-2
Machine learning algorithms

ABSTRACT

Lantana Camara L. (LC) invasive species has not been successfully mapped due to inadequate spectral information. This study aimed at assessing the performance of leaf-level *in-situ* hyperspectral data and derived indices in discriminating LC among co-occurring species during the dry and wet seasons. In addition, the performance of simulated Sentinel-2 bands, Sentinel-2 derived indices and machine learning algorithms in discriminating it was explored. Spectrally distinct features for species discrimination were selected using the guided regularized random forest (GRRF) and their separability quantified with Jeffries–Matusita distance method. We found that ratio-based and difference indices constructed with first and second-order derivative hyperspectral reflectance wavelengths perfectly separated LC from co-occurring species in the dry and wet seasons with $\geq$ 97% of separability accuracy. Similarly, a set of derived ratio-based and difference Sentinel-2 indices yielded $>$ 95% and $<$ 80% of LC separability accuracy in wet and dry seasons respectively. The SVM with radial basis function algorithm fitted with selected continuum removed derivative reflectance (band depth) narrow-bands yielded the highest overall accuracy (OA) of 84% and a Kappa of 0.75 for the dry season while the same algorithm fitted with selected first derivative narrow-bands yielded an OA of 82% and Kappa of 0.66 for the wet season. Conversely, the regularized logistic regression yielded the highest performance (OA of 77% and Kappa of 0.62) when fitted with combined selected Sentinel-2 variables for the dry season while the gradient boosting machine (GBM) fitted with combined Sentinel-2 variables had the highest performance (OA of 75% and Kappa of 0.51) for the wet season. These findings have important implications on the upscaling of LC's derived leaf-level indices to canopy-level and subsequent LC classification with hyperspectral and Sentinel-2 imagery datasets over heterogeneous environments.

1. Introduction

Establishment of alien invasive species in a given habitat brings changes to the prevailing plant community structures and compositions (Kandwal et al., 2009). *Lantana camara* L. (LC) is one of the global alien invasive species causing habitat transformations in the tropical and subtropical regions where it has been introduced outside its native range of central and northern South America (Shackleton et al., 2017; Taylor & Kumar, 2014; Witt et al., 2018). Regional ecological studies have reported on LC's potential spread across habitats and its negative impacts on local household livelihoods (Shackleton et al., 2017). It's negative impacts may also be exacerbated by changing climatic conditions, for instance in Nyeri County, Kenya where it has potential to spread to new habitats (Waititu et al., 2022). LC's success as an invasive is attributed to its allelopathic activity that lead to suppressed growth of other vegetation (Goncalves et al., 2014; Kandwal et al., 2009; Ruwanza and Shackleton, 2016). Forest ecosystems are particularly prone to LC invasions especially in degraded parts where it can easily establish (Kandwal et al., 2009; Kimothi and Dasari, 2010).

Institutions tasked with monitoring forest ecosystems still lack information on its absolute spatial distribution for effective management. Mapping of the species can be done through remote sensing technology which has become an indispensable tool for mapping of invasive species since mid-1990 s (Huang and Asner, 2009). However, successful detection of species of interest is often hindered by limitations in various remote sensing datasets brought about by tradeoffs made between the sensor resolutions and the ground image coverage (Bradley, 2014). Successful mapping of plant functional types e.g., shrubland and

* Corresponding author at: Department of Spatial and Environmental Planning, Kenyatta University, Nairobi, Kenya.
E-mail addresses: waititu.julius@ku.ac.ke (J.M. Waititu), charles.mundia@dkut.ac.ke (C.N. Mundia), arthur.sichangi@dkut.ac.ke (A.W. Sichangi).

https://doi.org/10.1016/j.jag.2023.103307
Received 21 December 2022; Received in revised form 29 March 2023; Accepted 12 April 2023
Available online 28 April 2023

grassland with medium resolution datasets such as those from Landsat 8 and Sentinel-2 has been achieved (Dube et al., 2020; Ngadze et al., 2020). However, individual plant species separation may require use of their unique spectral signature, textural and phenological characteristics (Bradley, 2014). A high percentage cover of an individual species in an image pixel guarantees its detection in medium resolution images although species spectral uniqueness in relation to the surrounding influences how high the cover should be (Bradley, 2014). A number of studies have reported the usefulness of hyperspectral datasets in species discrimination. For example, Große-Stoltenberg et al. (2016) demonstrated the ability to discriminate *Acacia longifolia* from other species in a Mediterranean dune ecosystems using hyperspectral vegetation indices calculated from field hyperspectral data, while Skowronek et al. (2017) derived distribution maps of invasive bryophyte species, *Campylopus introflexus* using field hyperspectral data and airborne hyperspectral images. It is hypothesized that if a species has distinct phenological phases than co-occurring vegetation, it may provide an opportunity for successful detection (Bradley, 2014; Royimani et al., 2019). Therefore, collection of seasonal hyperspectral data is recommended since spectral response of plants is influenced by changes in their biochemical and eco-physiological parameters during the seasons (Große-Stoltenberg et al., 2016). As such, looking at plant's leaves seasonal spectral reflectance characteristics may provide a basis for their discrimination from one another (Peng et al., 2019).

Although hyperspectral datasets possess high spectral detail for object detection (Royimani et al., 2019) compared to multispectral datasets, hyperspectral imagery datasets are not yet freely available for large scale mapping of invasive species (Ghamisi et al., 2017). Multispectral imagery datasets such as those from Sentinel-2 sensor provide a global coverage and requires no purchasing cost thereby becoming popular alternatives for mapping and monitoring of invasive species e.g., LC (Dube et al., 2020), *Picea abies* and *Prosopis* spp. (Ng et al., 2017), and *Rubus cuneifolius* (Rajah et al., 2019). There are also many envisaged synergies between Sentinel-2 and hyperspectral imagery products for various applications since its bandwidths have been found to cover more than half of the identified useful hyperspectral bands in various vegetation studies (Transon et al., 2018). A number of studies have utilized field spectroscopy data and simulated Sentinel-2 bandwidths to retrieve vegetation spectral characteristics for discrimination e.g., (Mudereri et al., 2020) as well as specific plant biophysical parameters such as LAI e.g., (Mananze et al., 2018). To broaden the limited knowledge of LC detection and mapping, a similar approach can be applied.

As described in literature, successful retrieval of informative spectral information for species discrimination e.g. that of LC, may involve use of hyperspectral vegetation indices (HVIs) (Thenkabail et al., 2013), use of continuum removal of reflectance spectra to enhance the absorptive strengths in the visible range among vegetation types (Schmidt and Skidmore, 2003), use of continuum removal derivative reflectance (Mutanga & Skidmore, 2003) and use of derivative indices to detect vegetation changes such as plants' fluorescence emissions (Zarco-Tejada et al., 2003). In addition, features that have the highest frequency of being statistically different among species need to be selected (Abbasi et al., 2020; Mureriwa et al., 2016; Prospere et al., 2014; Schmidt and Skidmore, 2001). Selected spectral features are used as predictor variables in machine learning algorithms for improved species discrimination (Abbasi et al., 2020; Große-Stoltenberg et al., 2016; Mudereri et al., 2020; Prospere et al., 2014).

Information on LC's separability from other species is still very limited especially within the East African region. An attempt to determine LC's spectral response in a forest habitat was conducted in northeastern New South Wales, Australia by Taylor et al. (2012) where leaf-level hyperspectral reflectance measurements were used to select optimal wavelengths for LC detection as well as evaluating the performance of selected bands in classifying LC with Hyperion imagery datasets. For the first time, our study is seeking to use leaf-level hyperspectral measurements, transformed reflectance data, derived hyperspectral indices and derived Sentinel-2 spectral variables for LC discrimination during the dry and wet seasons. To the best of our knowledge, this approach to discriminating LC has not been examined. The study is guided by three objectives: (a) To assess spectral discrimination of LC among co-occurring species using leaf-level *in-situ* hyperspectral data in dry and wet seasons, (b) To explore the potential of simulated Sentinel-2 wavebands and derived indices in discriminating LC among co-occurring species in both seasons, and (c) To explore the potential of machine learning algorithms in LC discrimination using spectrally significant hyperspectral and Sentinel-2 spectral variables.

2. Materials and methods

2.1. Study area and sampling design

The study was conducted in a one-hectare site located within Muringato forest in Nyeri county, Kenya (0° 24′ 12″ S; 36° 56′ 25″ E) (Fig. 1). The site was identified through purposive sampling based on the abundance of LC. It consists of a large patch of LC intermixed with *Neonotonia wightii* (Wight & Arn.) J.A.Lackey (NW), *Cucumis maderaspatanus* L. (CM) and *Ocimum gratissimum* L. (OG) as identified in the sampling locations. On the sampling site's periphery, to the western and southern direction, large trees including *Croton macrostachyus*, *Cupressus lusitanica* Mill, and *Eucalyptus* spp. were noted but not sampled. LC, our target species, was found in ~ 88% of the total sampled locations followed by NW in ~ 28%, CM in ~ 15%, and OG in ~ 6%.

The site was delineated from Sentinel-2 imagery pixels that were overlaid over the target area (Fig. 1). Five transect lines of 100 m long oriented in an east – west direction and placed at 20 m systematic intervals were used. The use of line-transect sampling ensured minimal destruction of plant species on site as no plot was necessary to establish. In addition, sampling along transect lines ensures capturing of small-scale heterogeneity of the sampling site (Soubry and Guo, 2021). Predefined sampling points' locations along the transect lines were marked on the ground at 8 m systematic intervals using real time kinematic GPS receivers for subsequent *in-situ* individual plant leaves' spectral reflectance measurements. Placing sampling locations at an interval of 8 m between them ensured that sampled plants leaves came from different individual plants. Out of the 65 planned sampling locations, only 53 were sampled as some of them fell on bare ground and in inaccessible locations.

2.2. In-situ hyperspectral reflectance measurement

In-situ leaf-level spectral reflectance measurements were collected on 27th August 2021 and 8th January 2022 during the dry and wet seasons respectively. We applied a set of standard procedures in field spectroscopy described in reviewed literature (Große-Stoltenberg et al., 2016; Jiménez and Díaz-Delgado, 2015). Measurements were conducted between 10:00am and 03:00 pm under sunny conditions. Two portable field spectroradiometers from Apogee Instruments, Inc (https://www.apogeeinstruments.com) with 340 nm-820 nm and 635 nm-1100 nm wavelength ranges were used for spectral reflectance measurements. The instrument's reflectance heads had a 25° field of view, a spectra measurement interval of 1 nm and a wavelength resolution of 3 nm (full width half maximum). At each sampling location, three to five mature leaves were randomly selected by eye based on their overall appearance at the upper canopy (i.e. between 0.05 m − 0.3 m from the top) of target species, detached from the plant, placed on a black cardboard (Carter and Knapp, 2001; Mureriwa et al., 2016) and their upper-side reflectance measurements taken immediately with both spectroradiometers at ~ 5–10 cm above the leaf (depending on leaf size) in a downward-facing position. Leaves' biochemical traits including nitrogen, chlorophyll (a, b), carbon and equivalent water thickness have been found to vary across the different vertical canopy profiles and in turn influence their spectral reflectance (Gara et al., 2018). We chose to analyze the

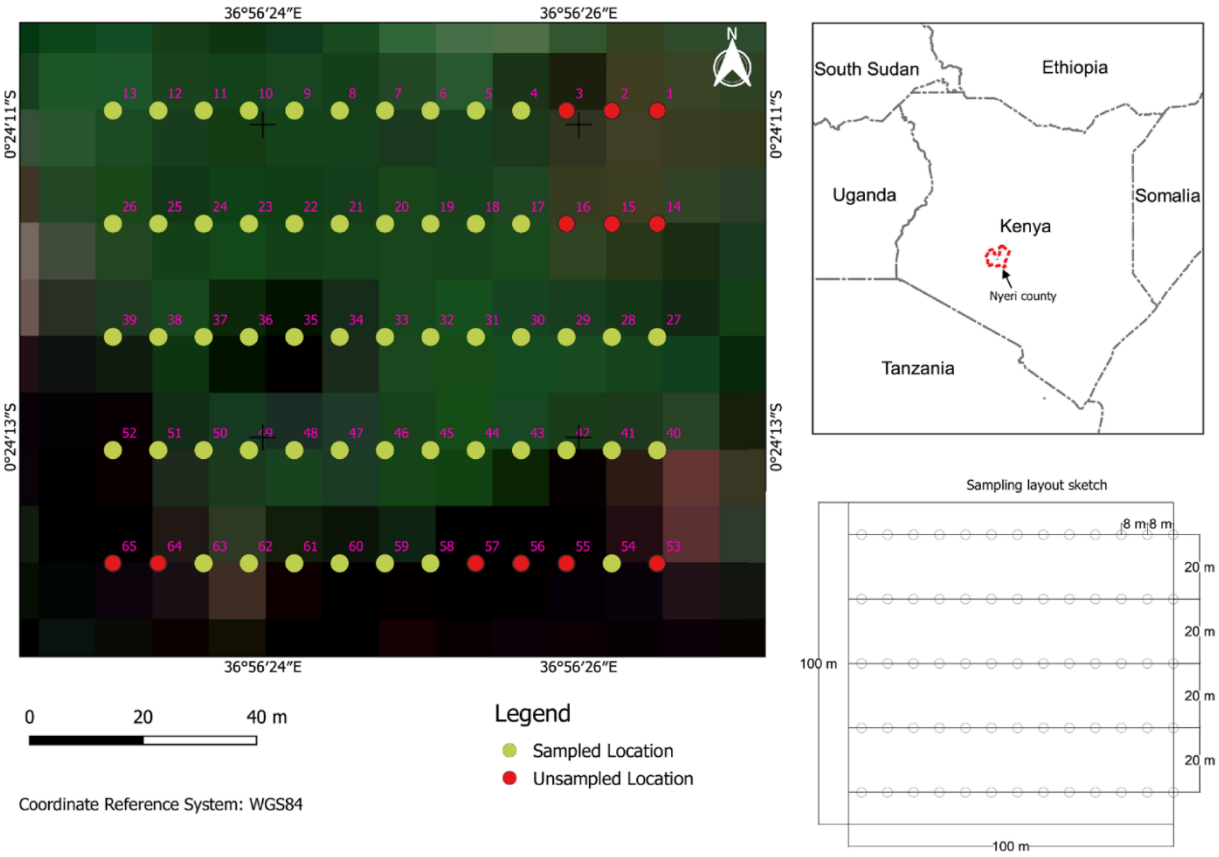


Fig. 1. Site location and sampling layout design. The background image is a true color composite of Sentinel-2 bands 4, 3, and 2 with 10 m × 10 m pixel size acquired on 13th January 2022.

spectra of sunlit top canopy leaves so as to allow for identification of key wavebands for discrimination among the species leaf reflectance corresponding to the top canopy. Before spectral reflectance measurements, the instrument was set to average three measurements of each leaf before recording of the final reflectance measurement. The instruments were regularly calibrated with a white reference panel (Diffuse reflectance target – 99%R) for radiometric corrections during field measurements since illumination conditions in the field change from time to time. A measurement of dark reference using a black cap (provided with the instrument) was taken to correct for internal spectroradiometer dark current noise during calibrations (Jiménez and Díaz-Delgado, 2015).

2.3. Spectra data processing

2.3.1. Data preparation

Spectral reflectance data processing was done using the "*hsdar*" R package (Lehnert et al., 2019) in R Statistical software (R Core Team, 2021). Steps taken in the processing and analysis of reflectance data are illustrated in Fig. 2.

The raw spectra data were filtered using a Savitsky-Golay filter with a length of 25 nm to eliminate random noise in the leaves' spectral response curves arising from lack of full control of the field measurement environment (Prospere et al., 2014). In addition, since hyperspectral sensors suffer from low signal to noise ratios, noise reduction from measured high spectral resolution curves is a critical component (Lehnert et al., 2019). The Savitsky-Golay filter works by fitting a polynomial function that eliminates small differences between neighboring bands (mainly from measurement inaccuracy) (Lehnert et al., 2019). Spectral reflectance narrow-bands below 400 nm and above 1050 nm were also masked and excluded from subsequent analysis due to errors. Spectral Angle Measure (SAM) (Dennison et al., 2004) and Spectral Information Divergence (SID) (Chang, 1999) were used for spectra similarity analysis to identify and remove outliers and by that obtain spectrally similar species spectra necessary for choosing the best species discrimination bands (Jiménez and Díaz-Delgado, 2015). Table 1 gives details of sampled species.

2.3.2. Spectral reference datasets

Seven spectral reflectance datasets namely filtered (R), first derivative (D), second derivative (S), continuum removed reflectance - band depth (CRR_bd), continuum removed reflectance - band depth ratio (CRR_ratio), continuum removed derivative reflectance - band depth (CRDR_bd), and continuum removed derivative reflectance - band depth ratio (CRDR_ratio) were created using the "*hsdar*" R package (Lehnert et al., 2019). Differences in the sun's illumination angles and conditions on the plant materials only affect the reflectance but not the absorption features which are enhanced through transforming spectra by computing derivatives and continuum removal (Erudel et al., 2017; Ferwerda and Skidmore, 2007). Continuum removal normalizes and enhances location of individual absorption features (Clark and Roush, 1984). We calculated the continuum removal using a convex hull fitted over the top of a spectrum, utilizing straight line segments that connect local spectra maxima (Mutanga & Skidmore, 2003; Schmidt & Skidmore, 2001). Continuum removed curves values range from zero to one and shows the enhanced absorption pits.

2.4. Extraction of absorption features

Using "*hsdar*" package (Lehnert et al., 2019), individual absorption features identified in Curran et al, (2001), Mutanga et al. (2005), Mutanga and Skidmore (2004) and Peng et al. (2019) studies were extracted from continuum removed spectra so as to ensure

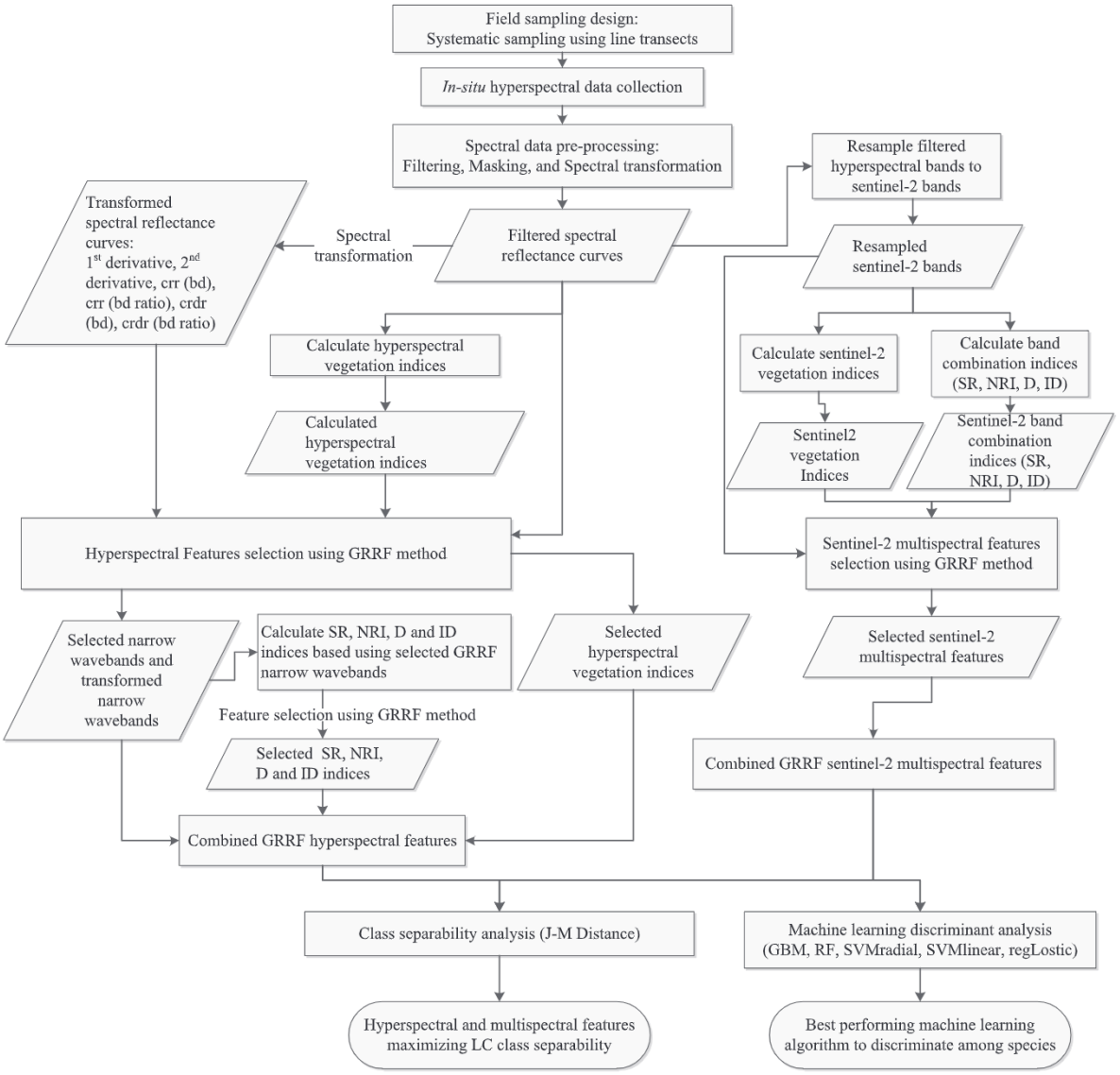


Fig. 2. Overview of methodology.

comparability of individual absorption features among study species as suggested by Mutanga and Skidmore (2003). Individual absorption features were isolated around their center wavelengths to ensure comparability of their absorption character (Lehnert et al., 2019). The size of the areas under the curve of the extracted individual absorption features were extracted for assessment of species characteristics (Meyer and Lehnert, 2018). Statistical differences of the size of absorption features between paired species were determined with a Welch's *t*-test with a significance alpha level $p = 0.05$ and Bonferroni adjustments. We tested the null hypothesis that there was no significant difference in the means of area under the curve of specific absorption features of paired species ($p > 0.05$) versus the alternative hypothesis that there were statistically significant differences between the means of area under curve of specific absorption features of paired species ($p \leq 0.05$).

2.5. Resampling of hyperspectral reflectance to Sentinel-2 sensor wavebands

Using the *spectralResampling*" function in "*hsdar*" package, the filtered spectra were resampled to simulate Sentinel-2 sensor wavebands. Simulation was achieved by using predefined spectral response function of the Sentine-2 sensor (Lehnert et al., 2019). Sentinel-2 wavebands that covered the spectra data wavelength range (340 – 1100 nm) were resampled i.e. Sentinel-2 bands 1 to 9.

2.6. Calculation of hyperspectral and multispectral vegetation indices

Hyperspectral vegetation indices provided in *"hsdar"* package plus additional hyperspectral indices given in Erudel et al. (2017), Liang et al. (2016), and Mudereri et al. (2020) were computed using the filtered spectra dataset. In addition, hyperspectral band ratio-based indices namely simple ratio indices (SR), normalized difference ratio indices (NRI), difference indices (D) and inverse difference indices (ID)) were constructed using GRRF selected narrow-bands and computed using the formulas given in Table 2. Using selected bands ensured that only spectrally distinct bands were considered in the ratio-based and difference indices. Eighty-five Sentinel-2 vegetation indices (VIs) selected from the online index database (https://www.indexdatabase.de) were computed using the resampled Sentinel-2 bands. In addition, all possible Sentinel-2 bands ratio-based and difference indices were computed as in the case of hyperspectral indices described above.

Table 1
Description of study species and the number of sampled spectra for the dry and wet season. Letter (a) represents, 340 – 840 nm and (b), 635 – 1100 nm wavelength ranges. Species photos were obtained during field survey.

Species	Functional group	Species photo	wavelength range (nm)	Season	No. of sampled locations	No. of raw spectra	No. of spectra retained after similarity test
LC	Ornamental shrub		(a)	dry	41	139	116
			(b)		41	139	116
			(a)	wet	47	216	210
			(b)		47	219	212
NW	Climbing herb		(a)	dry	15	66	55
			(b)		14	64	56
			(a)	wet	15	75	75
			(b)		15	75	75
CM	Climber or trailer		(a)	dry	10	44	33
			(b)		11	52	43
			(a)	wet	8	38	38
			(b)		8	39	39
OG	Perennial herb		(a)	dry	3	11	11
			(b)		3	12	12
			(a)	wet	3	12	12
			(b)		3	13	13

Table 2
New ratio-based and difference indices developed for LC discrimination among other species (adopted from Song and Wang (2022)). λ_i and λ_j represent reflectance wavelengths from respective spectral datasets described in section 2.3.2.

Type of index	Formula
Difference Indices (D)	$\lambda_i - \lambda_j$
Simple Ratio Indices (SR)	λ_i / λ_j
Normalized Difference Ratio Indices (NRI)	$(\lambda_i - \lambda_j) / (\lambda_i + \lambda_j)$
Inverse Difference Indices (ID)	$(1 / \lambda_i) - (1 / \lambda_j)$

2.7. Selection of spectrally significant features

Unique spectral features were selected using the GRRF method which was implemented in the "RRF" package (Deng, 2013) and R statistical software. Although other feature selection methods exist in literature, the GRRF method is able to identify features that produce better performance in classification and regression tasks (Izquierdo-Verdiguier and Zurita-Milla, 2020; Mureriwa et al., 2016). GRRF is regarded as an enhanced regularized random forest (RRF) where relative importance scores from ordinary random forest (RF), the mean decrease in Gini Index, is used to guide the feature selection process (Deng and Runger, 2013). The advantage of using the mean decrease in Gini Index is that, the statistic is non-parametric and assumes no theoretical parametric distribution (Cánovas-García and Alonso-Sarría, 2015). The method streamlines the backward elimination procedure for determination of variable importance (Mureriwa et al., 2016) and returns a set of uncorrelated and representative features (Izquierdo-Verdiguier and Zurita-Milla, 2020; Mudereri et al., 2020). Moreover, GRRF does not require *a priori* features to be selected or setting of a threshold of feature importance, but rather, make use of a regularization parameter for feature selection (Izquierdo-Verdiguier and Zurita-Milla, 2020). The regularization parameter (coefReg) depends on the normalized RF feature importance scores (impRF) and a gamma value (regularization constant) as indicated in equation (1) (normalizing RF feature importance score) and equation (2) (weighted average of normalized feature importance score).

$$imp = impRF/(max(impRF)) \quad (1)$$

$$coefReg = (1 - gamma) + gamma^{*}imp \quad (2)$$

Selected gamma values in equation (2) vary between 0 and 1. If values closer to 1 are chosen, higher penalties are executed leading to few features being selected and vice versa. Gamma values chosen in our feature selections varied between 0.5 and 1 so as to ensure at least two or more ($n \geq 2$) features were selected for every spectral reference dataset.

2.8. Species separability analysis

The degree of interclass spectral differences associated with GRRF

selected features were quantified using Jeffries–Matusita (J-M) distance index (Adam and Mutanga, 2009; Cánovas-García and Alonso-Sarría, 2015; Vaiphasa et al., 2005). The J-M distance between a pair of probability functions measures the average distance between the two class density functions (Schmidt and Skidmore, 2003). The J-M distance values for both single-band and multi-band spectra were computed using the "*spectral.separability*" function of the "*spatialEco*" R package developed by Evans and Murphy (2021). The function calculates the JM distance using equations (3) and (4).

$$JM_{i,j} = sqrt(2x(1 - e^{-BH_{i,j}})) \quad (3)$$

$$BH_{i,j} = \frac{1}{8}(\mu_i - \mu_j)^T \left(cov_i^{-1} - cov_j^{-1}\right)^{-1} (\mu_i - \mu_j) - \frac{1}{2} ln\left(\frac{(cov_i + cov_j)/2}{2cov_i^{\frac{1}{2}} cov_j^{\frac{1}{2}}}\right) \quad (4)$$

where $JM_{i,j}$ is the J-M distance between spectra bands *i* and *j*. $BH_{i,j}$ is the Bhattacharyya distance between *i* and *j*, which was calculated using the mean (μ_i) and covariance (cov_i) of the selected band spectra.

The J-M distance values range from 0 as the lower bound and $\sqrt{2}$ (~1.414) as the upper bound. The upper bound value of ~ 1.414 implies complete separability between class pairs i.e. 100% separability accuracy and while decreasing J-M values towards 0 implies diminishing separability (Adam and Mutanga, 2009; Ouyang et al., 2013). The J-M distance value tends to saturate when optimal number of spectral features has been achieved (Adam and Mutanga, 2009; Dalponte et al., 2013). We used a J-M distance value of ≥ 1.3718 (i.e. ≥ 97% separability accuracy) suggested by Adam & Mutanga (2009) to indicate a perfect separability between species class pairs.

2.9. Species discrimination using machine learning algorithms

Five machine learning algorithms namely gradient boosting machine (GBM) (Greenwell et al., 2020), RF (Liaw and Wiener, 2002), SVM with radial basis function kernel (SVMRadial) (Karatzoglou et al., 2022), SVM with linear kernel (SVMLinear) (Karatzoglou et al., 2022), and regularized logistic regression (regLogistic) (Helleputte, 2021) were used for species classification. GRRF selected features were used as predictor variables instead of all of the original variables. Doing so has been shown to improve species classification accuracies (Izquierdo-Verdiguier and Zurita-Milla, 2020; Pal and Foody, 2010). Machine learning classification models were built using "*caret*" Package (Kuhn, 2008) and R statistical software. We used a 10-fold cross-validation repeated 5 times, scaled and centered predictor variables prior to model fitting and used a tune length of 10 to determine model hyperparameters. Other than these parameters, the rest of tuning parameters were retained for all algorithms as provided in the caret package (Kuhn, 2020). All models were fitted using the same set of training spectra dataset (70%) and evaluated using the remaining (30%) spectra dataset. These two sets of balanced spectra datasets were generated randomly within each class using the "*createDataPartition*" function of "*caret*" package (Kuhn, 2020). The random seed number used for the dataset partition was also used for model fitting so as to facilitate model performance comparison. Model performances were assessed with boxplots of overall accuracy (OA) and Kappa statistics. Kappa statistics was appropriate for our imbalance multi class datasets (Kuhn, 2008) and its values were interpreted using a classification scheme provided by Landis and Koch (1977). We assessed whether there were significant differences in performance of machine learning models at 95% confidence interval using McNemar's test (Foody, 2004). McNemar's test was appropriate since the spectra data used to build machine learning classification models did not follow any specific distribution and that classification models were based on similar data samples (Ramezan et al., 2019). The test was implemented in R software by first creating a contingency table from extracted class instance predictions of the 30% test dataset per classification model using the "*extractPredictions*" function in hsdar package. Contingency table was in the form described in Table 3.

3. Results

3.1. Species spectral response

The species mean spectral reflectance curves are given in Figs. 3 and 4. Other than the OG species curves that showed higher reflectance magnitude in the dry season and lower magnitude in the wet season, the rest of the species curves were relatively similar. The species reflectance curves peak in the green–red region (~550 nm) of the spectrum corresponded to the resampled Sentinel-2 band 3 (green) and the depression at ~ 680 nm corresponds to Sentinel-2 band 4 (red) and Sentinel-2 band 5 (Red-edge 1). Reflectance values near ~ 680 nm depression showed distinct variations among the four species during the wet season with the lowest reflectance magnitude being for OG while the highest for NW. These variations in the species reflectance magnitudes during the wet conditions were also consistent in the blue region (Fig. 4a) and the green–red region (Fig. 4b). The spectral reflectance gradients in the red-edge region among the species were not visually different for the wet season although a slight difference in the OG curve was noted (Fig. 4c).

3.2. Characteristics of absorption features

Results of significance test of size of absorption features between species are shown in Figs. 5 and 6. In the dry season, area of absorption features located in the blue region related to chlorophyll absorption were significantly different between pairs of OG vs (LC, NW, CM). Area of absorption features in the green to red-edge regions (495 – 750 nm) were not significantly different between species in the dry season. However, there were significant differences in size of absorption features between LC vs (NW, CM, OG) in the longer wavelength of red-edge (760 – 800 nm; related to cell structure reflectance peak) and in the NIR region (980 – 1000 nm; related to starch content). In the wet season, the absorption feature located in the NIR region (980 – 1000 nm) was significantly different for LC vs (NW, CM, OG). Significant absorption features between LC vs NW were located across the visible (VIS) to NIR wavelength ranges while those for LC vs CM were located only in the NIR region. LC vs OG significant absorption features were located in the red and NIR regions. Areas of absorption features between NW vs OG were significantly different in the red and red-edge regions while for CM vs OG were located in the blue and red regions. Significant absorption features between CM vs NW were located in the blue, green and NIR regions.

3.3. GRRF selected features

3.3.1. Narrow-bands

A graphical representation of selected narrow-bands for discriminating among the study species is shown in Fig. 7. Most of these bands occurred in the VIS wavelength region, a region well known for plant foliar pigment absorption. However, few narrow-bands appeared in the

Table 3
Contingency table format for McNemar's significance test.

		Classifier model 2 (e.g., RF)	
Allocation		Correct	Incorrect
Classifier Model 1 (e.g., GBM)	Correct	True/True	True/False
	Incorrect	False/True	False/False

Where, True/True is the count of test instances that model 1 and model 2 predicted correct, True/False is the count of instances that model 1 predicted correct but model 2 predicted incorrect, False/True is the count of instances that model 1 predicted incorrect but model 2 predicted correct and False/False is the count of instances that both models predicted incorrect.

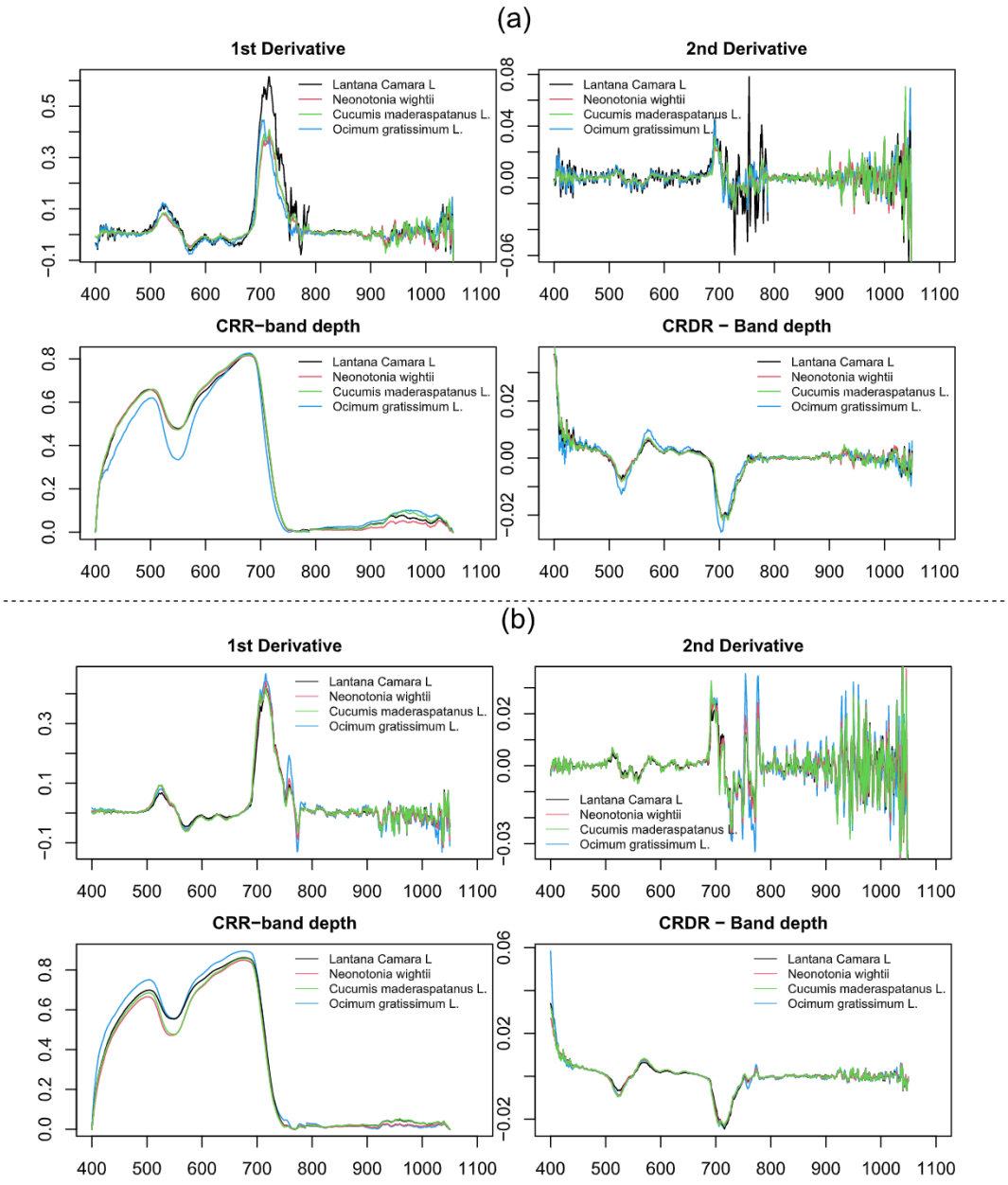


Fig. 3. Transformed spectral reflectance curves for the (a) dry season and (b)wet season.

blue region. Majority of the bands in the dry season were within the red and red-edge regions while in the wet season, were within the green and red-edge regions.

3.3.2. *Hyperspectral vegetation indices*

Out of 171 computed hyperspectral vegetation indices, 4 were selected for dry season and 7 for wet season. These indices were those related to various plant biophysical and biochemical parameters such as chlorophyll; red-edge inflection point (mREIP) and canopy chlorophyll Index (CCI), chlorophyll fluorescence variations; derivative indices (D2), curvature index (CI), photochemical reflectance index (PRI) and curvature index 2 (CI2), leaf water; normalized difference between maxima of first derivatives of reflectance at *dRE* (red edge) and *dG* (green) regions (EGFN) and plant reflectance water index (PWI), and plant stress; double peak index (DPI). A total of 113 and 102 hyperspectral ratio-based and difference indices were also selected for the dry and wet seasons respectively. Ratio-based and difference indices constructed with transformed narrow-bands featured prominently among the ones with the highest variable importance in both seasons (Fig. 8).

3.3.3. *Sentinel-2 multispectral features*

Three of the multispectral bands, bands 2, 3, and 4 were selected in both seasons while bands 5, 6 and 8A were selected in the dry season and bands 5, and 7 in the wet season. Out of 85 Sentinel-2 vegetation indices, 12 were selected in the dry season and 8 in the wet season. From a total of 120 possible Sentinel-2 bands ratio and difference indices, 35 were selected as the most important for the dry season and 40 for the wet season.

3.4. *Separability analysis*

Results of spectral separability analysis of class pairs are shown in Table 4. Indices selected for classes separability were exclusively from

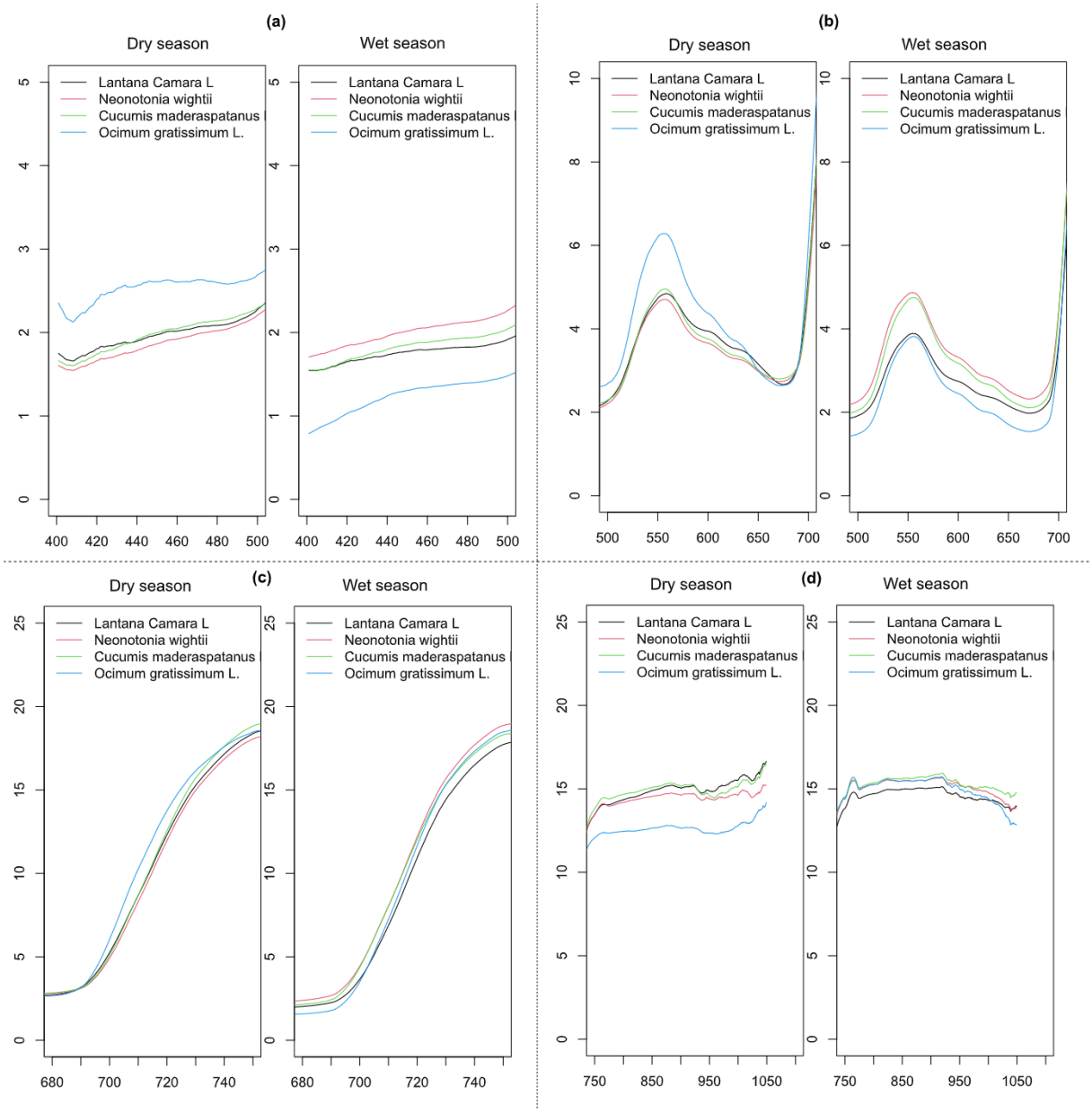


Fig. 4. Spectral reflectance curves of the study species in (a) the blue region (400–500 nm); (b) green–red region (500–700 nm); (c) red-edge region (680–750 nm); and (d) NIR (750–1100 nm). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the newly developed hyperspectral ratio-based and difference indices. These indices consisted of the ones constructed with first and second derivative bands located in green, red and red-edge regions. Perfect separability of class pairs in the wet season was achieved after combination of two or more best performing spectral variables. Narrow-bands located in the green and red-edge regions contributed more to the separability of LC vs NW, CM and OG in both seasons. The red region contributed only to the separation of LC vs CM in the dry season. Separation of LC vs OG required indices combining narrow-bands in NIR, green, red-edge and the blue regions. In addition, separation of LC vs NW and LC vs OG class pairs involved indices combining green, red-edge, blue and NIR region bands in the wet season.

Results of class separability analysis with Sentinel-2 variables for both seasons are shown in Table 5. The variables yielded perfect separability accuracy (≥97%) for the class pairs LC vs OG, NW vs OG and CM vs OG. Although the class pairs LC vs NW, LC vs CM and NW vs CM were not perfectly separable in both seasons, separability accuracy of these classes in the wet season had better separability (>90%) than in the dry season (<85%).

3.5. Species discrimination using machine learning algorithms

Based on the overall accuracy (OA) and Kappa values, the best performing model was the SVMRadial model fitted with selected CRDR_bd narrow-bands (OA of 84% and Kappa of 0.75) using the dry season dataset while the same algorithm fitted with selected first derivative narrow-bands had an OA of 82% and Kappa of 0.66 using the wet season dataset (Fig. 9a, b). The regLogistic model fitted with a combination of all selected Sentinel-2 variables yielded the highest performance (OA of 77% and Kappa of 0.62) followed by SVMRadial model fitted with selected inverse difference indices (OA of 72% and Kappa of 0.49) and selected normalized difference ratio indices (OA of 69% and Kappa of 0.47) using the dry season datasets (Fig. 9c). The GBM model fitted with combined selected Sentinel-2 variables had the highest performance in

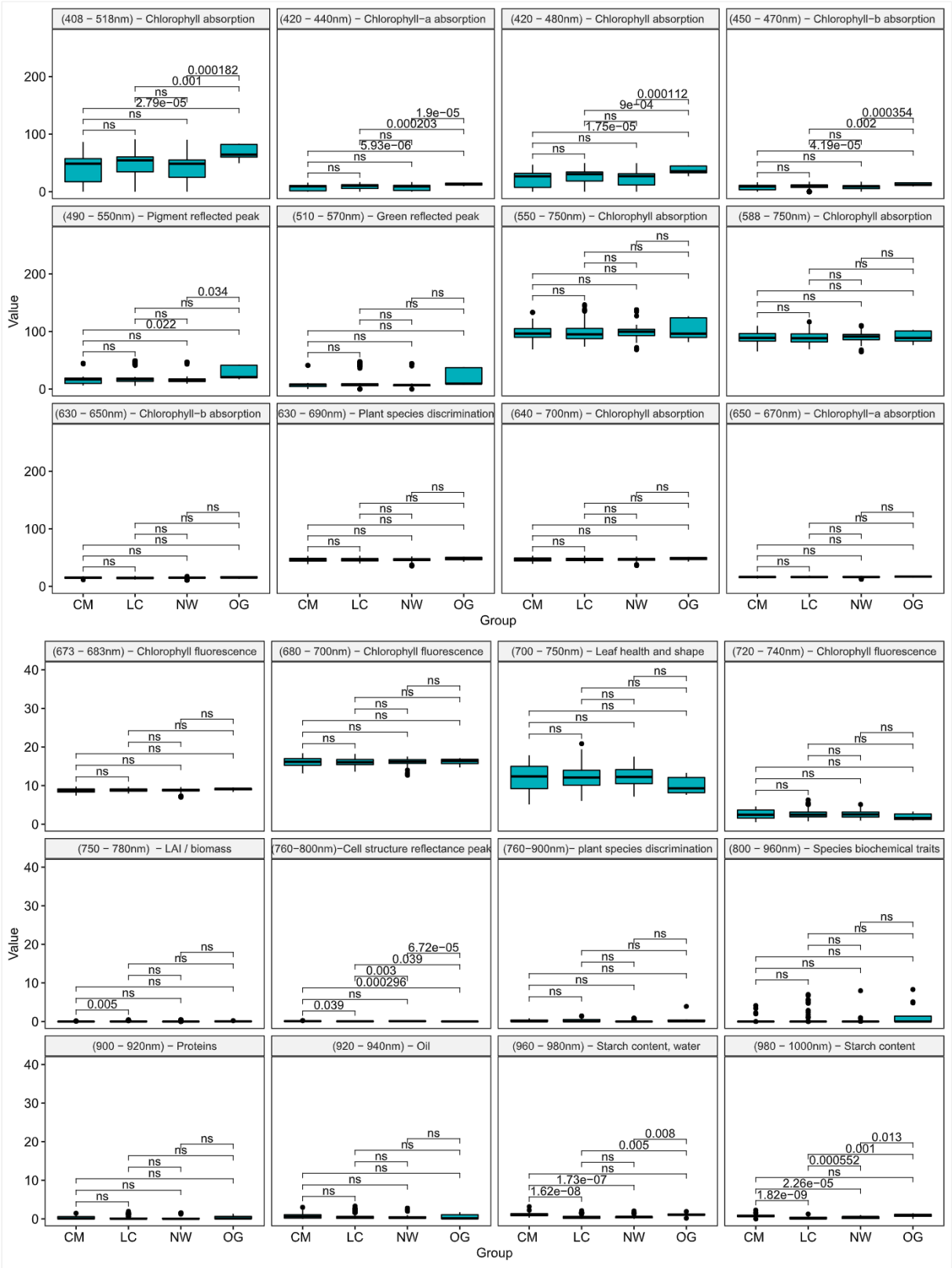


Fig. 5. Significance test of the size of absorption features between species in the dry season.

the wet season (OA of 75% and Kappa of 0.51) followed by GBM model fitted with selected Sentinel-2 normalized ratio difference indices (OA = 74%, Kappa = 0.49) and GBM model fitted with selected Sentinel-2 simple ratio indices (OA = 73%, Kappa = 0.48) (Fig. 9c).

Individual class producer's accuracy (PA) and user's accuracy (UA) shown in Tables 6, indicated that LC class had the highest PA and UA of > 80% in both seasons. The CM class performed better in the dry season (PA:67%, UA:86%) while NW class performed better in the wet season (PA:82%, UA:78%). Although, OG's class UA in the dry season was 100%, both errors of omission and commission were 100% regardless of

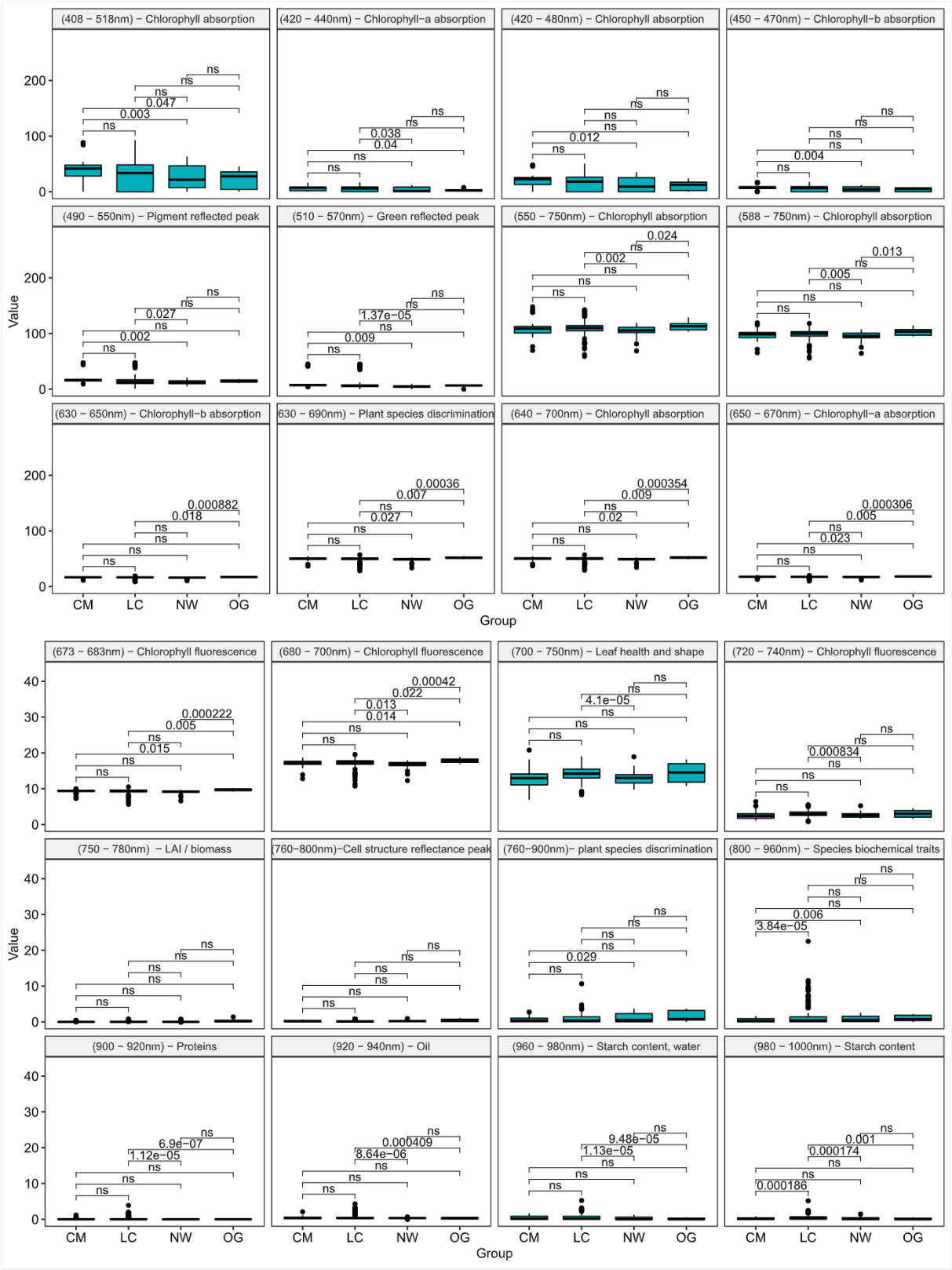


Fig. 6. Significance test of the size of absorption features between species for the wet season.

the classifier used for the wet season.

Performance comparison of paired machine learning algorithms using test hyperspectral datasets showed that all models were not significantly different ($p > 0.05$) from each other except for the SVMRadial vs RF using CRDR_bd narrow-bands ($p < 0.05$) in the dry season and GBM vs regLogistic using combined Sentinel-2 variables in the dry season (Table 7).

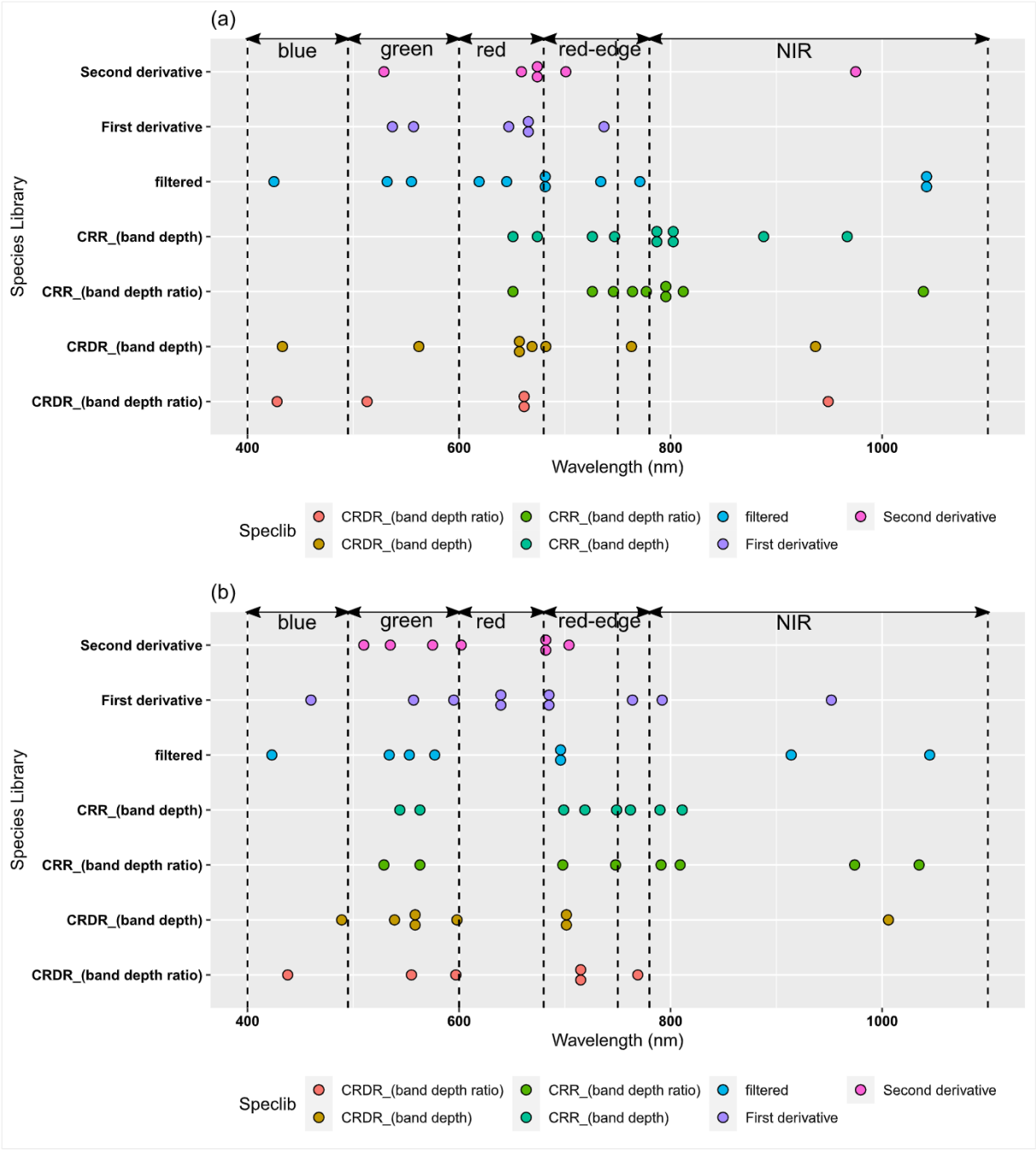


Fig. 7. Graphical representation of selected narrow-bands for the dry season (a) and for the wet season (b) divided into wavelength ranges: blue (400–495 nm), green (495 – 600 nm), red (600 – 680 nm), shorter wavelength of red-edge (680 – 750 nm), longer wavelength of red-edge (750 – 780 nm) and the NIR (780 – 1100 nm). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

A number of studies have successfully demonstrated the use of spectral signatures in discriminating vegetation species (Erudel et al., 2017; Große-Stoltenberg et al., 2016; Prospere et al., 2014; Schmidt and Skidmore, 2001). Usually, where plants' spectral signatures show considerable differences, discrimination among them is possible (Jain and Singh, 2003). In other words, species discrimination is realized where spectral variations between species is greater than that within species (Erudel et al., 2017; Prospere et al., 2014; Schmidt and Skidmore, 2001).

In our current study, we have identified potential spectrally significant features for discriminating LC from co-occurring species using *in-situ* leaf-level hyperspectral datasets and resampled Sentinel-2 multispectral datasets.

4.1. Absorption features

The results of absorption features indicating significant differences between LC vs (NW, CM, OG) in the NIR region (980 – 1000 nm; related to starch content) and in the longer wavelength of red-edge (760 – 800 nm; related to cell structure reflectance peak) indicate possible differences of the arrangement of cells within the LC mesophyll layer versus other species layers (Ollinger, 2011).

The chlorophyll absorption around 470 nm (Curran et al., 2001) explains the significant differences of the absorption features in the blue

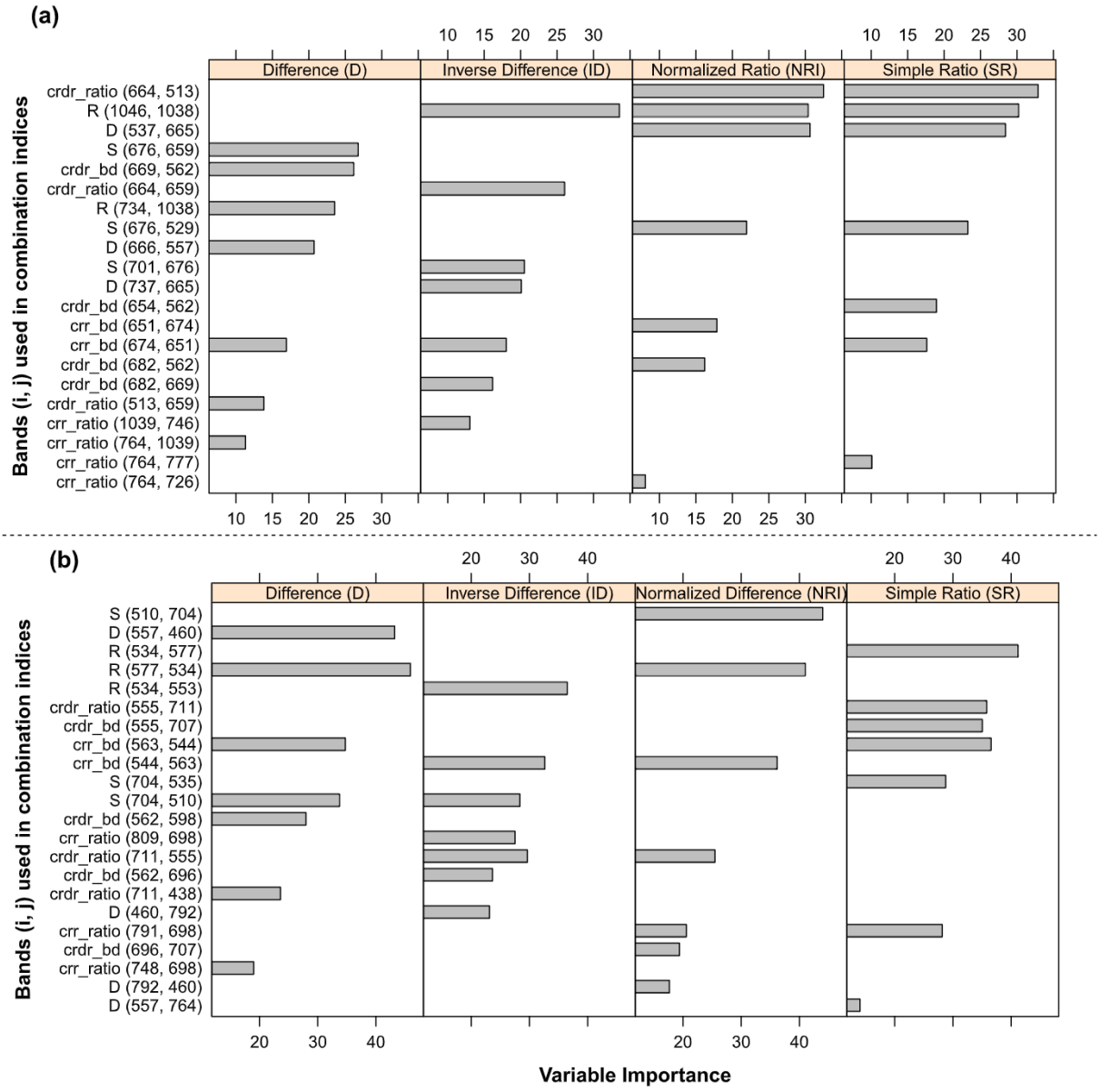


Fig. 8. Hyperspectral ratio-based and difference indices with the highest variable importance score for the dry (a) and wet season (b). The y axis shows hyperspectral bands i and j substituted in the ratio-based and difference indices formulas e.g. in R (1046, 1038), band i is 1046 and band j is 1038. R, filtered reflectance; D, first derivative reflectance; S, second derivative reflectance; crr, continuum removed reflectance; crdr, continuum removed derivative reflectance; bd, band depth; ratio, band depth ratio.

region between OG vs (LC, NW, CM) in dry season. Low chlorophyll in leaves translates to high reflectance in the blue and the red regions and vice versa. The influence of water content in the leaf's spectra measured in wet season is revealed in the identified significant absorption features in the green and red wavelength regions. Wavelengths near green peak (~550 nm) and red-edge (~730 nm) are correlated with leaves water content (Ng et al., 2007).

4.2. GRRF selected features

The spectral differences are usually as a result of varying leaf structural parameters e.g. leaf thickness, leaf pigment content and macronutrient content. Most of the selected spectrally significant hyperspectral bands are particularly in the green and red-edge regions for the wet season and in the red and red-edge regions for the dry season. Although no leaf biochemical properties were determined in this study, it is likely that selected significant wavelengths were as a result of variations of leaf macronutrient contents e.g. nitrogen and other leaf properties such as chlorophyll content and water content among the study species. Mutanga et al. (2004) indicated that there is a direct relationship between absorption bands in the visible region and the nitrogen content in the leaves. The red-edge region is considered important in calculating vegetation indices for vegetation studies (Mudereri et al., 2020; Odindi et al., 2016). Furthermore, the characteristic of red-edge region is said to be related to canopy biomass or LAI, and leaf chlorophyll content (Mutanga & Skidmore, 2004).

4.3. Separability analysis

Significantly different spectral variables selected as potential candidates to discriminate among species are not necessarily variables that maximize the classification between species class pairs (Schmidt &

Table 4
Spectral variables maximizing separability between species class pairs for the dry and wet seasons. Hyperspectral variables highlighted in bold were among those with highest variable importance shown in Fig. 8.

Class pair	Index type	Formula of hyperspectral variable	Wavelength region	J-M value	Separability accuracy (%)	Season
LC vs NW	NRI	$S_{701} - S_{529}/S_{701} + S_{529}$	Green, Red-Edge	1.4000	99.00	Dry
	ID	$1/D_{680} - 1/D_{557}$	Green, Red-Edge	1.3905	98.33	Wet
	ID	$1/D_{460} - 1/D_{792}$	Blue, NIR			
LC vs CM	SR	$\boldsymbol{D_{537}/D_{665}}$	Green, Red	1.3955	98.68	Dry
	ID	$1/\boldsymbol{D_{737}} - 1/\boldsymbol{D_{665}}$	Red-Edge, Red			
	NRI	$\boldsymbol{D_{537} - D_{665}/D_{537} + D_{665}}$	Green, Red			
	ID	$1/D_{680} - 1/D_{557}$	Green, Red-Edge	1.3820	97.72	Wet
	ID	$1/D_{535} - 1/D_{687}$	Green, Red-Edge			
	ID	$1/D_{690} - 1/D_{595}$	Green, Red-Edge			
LC vs OG	Band	CRR_ratio_{812}	NIR	1.4142	100.00	Dry
	NRI	$\boldsymbol{D_{792} - D_{460}/D_{792} + D_{460}}$	Blue, NIR	1.3846	97.90	Wet
	SR	$\boldsymbol{S_{704}/S_{535}}$	Green, Red-Edge			
NW vs CM	NRI	$S_{701} - S_{529}/S_{701} + S_{529}$	Green, Red-Edge	1.4098	99.68	Dry
	NRI	$D_{557} - D_{952}/D_{557} + D_{952}$	Green, NIR	1.3846	97.87	Wet
	NRI	$S_{602} - S_{575}/S_{602} + S_{575}$	Green, Red			
	ID	$1/D_{460} - 1/D_{792}$	Blue, NIR			
	NRI	$\boldsymbol{S_{510} - S_{704}/S_{510} + S_{704}}$	Green, Red-Edge			
NW vs OG	NRI	$S_{701} - S_{529}/S_{701} + S_{529}$	Green, Red-Edge	1.4020	99.14	Dry
	NRI	$\boldsymbol{D_{792} - D_{460}/D_{792} + D_{460}}$	Blue, NIR	1.3844	97.89	Wet
	NRI	$D_{952} - D_{764}/D_{952} + D_{764}$	Red Edge, NIR			
CM vs OG	Band	CRR_ratio_{812}	NIR	1.4142	100.00	Dry
	NRI	$S_{792} - S_{460}/S_{792} + S_{460}$	Blue, NIR	1.3864	98.03	Wet
	NRI	$S_{952} - S_{764}/S_{952} + S_{764}$	Red Edge, NIR			

Table 5
Separability accuracies between class pairs using Sentinel-2 spectral variables. Values in bold indicate separability accuracy of ≥ 97%.

Season	Index type	Sentinel-2 predictor variables	Wavelength region	Class pair	J-M value and Separability accuracy (%)
Dry	NRI	$band_4 - band_3/band_4 + band_3$	Red, Green	LC vs NW	1.1029 (77.99%)
	NRI	$band_6 - band_8/band_6 + band_8$	Red-Edge_2, NIR	LC vs CM	1.1200 (79.19%)
	SR	$band_3/band_4$	Red, Green	LC vs OG	**1.4139 (99.98%)**
	SR	$band_7/band_6$	Red-Edge_2, Red-Edge_3	NW vs CM	1.1202 (79.21%)
	D	$band_4 - band_3$	Red, Green	NW vs OG	**1.4130 (99.92%)**
	D	$band_8 - band_6$	Red-Edge_2, NIR	CM vs OG	**1.4127 (99.89%)**
	ID	$1/band_8 - 1/band_6$	Red-Edge_2, NIR		
	ID	$1/band_8 - 1/band_{8A}$	Nir, Narrow Nir		
	MCARI/MTVI2	$\left(\frac{band_5}{band_4}\right)((band_5 - band_4) - 0.2(band_5 - band_3))$ / $\left(1.5\frac{1.2(band_8 - band_3) - 2.5(band_4 - band_3)}{\sqrt{(2*band_8 + 1)^2 - (6*band_8 - 5\sqrt{band_4}) - 0.5}}\right)$	Green, Red, Red-Edge_1, NIR		
Wet	Band	$band_2$	Blue	LC vs NW	1.3567 (95.94%)
	mSR	$band_8 - band_1/band_4 + band_1$	Coastal Aerosol, Red, NIR	LC vs CM	1.3400 (94.75%)
	NRI	$band_8 - band_7/band_8 + band_7$	Red-Edge_3, NI	LC vs OG	**1.4106 (99.74%)**
	SR	$band_7/band_8$	Red-Edge_3, NI	NW vs CM	1.3157 (93.04%)
	D	$band_8 - band_7$	Red-Edge_3, NIR	NW vs OG	**1.4138 (99.97%)**
	ID	$1/band_4 - 1/band_1$	Coastal Aerosol, Red	CM vs OG	**1.4109 (99.77%)**
	ID	$1/band_7 - 1/band_2$	Blue, Red-Edge_3		
	ID	$1/band_3 - 1/band_5$	Green, Red-Edge_1		
	ID	$1/band_8 - 1/band_7$	Red-Edge_3, NIR		
	SR	$band_9/band_5$	Water Vapor, Red-Edge_1		

*MCARI, Modified Chlorophyll Absorption in Reflectance Index; MTVI2, Modified Triangular Vegetation Index 2; mSR, Modified Simple Ratio.

Skidmore, 2003). Results of the most important variables in Fig. 8 and our separability results presented in Table 4 concurs with this observation. Notably, perfect separability of species class pairs is achieved by hyperspectral indices combining derivative narrow-bands especially those located in the red-edge region. Derivative spectra resolve overlapping species spectra for clear separability features identification (Taylor et al., 2012). Indices selected for perfect separability of LC combines narrow-bands located within the green and red-edge region which are consistent with those selected by Taylor et al. (2012). Hyperspectral indices combining derivative reflectance bands in the red-edge have the ability to detect subtle changes in plants due to fluorescence emissions (Zarco-Tejada et al., 2003). In addition, the usefulness of derivative spectra has been demonstrated in a number of studies e.g., Arun Prasad & Gnanappazham (2014) where they indicated that *Rhizophoracea* mangroves were discriminable with derivative spectra while in the work of Song & Wang (2022), a newly developed difference spectral index ($R_{\lambda 1} - R_{\lambda 2}$) combining first derivative spectra at 540 nm and 1396 nm wavelengths was used to assess the ratio of chlorophyll/carotenoid in cool temperate deciduous forests. Sims and Gamon (2002) also highlighted the advantage of indices combining first derivative spectral wavelengths in estimating leaf pigment properties. In the dry season, the normalized difference ratio indices (S_{701} - S_{529}) / (S_{701} + S_{529}) constructed with second derivative spectral bands located in the green and red-edge regions separates three class pairs namely LC vs NW, NW vs CM and NW vs OG in the dry season. This spectral index is constructed with wavelengths located slightly below (~550 nm) or above the chlorophyll absorbance in the red-edge region (~700 nm) which are correlated with chlorophyll content prediction (Sims and Gamon, 2002). The chlorophyll content between these species during the dry season is therefore likely to be significantly different.

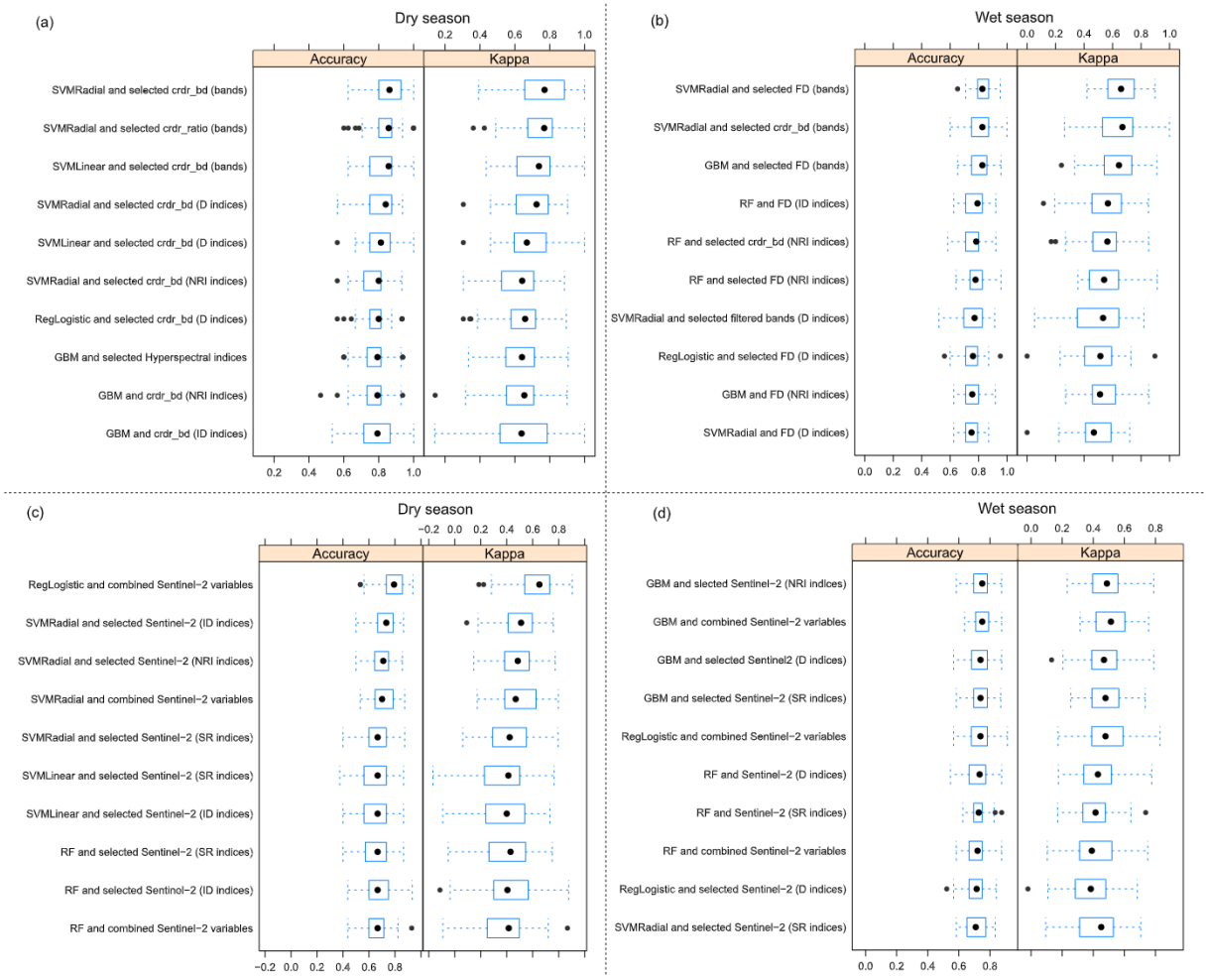


Fig. 9. Model performance of machine learning algorithms and selected hyperspectral (a, b) and Sentinel-2 multispectral (c, d) variables, ordered from the first to the tenth best performing model, for the dry and wet seasons.

Table 6
Model evaluation accuracies obtained from the 30% hyperspectral test datasets. Accuracies of > 70% are highlighted in bold.

Season		Dry						Wet					
Algorithm		SVMRadial		SVMRadial		SVMLinear		SVMRadial		SVMRadial		GBM	
Selected narrow-bands		CRDR_bd		CRDR_ratio		CRDR_bd		1st derivative		CRDR_bd		1st derivative	
		PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
Species	CM	67%	**86%**	56%	**71%**	44%	67%	45%	**71%**	27%	**100%**	55%	67%
	LC	**82%**	**97%**	**88%**	**91%**	**88%**	**94%**	**97%**	**88%**	**94%**	**84%**	**94%**	**87%**
	NW	**88%**	56%	**81%**	68%	**81%**	57%	**82%**	**78%**	**86%**	**76%**	**77%**	**81%**
	OG	33%	**100%**	**100%**	**100%**	33%	**100%**	0%	0%	0%	0%	0%	0%
OA		**79%**		**82%**		**77%**		**85%**		**82%**		**83%**	

Separability between LC vs OG and CM vs OG maximized by the CRR_ratio_{812nm} may signify differences in the light scattering effects of OG leaves at ~ 800 nm as a function of the species leaf internal structures (Sims and Gamon, 2002). In the wet season, it is noted that the index, $(D_{792} - D_{460} / D_{792} + D_{460})$, combining blue-green edge and NIR bands, contains useful spectral information for separability of OG vs LC, NW and CM in the wet season. Such blue normalized difference vegetation index correlates to leaf area index (Lukas et al., 2022). The lack of the same set of predictor variables to separate LC from co-occurring species for both seasons alludes to its seasonal spectral variability. It is therefore important to consider such variations when mapping LC invasions.

4.4. Machine learning discriminant analysis

As expected, our results show higher accuracies in species classification using selected hyperspectral variables compared to the resampled Sentinel-2 variables. In addition, our study shows that performance of machine learning algorithms is directly influenced by the predictor variables used. The svmRadial produces a high performance when fitted with CRDR_bd and 1st derivative narrow-bands while the regLogistic and GBM performs substantially well (Kappa of 0.61 – 0.80) when fitted with combined and Sentinel-2 NRI indices respectively. High performance by svmRadial concurs with findings of Große-Stoltenberg et al. (2016) that showed svmRadial outperformed RF model in distinguishing *Acacia longifolia* from surrounding vegetation at leaf-level. However, the

Table 7
McNemar's test results from comparison of model performance using best performing selected hyperspectral and Sentinel-2 variables. *indicate significant result at 95% confidence interval ($p \leq 0.05$).

			Hyperspectral variables				Sentinel-2 variables			
Season			dry		wet		dry		wet	
Selected predictor variables			CRDR_bd narrow-bands		1st derivative narrow-bands		Combined variables		NRI indices	
Paired models			Chi-square	p-value	Chi-square	p-value	Chi-square	p-value	Chi-square	p-value
SVMRadial	vs	GBM	3.05	0.081	3.273	0.070	1.500	0.221	0.050	0.823
SVMRadial	vs	RF	**4.00**	**0.046***	0.571	0.450	0.000	1.000	0.050	0.823
SVMRadial	vs	SVMLinear	0.00	1.000	0.900	0.343	0.000	1.000	0.842	0.359
SVMRadial	vs	regLogistic	23.05	0.081	0.900	0.343	0.571	0.450	1.389	0.239
GBM	vs	RF	0.17	0.683	0.900	0.343	0.571	0.450	0.000	1.000
GBM	vs	SVMLinear	1.75	0.186	0.267	0.606	0.444	0.505	0.308	0.579
GBM	vs	regLogistic	0.00	1.000	0.267	0.606	**5.143**	**0.023***	0.563	0.453
RF	vs	SVMLinear	2.89	0.089	0.000	1.000	0.000	1.000	0.364	0.546
RF	vs	regLogistic	0.10	0.752	0.000	1.000	1.500	0.221	0.750	0.386
SVMLinear	vs	regLogistic	1.88	0.170	0.000	1.000	2.250	0.134	0.000	1.000

RF model provided optimal results when canopy-level hyperspectral data was used in that study. Both svmRadial and RF algorithms are less sensitive to high dimensional phenomena in hyperspectral datasets (Große-Stoltenberg et al., 2016). Mudereri et al. (2020) found RF and GBM to be the best performing algorithms in classifying striga weed infestation classes using canopy-level hyperspectral and resampled Sentinel-2 data. High classification performance by regLogistic when fitted with combined Sentinel-2 variables can be attributed to the high spectral information provided by the high number of combined Sentinel-2 variables. This suggests that combining important Sentinel-2 predictor variables should be considered when discriminating LC among other species otherwise fewer variables may lack enough spectral information to discriminate it among other species and hence poor performance by machine learning algorithms (Royimani et al., 2019). Although regLogistic has been reported as the best in classifying *cyperus esculentus* among similar weeds using hyperspectral datasets (Lauwers et al., 2020), our study shows it does perform well with Sentinel-2 datasets. McNemar's test show that performance of algorithms used in this study do not significantly differ from one another using both hyperspectral and Sentinel-2 datasets. This suggests on the robustness of these algorithms in classification of LC species from co-occurring species using selected hyperspectral and Sentinel-2 predictor variables.

4.5. Limitations

Although a number of spectrally significant features were developed for perfect separation of LC from co-occurring species in both dry and wet seasons, there were some limitations of this study. It is well known that vegetation reflectance is influenced by several traits including leaf water content, leaf biochemicals and other components of the plant structure (Ollinger, 2011). However, this study was not able to relate the identified spectral features with specific leaf traits. Possible changes in leaf bio-chemicals in the dry and wet season could explain the differences noted in selected features for LC's separation in the two seasons. Furthermore, this study did not assess reflectance properties of other plants' canopy components such as stem/branches. Canopy-level reflectance consists of the plant's stem, branches and leaf distribution and their orientation. However, it is still challenging to determine the leaf orientation in a canopy and instead leaf angle distribution defined for aggregated plant functional groups is often used (Ollinger, 2011). Nevertheless, our interest was on the sunlit upper canopy leaves for LC discrimination. Since we envision mapping of LC with hyperspectral and multispectral imagery datasets as the next step, it would be important to determine whether the leaf-level spectral indices derived in this study can be scaled up to plant's canopy-level and plant's community-level. In addition, it would be important to assess whether the said indices can be replicated in different forest environments.

5. Conclusions

The LC's leaf spectral separability among co-occurring species in dry and wet seasons has not been adequately documented. This study examined *in-situ* leaf-level hyperspectral data and simulated Sentinel-2 wavebands to select spectrally significant features for discrimination of LC among co-occurring species. Findings of this study demonstrate the potential of hyperspectral variables and Sentinel-2 multispectral variables for discrimination of LC among co-occurring species in both dry and wet seasons. Derived LC's separability indices were those constructed with transformed narrow-bands suggesting on the importance of transforming hyperspectral reflectance measurements to enhance detection of subtle leaf spectral variations among plant's species. The fact that different sets of spectral indices were selected for LC's separability during the dry and wet seasons suggests on LC's leaves' seasonal spectral variations. These results are encouraging for future work on LC classification and mapping with multispectral and hyperspectral images. Further research is warranted so as to assess whether the developed leaf-level spectral indices can be scaled up to canopy-level and to the entire plants' community-level in a similar and/or a variety of environments.

Funding

This work was supported by the International Foundation for Science [grant number 1–1-D-6489–1, 2020].

CRediT authorship contribution statement

Julius Maina Waititu: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Funding acquisition. **Charles Ndegwa Mundia:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Arthur W. Sichangi:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Waititu, Julius Maina (2022), "Hyperspectral data - muringato forest sampling site", Mendeley Data, V1, https://doi.org/10.17632/kvnxgy75mm.1.

References

Abbasi, M., Verrelst, J., Mirzaei, M., Marofi, S., Bakhtiari, H.R.R., 2020. Optimal spectral wavelengths for discriminating orchard species using multivariate statistical techniques. Remote Sens. 12, 1–16. https://doi.org/10.3390/RS12010063.

Adam, E., Mutanga, O., 2009. Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry. ISPRS J. Photogramm. Remote Sens. 64, 612–620. https://doi.org/10.1016/j.isprsjprs.2009.04.004.

Arun Prasad, K., Gnanappazham, L., 2014. Species Discrimination of Mangroves using Derivative Spectral Analysis. ISPRS Ann. Photogramm Remote Sens. Spat Inf. Sci. II–8, 45–52. https://doi.org/10.5194/isprsannals-II-8-45-2014.

Bradley, B.A., 2014. Remote detection of invasive plants: A review of spectral, textural and phenological approaches. Biol. Invasions 16, 1411–1425. https://doi.org/10.1007/s10530-013-0578-9.

Cánovas-García, F., Alonso-Sarría, F., 2015. Optimal Combination of Classification Algorithms and Feature Ranking Methods for Object-Based Classification of Submeter Resolution Z/I-Imaging DMC Imagery. Remote Sens. 7, 4651–4677. https://doi.org/10.3390/rs70404651.

Carter, G.A., Knapp, A.K., 2001. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. Am. J. Bot. 88, 677–684. https://doi.org/10.2307/2657068.

Chein-I Chang, 1999. Spectral information divergence for hyperspectral image analysis, in: IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293). IEEE, pp. 509–511. https://doi.org/10.1109/IGARSS.1999.773549.

Clark, R.N., Roush, T.L., 1984. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. J. Geophys. Res. Solid Earth 89, 6329–6340. https://doi.org/10.1029/JB089iB07p06329.

Curran, P.J., Dungan, J.L., Peterson, D.L., 2001. Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: Testing the Kokaly and Clark methodologies. Remote Sens. Environ. 76, 349–359. https://doi.org/10.1016/S0034-4257(01)00182-1.

Dalponte, M., Orka, H.O., Gobakken, T., Gianelle, D., Naesset, E., 2013. Tree Species Classification in Boreal Forests With Hyperspectral Data. IEEE Trans. Geosci. Remote Sens. 51, 2632–2645. https://doi.org/10.1109/TGRS.2012.2216272.

Deng, H., Runger, G., 2013. Gene selection with guided regularized random forest. Pattern Recognit. 46, 3483–3489. https://doi.org/10.1016/j.patcog.2013.05.018.

Deng, H., 2013. Guided Random Forest in the RRF Package, pp. 1–2. http://arxiv.org/abs/1306.0237.

Dennison, P.E., Halligan, K.Q., Roberts, D.A., 2004. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. Remote Sens. Environ. 93, 359–367. https://doi.org/10.1016/j.rse.2004.07.013.

Dube, T., Shoko, C., Sibanda, M., Madileng, P., Maluleke, X.G., Mokwatedi, V.R., Tibane, L., Tshebesebe, T., 2020. Remote Sensing of Invasive Lantana camara (Verbenaceae) in Semiarid Savanna Rangeland Ecosystems of South Africa. Rangel. Ecol. Manag. 73, 411–419. https://doi.org/10.1016/j.rama.2020.01.003.

Erudel, T., Fabre, S., Houet, T., Mazier, F., Briottet, X., 2017. Criteria Comparison for Classifying Peatland Vegetation Types Using In Situ Hyperspectral Measurements. Remote Sens. 9, 748. https://doi.org/10.3390/rs9070748.

Evans, J.S., Murphy, M.A., 2021. spatialEco. R package version 1.3-6. https://github.com/jeffreyevans/spatialEco.

Ferwerda, J.G., Skidmore, A.K., 2007. Can nutrient status of four woody plant species be predicted using field spectrometry? ISPRS J. Photogramm. Remote Sens. 62, 406–414. https://doi.org/10.1016/j.isprsjprs.2007.07.004.

Foody, G.M., 2004. Thematic Map Comparison: Evaluating the statistical significance of differences in classification accuracy. Photogramm. Eng. Remote Sens. 70, 627–633. https://doi.org/10.14358/PERS.70.5.627.

Gara, T.W., Darvishzadeh, R., Skidmore, A.K., Wang, T., 2018. Impact of vertical canopy position on leaf spectral properties and traits across multiple species. Remote Sens. 10, 1–17. https://doi.org/10.3390/rs10020346.

Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., Plaza, A., 2017. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. IEEE Geosci. Remote Sens. Mag. 5, 37–78. https://doi.org/10.1109/MGRS.2017.2762087.

Goncalves, E., Herrera, I., Duarte, M., Bustamante, R.O., Lampo, M., Velásquez, G., Sharma, G.P., García-Rangel, S., Gonzalez-Andujar, J.L., 2014. Global invasion of Lantana camara: Has the climatic niche been conserved across continents? PLoS One 9 (10), e111468.

Greenwell, B., Boehmke, B., Cunningham, J., Developers, G.. gbm: Generalized Boosted Regression Models. R package version 2.1.8. https://cran.r-project.org/package=gbm.

Große-Stoltenberg, A., Hellmann, C., Werner, C., Oldeland, J., Thiele, J., 2016. Evaluation of continuous VNIR-SWIR spectra versus narrowband hyperspectral indices to discriminate the invasive Acacia longifolia within a mediterranean dune ecosystem. Remote Sens. 8, 334. https://doi.org/10.3390/rs8040334.

Helleputte, T., 2021. LiblineaR: Linear Predictive Models Based on the LIBLINEAR C/C++ Library. R package version 2.10-12. https://www.csie.ntu.edu.tw/~cjlin/liblinear/.

Huang, C., Asner, G.P., 2009. Applications of remote sensing to alien invasive plant studies. Sensors (Switzerland) 9, 4869–4889. https://doi.org/10.3390/s90604869.

Izquierdo-Verdiguier, E., Zurita-Milla, R., 2020. An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. Int. J. Appl. Earth Obs. Geoinf. 88, 102051 https://doi.org/10.1016/j.jag.2020.102051.

Jain, S.K., Singh, V.P., 2003. Emerging Techniques for Data Acquisition and Systems Modeling. Dev. Water Sci. 123–205. https://doi.org/10.1016/S0167-5648(03)80057-6.

Jiménez, M., Díaz-Delgado, R., 2015. Towards a Standard Plant Species Spectral Library Protocol for Vegetation Mapping: A Case Study in the Shrubland of Doñana National Park. ISPRS Int. J. Geo-Information 4, 2472–2495. https://doi.org/10.3390/ijgi4042472.

Kandwal, R., Jeganathan, C., Tolpekin, V., Kushwaha, S.P.S., 2009. Discriminating the invasive species', Lantana'using vegetation indices. J. Indian Soc. Remote Sens. 37 (2), 275–290.

Karatzoglou, A., Smola, A., Hornik, K., 2022. kernlab: Kernel-Based Machine Learning Lab. R package version 0.9-31. https://cran.r-project.org/package=kernlab.

Kimothi, M.M., Dasari, A., 2010. Methodology to map the spread of an invasive plant (Lantana camara L.) in forest ecosystems using Indian remote sensing satellite data. Int. J. Remote Sens. 31, 3273–3289. https://doi.org/10.1080/01431160903121126.

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. J. Stat. Softw. 28 https://doi.org/10.18637/jss.v028.i05.

Kuhn, M.. caret: Classification and Regression Training. R package version 6.0-86. https://cran.r-project.org/package=caret.

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, 159. https://doi.org/10.2307/2529310.

Lauwers, M., De Cauwer, B., Nuyttens, D., Cool, S.R., Pieters, J.G., 2020. Hyperspectral Classification of Cyperus esculentus Clones and Morphologically Similar Weeds. Sensors 20, 2504. https://doi.org/10.3390/s20092504.

Lehnert, L.W., Meyer, H., Obermeier, W.A., Silva, B., Regeling, B., Bendix, J., 2019. Hyperspectral Data Analysis in R : The hsdar Package. J. Stat. Softw. 89 https://doi.org/10.18637/jss.v089.i12.

Liang, L., Qin, Z., Zhao, S., Di, L., Zhang, C., Deng, M., Lin, H., Zhang, L., Wang, L., Liu, Z., 2016. Estimating crop chlorophyll content with hyperspectral vegetation indices and the hybrid inversion method. Int. J. Remote Sens. 37, 2923–2949. https://doi.org/10.1080/01431161.2016.1186850.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2 (3), 18–22. https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf.

Lukas, V., Huňady, I., Kintl, A., Mezera, J., Hammerschmiedt, T., Sobotková, J., Brtnický, M., Elbl, J., 2022. Using UAV to Identify the Optimal Vegetation Index for Yield Prediction of Oil Seed Rape (Brassica napus L.) at the Flowering Stage. Remote Sens. 14 (19), 4953.

Mananze, S., Pôças, I., Cunha, M., 2018. Retrieval of Maize Leaf Area Index Using Hyperspectral and Multispectral Data. Remote Sens. 10, 1942. https://doi.org/10.3390/rs10121942.

Meyer, H., Lehnert, L.W., 2018. Introduction to "hsdar". https://cran.r-project.org/web/packages/hsdar/vignettes/Hsdar-intro.pdf.

Mudereri, B.T., Dube, T., Niassy, S., Kimathi, E., Landmann, T., Khan, Z., Abdel-Rahman, E.M., 2020. Is it possible to discern Striga weed (Striga hermonthica) infestation levels in maize agro-ecological systems using in-situ spectroscopy? Int. J. Appl. Earth Obs. Geoinf. 85, 102008 https://doi.org/10.1016/j.jag.2019.102008.

Mureriwa, N., Adam, E., Sahu, A., Tesfamichael, S., 2016. Examining the Spectral Separability of Prosopis glandulosa from Co-Existent Species Using Field Spectral Measurement and Guided Regularized Random Forest. Remote Sens. 8, 144. https://doi.org/10.3390/rs8020144.

Mutanga, O., Skidmore, A.K., 2003. Continuum-removed absorption features estimate tropical savanna grass quality in situ, in: M. Habermeyer, A. Mülle, S.H. (Ed.), Proceedings of the 3rd EARSeL Workshop on Imaging Spectroscopy, Herrsching, Germany, 13-16 May 2003 / Ed. by M. Habermeyer, A. Mülle and S. Holzwarth. EARSeL. Pp. 543-558. EARSeL, Herrsching, Germany, pp. 543–558.

Mutanga, O., Skidmore, A.K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. Int. J. Remote Sens. 25 (19), 3999–4014. https://doi.org/10.1080/01431160310001654923.

Mutanga, O., Skidmore, A.K., Prins, H.H.T., 2004. Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. Remote Sens. Environ. 89, 393–408. https://doi.org/10.1016/j.rse.2003.11.001.

Mutanga, O., Skidmore, A.K., Kumar, L., Ferwerda, J., 2005. Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain. Int. J. Remote Sens. 26, 1093–1108. https://doi.org/10.1080/01431160512331326738.

Ng, C.Q.J., Toh, Y.Y., Lam, C.Y.L., Chang, C.W., Liew, S.C., 2007. Effects of leaf water content on reflectance. In: 28th Asian Conference on Remote Sensing, pp. 631–636.

Ng, W.-T., Rima, P., Einzmann, K., Immitzer, M., Atzberger, C., Eckert, S., 2017. Assessing the Potential of Sentinel-2 and Pléiades Data for the Detection of Prosopis and Vachellia spp. in Kenya. Remote Sens. 9, 74. https://doi.org/10.3390/rs9010074.

Ngadze, F., Mpakairi, K.S., Kavhu, B., Ndaimani, H., Maremba, M.S., 2020. Exploring the utility of Sentinel-2 MSI and Landsat 8 OLI in burned area mapping for a heterogenous savannah landscape. PLoS One 15, e0232962.

Odindi, J., Mutanga, O., Rouget, M., Hlanguza, N., 2016. Mapping alien and indigenous vegetation in the KwaZulu-Natal Sandstone Sourveld using remotely sensed data. Bothalia - African. Biodivers. Conserv. 1–9 https://doi.org/10.4102/abc.v46i2.2103.

Ollinger, S.V., 2011. Sources of variability in canopy reflectance and the convergent properties of plants. New Phytol. 189, 375–394. https://doi.org/10.1111/j.1469-8137.2010.03536.x.

Ouyang, Z.-T., Gao, Y.u., Xie, X., Guo, H.-Q., Zhang, T.-T., Zhao, B., Convertino, M., 2013. Spectral Discrimination of the Invasive Plant Spartina alterniflora at Multiple Phenological Stages in a Saltmarsh Wetland. PLoS One 8 (6), e67315.

Pal, M., Foody, G.M., 2010. Feature Selection for Classification of Hyperspectral Data by SVM. IEEE Trans. Geosci. Remote Sens. 48, 2297–2307. https://doi.org/10.1109/TGRS.2009.2039484.

Peng, Y.u., Fan, M., Bai, L., Sang, W., Feng, J., Zhao, Z., Tao, Z., 2019. Identification of the best hyperspectral indices in estimating plant species richness in sandy grasslands. Remote Sens. 11 (5), 588.

Prospere, K., McLaren, K., Wilson, B., 2014. Plant Species Discrimination in a Tropical Wetland Using In Situ Hyperspectral Data. Remote Sens. 6, 8494–8523. https://doi.org/10.3390/rs6098494.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria. R Foundation for Statistical Computing. version 4.0.5. http://www.r-project.org/.

Rajah, P., Odindi, J., Mutanga, O., Kiala, Z., 2019. The utility of Sentinel-2 Vegetation Indices (VIs) and Sentinel-1 Synthetic Aperture Radar (SAR) for invasive alien species detection and mapping. Nat. Conserv. 35, 41–61. https://doi.org/10.3897/natureconservation.35.29588.

Ramezan, C.A., Warner, T.A., Maxwell, A.E., 2019. Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. Remote Sens. 11 (2), 185.

Royimani, L., Mutanga, O., Odindi, J., Dube, T., Matongera, T.N., 2019. Advancements in satellite remote sensing for mapping and monitoring of alien invasive plant species (AIPs). Phys. Chem. Earth, Parts A/B/C 112, 237–245. https://doi.org/10.1016/j.pce.2018.12.004.

Ruwanza, S., Shackleton, C.M., 2016. Effects of the invasive shrub, Lantana camara, on soil properties in the Eastern Cape South Africa. Weed Biol. Manag. 16, 67–79. https://doi.org/10.1111/wbm.12094.

Schmidt, K.S., Skidmore, A.K., 2001. Exploring spectral discrimination of grass species in African rangelands. Int. J. Remote Sens. 22, 3421–3434. https://doi.org/10.1080/01431160152609245.

Schmidt, K.S., Skidmore, A.K., 2003. Spectral discrimination of vegetation types in a coastal wetland. Remote Sens. Environ. 85, 92–108. https://doi.org/10.1016/S0034-4257(02)00196-7.

Shackleton, R.T., Witt, A.B.R., Aool, W., Pratt, C.F., 2017. Distribution of the invasive alien weed, Lantana camara, and its ecological and livelihood impacts in eastern Africa. African J. Range Forage Sci. 34, 1–11. https://doi.org/10.2989/10220119.2017.1301551.

Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sens. Environ. 81, 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X.

Skowronek, S., Ewald, M., Isermann, M., Van De Kerchove, R., Lenoir, J., Aerts, R., Warrie, J., Hattab, T., Honnay, O., Schmidtlein, S., Rocchini, D., Somers, B., Feilhauer, H., 2017. Mapping an invasive bryophyte species using hyperspectral remote sensing data. Biol. Invasions 19, 239–254. https://doi.org/10.1007/s10530-016-1276-1.

Song, G., Wang, Q., 2022. Developing Hyperspectral Indices for Assessing Seasonal Variations in the Ratio of Chlorophyll to Carotenoid in Deciduous Forests. Remote Sens 14 (6), 1324. https://doi.org/10.3390/rs14061324.

Soubry, I., Guo, X., 2021. Identification of the Optimal Season and Spectral Regions for Shrub Cover Estimation in Grasslands. Sensors 21, 3098. https://doi.org/10.3390/s21093098.

Taylor, S., Kumar, L., 2014. Impacts of climate change on invasive Lantana camara L. distribution in South Africa. African J. Environ. Sci. Technol. 8, 391–400. https://doi.org/10.5897/AJEST2014.1705.

Taylor, S., Kumar, L., Reid, N., Lewis, C.R.G., 2012. Optimal band selection from hyperspectral data for *Lantana camara* discrimination. Int. J. Remote Sens. 33, 5418–5437. https://doi.org/10.1080/01431161.2012.661093.

Thenkabail, P.S., Mariotto, I., Gumma, M.K., Middleton, E.M., Landis, D.R., Huemmrich, K.F., 2013. Selection of Hyperspectral Narrowbands (HNBs) and Composition of Hyperspectral Twoband Vegetation Indices (HVIs) for Biophysical Characterization and Discrimination of Crop Types Using Field Reflectance and Hyperion/EO-1 Data. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 6, 427–439. https://doi.org/10.1109/JSTARS.2013.2252601.

Transon, J., d'Andrimont, R., Maugnard, A., Defourny, P., 2018. Survey of hyperspectral Earth Observation applications from space in the Sentinel-2 context. Remote Sens. 10, 1–32. https://doi.org/10.3390/rs10020157.

Vaiphasa, C., Ongsomwang, S., Vaiphasa, T., Skidmore, A.K., 2005. Tropical mangrove species discrimination using hyperspectral data: A laboratory study. Estuar. Coast. Shelf Sci. 65, 371–379. https://doi.org/10.1016/j.ecss.2005.06.014.

Waititu, J.M., Mundia, C.N., Sichangi, A.W., Koch, F.H., 2022. Assessing distribution changes of selected native and alien invasive plant species under changing climatic conditions in Nyeri County. Kenya. PLoS One 17 (10), e0275360.

Witt, A., Beale, T., van Wilgen, B.W., 2018. An assessment of the distribution and potential ecological impacts of invasive alien plant species in eastern Africa. Trans. R. Soc. South Africa 73, 217–236. https://doi.org/10.1080/0035919X.2018.1529003.

Zarco-Tejada, P., Pushnik, J., Dobrowski, S., Ustin, S., 2003. Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. Remote Sens. Environ. 84, 283–294. https://doi.org/10.1016/S0034-4257(02)00113-X.