JARS Journal of Applied Remote Sensing

RESEARCH PAPER

Discriminating Robusta coffee (*Coffea canephora*) cropping systems using leaf-level hyperspectral data

Getachew Kebede,[a,b,*] Bester Tawona Mudereri,[a,c] Elfatih M. Abdel-Rahman,[a,b] Onisimo Mutanga,[b] Tobias Landmann,[a] John Odindi,[b] Natacha Motisi,[a,d] Fabrice Pinard,[a,d] and Henri E. Z. Tonnang[a,b]

[a]International Centre of Insect Physiology and Ecology (icipe), Nairobi, Kenya
[b]University of KwaZulu-Natal Pietermaritzburg, School of Agricultural, Earth, and Environmental Sciences, South Africa
[c]International Potato Centre, Kacyiru, Kigali, Rwanda
[d]Centre de Coopération Internationale en Recherche Agronomique pour le Développement, UMR PHIM, Nairobi, Kenya

ABSTRACT. The coffee agro-ecosystems are increasingly being transformed into small-scale coffee-growing agricultural systems. In this context, the challenge of accurately classifying coffee cropping systems (CSs) becomes more significant, particularly in regions such as Uganda where dense vegetation and diverse topography complicate traditional land surveys. We harness the capabilities of remote sensing to provide hyperspectral data crucial for distinguishing between various coffee CSs and other land covers. Specifically, we focus on the spectral analysis of three types of Robusta coffee CSs—those integrating agroforestry, those combined with banana cultivation, and those in full sun exposure. Using *in situ* hyperspectral measurements captured by the FieldSpec 2™ spectroradiometer across the 325 to 1075 nm range of the electromagnetic spectrum, we aimed to (1) analyze the unique spectral properties and behaviors of these Robusta coffee CSs and (2) effectively discriminate among them using advanced hyperspectral datasets alongside the machine learning (ML) classification algorithms. The key to this process was the use of narrow spectral bands (NSBs) and various narrow-band vegetation indices (VIs), serving as predictor variables. A selection of critical variables (NSB = 9 and VIs = 8) was identified through the guided regularized random forest (RF) technique and then applied to four ML algorithms—RF, stochastic gradient boosting (GB), linear discriminant analysis, and support vector machine for classification experiments. The findings indicated high discrimination accuracy, with the RF and GB algorithms achieving overall accuracies of 93% and 90.5%, respectively, when using the selected VIs, and 87.3% (RF) and 83% (GB) when applying the chosen NBSs. These results underline the efficacy of integrating hyperspectral datasets and ML algorithms in reliably categorizing Robusta coffee CSs, a crucial step toward enhancing sustainable coffee cultivation practices.

 [DOI: 10.1117/1.JRS.18.044503]

Keywords: variable selection; Uganda; *in situ* hyperspectral data; machine learning; Africa

Paper 240076G received Jan. 30, 2024; revised Aug. 19, 2024; accepted Sep. 5, 2024; published Oct. 4, 2024.

*Address all correspondence to Getachew Kebede, garagaw@icipe.org

1931-3195/2024/$28.00

1 Introduction

Coffee *Coffea* spp. is recognized as an important global commodity, second only to petroleum in terms of income-generating products exported from developing nations. It is the main source of income for 25 million families living in hilly areas of Latin America, Southeast Asia, and East Africa.[1] As a key export commodity, coffee contributes substantially to income generation in developing countries. The global coffee industry however faces challenges in mapping diverse and fragmented coffee land cover due to varied cultivation practices, small-scale cropping systems (CSs), and canopy cover that complicates spatial assessment using satellite sensors.[2] A CS is characterized by the sequence or spatial arrangement of crops.

Uganda is Africa's second-largest coffee producer after Ethiopia, with 1.5 million small-scale growers, making up 10% of the world's coffee farmers.[3] In Ugandan, farmers produce both Arabica, *Coffea arabica* (23%), and Robusta, *Coffea canephora* (77%) coffee, often intercropped with food crops such as banana and forest trees as agroforestry systems (AFSs).[4] In the country, coffee is grown together with a variety of AFS trees and horticultural crops on small holdings. These farms typically consist of various tree species planted by farmers or regenerated naturally.[5] In general, ~60% of coffee producers in Uganda own holdings that are smaller than 0.5 ha[3] and mainly grow Robusta coffee. Thus, the country is dominated by smallholder coffee farmers with a few large-scale producers. Besides, the smallholder farmers largely produce coffee on highly fragmented lands.[6] Coffee productivity and production in the country are also hindered by aging coffee trees, unsustainable land management, pests (insects, diseases, and weeds), drought, and climate changes. The impact of some of these factors depends on the CS.[7]

Studies have shown that remote sensing, utilizing multi-temporal data and supplementary variables is pivotal in identifying agricultural systems for various crops,[6,8,9] including coffee AFSs in different countries.[10–12] Notwithstanding, there is a lack of data on land under coffee CS in Uganda, especially in areas associated with dense crops and tree canopies.[2] Coffee CS can also be difficult to determine due to several factors that include coffee farm or patch layout, the species heterogeneity in and between the coffee farms, the diversity of flora within coffee CS, and the topographic variability. Furthermore, coffee CS is difficult to distinguish from other land cover types due to the spatial heterogeneity of coffee-growing landscapes, the number of vegetation layers, tree density, species arrangement and distribution, the diversity of full sun, and shade-grown coffee spectral characteristics.[2]

In the case of sun-grown coffee plantations, which have few shade trees, a combination of spectral vegetation indices (VIs) and spectral bands of Landsat Thematic Mapper (TM) has achieved a coffee AFS mapping accuracy of 89% to 90%.[13] In situations of intermediate complexity, using high-resolution Cartosat-1 (2.5 m) and Resourcesat LISS-IV-IV multispectral (5.0 m) datasets in areas with homogeneous topography, Hebbar et al.[14] identified commercial poly and monoculture coffee systems with a relatively low classification error (accuracy = 90%). In a more complex AFS, the use of supplementary information, including slope, temperature, precipitation, and soil fertility coupled with Satellite pour l'Observation de la Terre (SPOT 5) imagery, improved the accuracy of the CS identification.[15] In another study, Kelley et al.[16] used spectral indices and land surface temperature derived from multi-seasonal Landsat 8 imagery to detect coffee AFS (with 30% of shade-grown trees) with an accuracy of 82.1% to 80.0%. On the other hand, previous studies have demonstrated that full sun coffee CS can be accurately classified using a combination of reflectance and textural characteristics with an accuracy of ~86%.[17]

Despite the successful application of various remote sensing systems to classify and map various crops,[18–20] including coffee CS in different agroecologies, there is still a dearth of information on Robusta coffee CS in Uganda. As previously mentioned, in the country, coffee is intercropped with other crops such as banana to maximize the land profit. Therefore, information on whether the coffee CS is a full sun or shade is of paramount importance for land use managers and policymakers. Furthermore, in Uganda, the most dominant crop that is intercropped with coffee is banana;[7] hence, it is of interest to know if the shade-grown coffee is intercropped with banana or other tree species (here, we refer to it as AFS). This study discriminates the Robusta coffee in Uganda under various CSs, including full sun without shade trees, Robusta coffee with banana plants, and Robusta coffee with AFS, using spectral assessments of features and vegetation communities that make up the various Robusta CSs.

The leading hypothesis of this study was that *in situ* hyperspectral data collected at the leaf level from representative vegetation communities could distinguish among different categories of Robusta coffee CS. This hypothesis was informed by the fact that many biochemical and physical characteristics of plants, such as pigments, nutrients, water, cell size and structure, and intercellular space, have spectral features that are obscured by broadband multispectral data but can be detectable by hyperspectral data.[17,21] These spectral patterns allow for accurate species and group of species identification.[22] *In situ* data collection also offers the advantage of detecting minor spectral variations not easily discernible from airborne and spaceborne platforms,[23] providing efficient spectral assessments under field conditions.[24]

Despite the benefits of *in situ* hyperspectral and satellite multispectral datasets, these datasets alone may not be adequate for distinguishing among heterogeneous Robusta coffee CS. Combining the magnitude of the spectral details provided by hyperspectral data with the strength of machine learning (ML) algorithms has been successfully applied to many applications.[25–28] This combination can be used to enhance the classification of these complex coffee CS characteristics. Robust ML classification methods such as support vector machines (SVMs),[29] linear discriminant analysis (LDA),[30] gradient boosting (GB),[31] and random forest (RF)[32] address issues of dimensionality and multicollinearity in hyperspectral datasets, yielding accurate and relevant input predictors to the feature of interest (e.g., coffee CS).[33] However, one of the prominent problems in hyperspectral data processing and analysis is the dimensionality and multicollinearity inherent in the data.[33] Multicollinearity associated with a small number of training samples (n) relative to a large number of hyperspectral variables (p) is a common cause of poor predictive model performance.[27,28] The guided regularized random forest (GRRF) has shown to be a successful method in reducing the dimensionality of the hyperspectral data and simultaneously handles the multicollinearity in the dataset by selecting a few, yet relevant predictor variables.[34,35]

However, previous studies have shown no consensus on the most effective ML classification method or the most effective dimension reduction technique for distinguishing among complex and diverse land cover types.[30,31] This study extends beyond the visible spectrum to utilize and analyze the reflective characteristics of diverse elements within coffee CS. It assumes that the three forms of Robusta coffee CS, i.e., (i) Robusta coffee with AFS, (ii) Robusta coffee with banana, and (iii) Robusta coffee with full sun, can be distinguished based on their species composition and specific biochemical and physical attributes. Specifically, this study aims to (1) investigate the spectral uniqueness of the three Robusta coffee CSs and (2) discriminate among these CSs using relevant hyperspectral datasets and ML classification algorithms.

2 Methodology

2.1 Study Site

The study was carried out in the main Robusta coffee-growing districts of central, eastern, western, and southwestern Uganda. The study site is situated at an altitude between 612 and 5239 m above sea level, with coordinates of latitudes 1° 32′ N and 1° 21′ S and longitudes 29° 31′ W and 34° 27′ E (Fig. 1). The tropical climate at the study site is characterized by an annual bimodal rainfall model. The temperature range is relatively higher, from 18°C to 22°C,[36] and the overall amount of annual precipitation is 750 to 1500 mm.[3]

The main growing areas for Robusta coffee cover 94,076 km^2 and are divided into 34 districts.[3] The Robusta coffee CS cropland combined with and without trees, grassland, water bodies, and built-up areas dominates the agro-natural ecosystem in the study site, whereas the agricultural activities are primarily subsistence and small-scale farming, including AFS (timber and fruit trees), and intercrops with horticultural crops such as banana, cassava, and yam are included in the Robusta coffee CSs. Climate change, pests, and disease are the key factors limiting Robusta coffee production at the study site.

2.2 Robusta Coffee Cropping System Characterization

Field surveys were conducted between January 28, 2023, and February 10, 2023, to record leaf-level spectral signatures of Robusta coffee CSs and bare soil. At the study site, 60 plots were sampled in four main Robusta coffee locations (15 samples from each location,

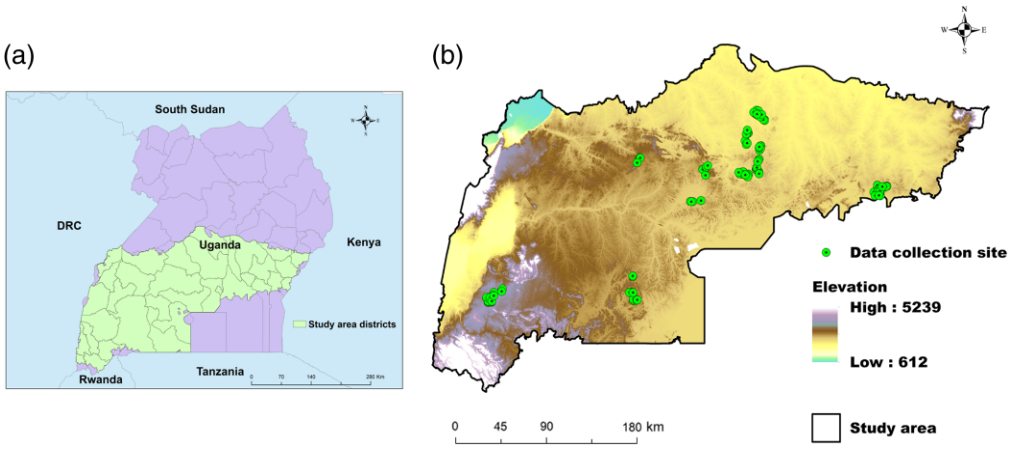

Fig. 1 Location of the main Robusta coffee-growing districts ($n = 34$) in Uganda (a) and the distribution of all Robusta coffee CSs sample plots ($n = 60$) on a 30 m resolution digital elevation model (b) obtained from the United States Geological Survey (USGS).

i.e., central, eastern, western, and southwestern parts of the study site). Different plant species that co-existed with Robusta coffee CS were also sampled for spectral data collection.

The sampled Robusta coffee CSs were selected based on the abundance (count) of Robusta coffee, AFS, and banana (*Musa hybrid*) plants. Specifically, the three target Robusta coffee CSs were (i) Robusta coffee with AFS, where tall (18 m on average) timber and fruit trees co-exist with the coffee crop. In this CS, coffee is an understory layer. The primary timber trees in this CS included *Grevillea robusta*, *Albizia adianthifolia*, *Ficus natalensis*, *Maesopsis eminii*, and *Markhamia lutea*, whereas the fruit trees were *Artocarpus heterophyllus*, *Mangifera indica*, *Persea americana*, *Carica papaya*, and *Citrus reticulata*. This CS also included some shrubs such as *Hibiscus syriacus*, (ii) Robusta coffee intercropped with banana. This CS was dominated by Robusta coffee and banana. In addition, intercrops of short fruit trees and vegetables were also co-existing as an understory layer. The main short fruits and intercrops in this category included *Citrus sinensis*, *Theobroma cacao*, *Manihot esculenta*, *Dioscorea bulbifera*, *Zea mays L.*, *Ipomoea batatas*, and (iii) Robusta coffee cultivated under full sun conditions without shade trees layer. However, some grass species, specifically *Megathyrsus maximus*, were found in the understory along with bare soils. Details of each species and Robusta coffee CS are given in Tables 1 and 2.

2.3 Spectral Data Collection

The leaf reflectance spectra for the Robusta coffee CSs were collected using a portable FieldSpec Handheld 2™ Spectroradiometer.[37] The spectroradiometer is a non-imaging sensor that measures electromagnetic radiation within a range of 325 to 1075 nm and a 25-deg full conical angle field of view.[37] We measured three leaf spectra per species (20 plant species) and five samples of bare soil after optimizing and calibrating the measured radiance using a white reference panel made of spectralon material (~100% reflectance). This material was used to

Table 1 Description of the three Robusta coffee cropping systems (CSs).

S. no	Robusta coffee CS	Description
2	Robusta coffee with AFS	Robusta coffee plantations primarily comprise AFS trees, predominantly timber, and fruit trees
3	Robusta coffee with banana	The majority of Robusta coffee plantations are characterized by the presence of banana species and the absence of any AFS trees
4	Robusta coffee full sun	Robusta coffee plantations devoid of any AFS trees and banana and solely exposed to direct sunlight

Table 2 Co-existing species/classes with the three Robusta coffee CS

Robusta coffee CS	User label	Botanical name/class	Local name	Code
Robusta coffee with AFS	1	*Grevillea robusta*	Grevillea	GR
		Albizia adianthifolia	Albizia	AA
		Ficus natalensis	Ficus	FN
		Maesopsis eminii	Musizi	ME
		Markhamia lutea	Musambya	ML
		Artocarpus heterophyllus	Jackfruit	AH
		Mangifera indica	Mango	MI
		Hibiscus syriacus	Mukuge	HS
		Persea americana	Avocado	PA
		Carica papaya	Papaya	CP
		Citrus reticulata	Tangerine	CR
Robusta coffee with banana	2	*Musa Hybrid*	Banana	MH
		Citrus sinensis	Orange	SS
		Theobroma cacao	Cocoa tree	TC
		Manihot esculenta	Cassava	MEE
		Dioscorea bulbifera	Yam	DB
		Zea mays L.	Maize	ZM
		Ipomoea batatas	Sweet potato	IB
Robusta coffee full sun	3	*C. canephora*	Robusta coffee	CC
		Bare soil (BS)	Bare soils	BS
		Megathyrsus maximus	Grass	MM

calibrate the spectroradiometer before taking the actual readings or when the instrument was saturated. These calibrations were done every 20 readings to preserve consistent spectral data and correct for solar light and weather fluctuations.[38] To ensure measurement accuracy, errors were minimized by taking multiple readings with the spectroradiometer positioned 5 to 7 cm above the leaf adaxial surface (upper), depending on the leaf size of different plant species,[22] and a total of 1260 spectral measurements were collected. All measurements were taken under sunny conditions between 10:00 a.m. and 2:00 p.m. local time (Greenwich Mean Time: GMT +3), and each leaf spectrum was sampled three times to reduce handling errors and improve data accuracy.[22] Furthermore, the spectral measurements were filtered using the "noiseFiltering" function and smoothened using the "Savitzky–Golay" filter with a window size of 3×3 in the "hsdar" package [39] in R software.[40] These preprocessing steps help to reduce the noise in the data to improve the model predictability.[39] A flowchart depicting the overall process is presented in Fig. 2.

2.4 Calculation of Spectral Vegetation Indices

In situ hyperspectral data can detect subtle spectral shifts not apparent in airborne- or spaceborne-based data.[23] Huang et al.[51] proposed that *in situ* hyperspectral sensors can swiftly measure ground target spectral properties and construct band indices, revealing intricate physical and biological aspects of plants. Narrow-banded spectral vegetation indices (NVI) that combine two or more hyperspectral bands mimic the fine vegetative properties such as pigment concentration or leaf water content and accurately discriminate among different vegetation species and CS.[35,52] This study calculates 12 NVIs in Table 3 to compare their performance with the narrow spectral

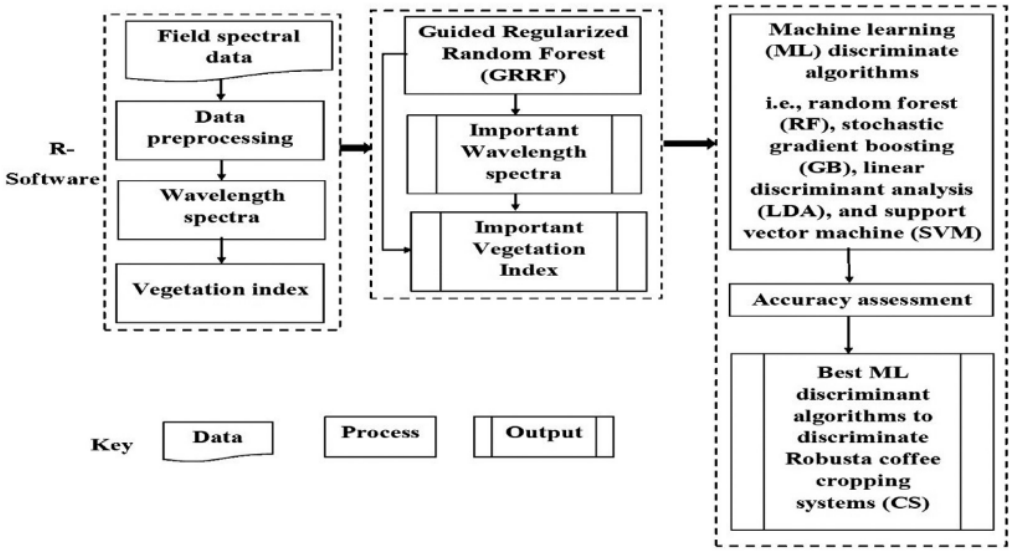

Fig. 2 Flowchart of the methodology adopted for data preprocessing, variable selection, and evaluation of the performance of the ML discriminate algorithms (RF, random forest; LDA, linear discriminant analysis; GB, gradient boosting; and SVM, support vector machines).

Table 3 Hyperspectral NVIs used in this study.

S.no	NVI	Equation	Significance	Reference
1	GI = greenness index	B554/B677	Indicator of prolonged vegetation stress due to changes in canopy structure	41
2	EVI = enhanced vegetation index	2.5 × ((B800 − B670) / (B800 − (6 × B670) − (7.5 × B475) + 1)	Indicator or biomass and leaf area index (LAI)	42
3	WI =Water index	(B900) / (B970)	Indicator of vegetation water status	43
4	REP = Red-edge position	700 + 40 (B670+ B780) / 2 − B700 / (B740 − B700)	Indicator of sharp change in vegetation reflectance	44
5	LCI = Leaf chlorophyll index	(B850 − B710) / (B850 + B680)	Indicator of total chlorophyll content	45
6	MSRI = Modified simple ratio index	(B800 − B445) / (B680 − B445)	Significant indicator of chlorophyll	46
7	VREI = Vogelmann red-edge index	(B734 − B747) / (B715 − B726)	Chlorophyll concentration, canopy leaf area, and water content	47
8	SRPI = Simple ratio pigment index	(B430) / (B680)	Carotenoid/chlorophyll-a content	43
9	Narrow-banded NDVI = Normalized difference vegetation index	(B830 − B670) / (B830 + B670)	Canopy greenness, LAI, a fraction of photo synthetically active radiation	48
10	GMI = Gitelson and Merzylak index	(B750) / (B700)	Leaf chlorophyll content	49
11	PSRI = plant senescing reflectance index	(B678 − B500) / B750	Leaf senescence	50
12	PRI= Photochemical reflectance index	(B531 − B570) / (B531 + B570)	Conversion of xanthophyll-cycle pigments, photosynthetic light use efficiency, LAI	43

B is used for reflectance at a specific band in nm.

bands (NSB) in discriminating Robusta coffee CSs. The "hsdar" package[39] in R software[40] was used to construct the 12 NVIs. Specifically, these 12 NVIs and NSB were used as predictor variables to distinguish among the different Robusta coffee CS.

2.5 Selection of the Predictor Variables

There are several feature selection methods available, including the regularized random forest (RRF), which was proposed as a method using a single ensemble approach.[35] Unlike the multiple ensemble approaches, the RRF uses only the information from one node, hence has a feature representation problem.[54,55] Also, RRF evaluates the features based on a subset of the training data at each tree node, potentially leading to a greedy selection process.[56] To address these problems, the GRRF was proposed to use the importance scores of the ordinary RF to guide the feature selection process. Hence, it penalizes the gain information of each variable in relation to the response variable to guide the variable selection process.[34,57] The regularization reserves the gain and reduces the time required for training the model.[35] It also helps in ensuring the selection of non-correlated and representative variables. The significance of a variable in RF is determined by calculating the "Gini index" across all nodes in all decision trees generated within the RF ensemble. This variable is then utilized to assess the purity of the feature at each node, aiding in the decision-making process of the RF trees.[32] In this study, we therefore used the GRRF to select a few yet relevant hyperspectral features for classifying the three Robusta coffee CSs. We limited the selection of the hyperspectral predictor variables (NSB or VIs) in the GRRF experiment using an optimal gamma value of 0.5. Moreover, all the selected variables were centered and rescaled for consistency before they were used to discriminate among the three Robusta coffee CSs. We utilized an ordinary RRF package[54] in R software to perform the GRRF experiment. Despite its effectiveness and efficiency as a feature selection method, GRRF is not a good model for prediction;[57] hence, we used other ML algorithms for discriminating among the Robusta coffee CSs.

2.6 Machine Learning Discriminant Algorithms

In this study, ML discriminant models, including RF, GB, LDA, and SVM were used to distinguish among the three Robusta coffee CSs using the selected NSB and VIs. These classification algorithms were selected due to their proven effectiveness in accurately discriminating vegetation-related classes using hyperspectral datasets.[58,59] The RF algorithm builds multiple decision trees (*ntree*) from bootstrapped samples, avoiding overfitting, working well against noisy data, requires minimal training time, and is suitable for both normally and non-normally distributed datasets.[32,60] The algorithm assigns class labels based on the majority votes from all *ntree*, which are differently built using different features (*mtry*) at each point. The GB enhances the prediction of the classes but may face scalability issues compared with RF, particularly with datasets of numerous classes.[31] The algorithm needs a setting of two major parameters, that is, the shrinkage value and the number of boosting bags. The shrinkage parameter ranges between 0 and 1 to reduce the overfitting. On the other hand, LDA aims to reduce the dimensionality in the data while maximizing the class discrimination power using optimum shrinkage and solver values.[61] The SVM excels in handling overfitting, especially in high-dimensional feature spaces, and effectively uses kernel functions for nonlinear data separation.[56] However, selecting the appropriate kernel function remains a challenge.[62] The SVM parameters that need to be optimized are kernel type and regularization parameter (C). Notably, these ML algorithms do not require traditional regression assumptions, enhancing their utility in various application scenarios.[63] A k-fold ($k = 10$) cross-validation method[64,65] was employed to fine-tune and optimize the algorisms hyperparameters to reduce the overfitting and to enhance the overall algorithm performance. To ensure consistency, the tune length parameter of the four ML algorithms was set to 10, allowing the assessment of 10 values for each parameter (e.g., the number of trees for the RF algorithm). In addition, all variables were centered and rescaled before the classification experiments to maintain consistency across analyses. To accomplish this, we utilized the "Caret" package[66] in R software[40] to train the four ML algorithms using 70% of the dataset ($n = 110$ for Robusta coffee with AFS, $n = 71$ for the Robusta coffee with banana, and $n = 30$ for Robusta coffee full sun) classification models. This package provides a common syntax for various ML approaches. Table 4 shows the "Caret" packages that were employed to run the RF, SVM, LDA, and GB algorithms and their parameter values.

Table 4 Packages in "Caret" that were used in R software for the four discriminant algorithms and their parameters that were optimized in this study. *ntree* is the number of trees, and *mtry* is the number of variables at each split in the RF algorithm, kernel type, and regularization parameter (*C*) in the SVM algorithm; shrinkage and solver in the LDA algorithm; and shrinkage and the number of boosting bags in the GB algorithm.

Algorithm	Caret code	Package	Reference	Optimum parameter
RF	"rf"	Ranger	64	*ntree* = 500 and *mtry* = 3
SVM	"svmRadial"	Kernlab	62	Kernel = linear and C = 0.5
LDA	"lda"	Mass	65	Solver = singular value decomposition (*svd*) and shrinkage = 0.5
Stochastic GB	"gbm"	gbm and plyr	67	Shrinkage = 0.5 and number of boosting bags = 500

2.7 Validation of Machine Learning Discriminant Algorithms

To validate the performance of the ML algorithms, we used the same independent 30% test dataset that was held out to determine the overall and individual class accuracies. Specifically, we assessed the accuracy of correctly classified samples within each Robusta coffee CS category, i.e., the producer's accuracy (PA); the proportion of correctly classified samples for a specific Robusta coffee CS, i.e., the user's accuracy (UA); and the accuracy of correctly classified samples among all samples of the Robusta coffee CS, i.e. the overall accuracy (OA). The models' inter-class prediction performances were evaluated using confusion matrices from the best predictor variables for each algorithm. The McNemar test, at a 95% confidence interval (CI), compared the performance of the four models in discriminating the three Robusta coffee CSs using GRRF-selected variables considered in this study to compare the performance of RF, SVM, LDA, and GB algorithms for discriminating the Robusta coffee CS categories. This was done for the selected NSB and VIS.

3 Results

3.1 Robusta Coffee Cropping Systems and Co-existing Species Spectral Profiles

The results in Fig. 3 show the mean reflectance of the three Robusta coffee CSs, namely, Robusta coffee with AFS, Robusta coffee with banana, and Robusta coffee in full sun. These CSs demonstrate a high level of distinguishability across the wavelength range of 713 to 1075 nm. Specifically, Robusta coffee in full sun shows a distinct difference between 760 and 1075 nm, whereas Robusta coffee with AFS and Robusta coffee with banana exhibit specificity within the red-edge and near-infrared (NIR) region.

The spectral profiles of Robusta coffee CS classes and their co-existing species were displayed according to the average leaf spectra in Fig. 4. The spectra showed typical vegetation spectral profiles for all plant species and soil spectral characteristics for bare soil class. The average reflectance values for the individual co-existing species in each category of Robusta coffee CS are illustrated in Fig. 4. The figure shows the spectral regions with high visual discriminatory power for each CS and its co-existing species.

3.2 Variable Selection to Discriminate Robusta Coffee Cropping Systems

The GRRF identified bands 602, 539, 525, 522, and 598 nm in the visible region and the 758, 761, 757, and 786 nm bands in the NIR region as the most relevant wavelengths for distinguishing the three types of Robusta coffee CS in Fig. 5(a). In addition, red edge position (REP), visible red edge index (VREI), greenness and moisture stress index (GMI), modified simple ratio index (MSRI), greenness index (GI), photochemical reflectance index (PRI), water index (WI), and

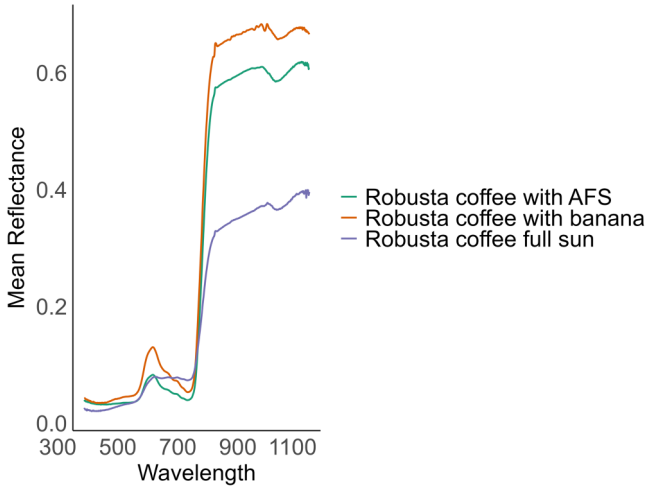

Fig. 3 Mean reflectance of the three Robusta coffee CSs, i.e., Robusta coffee with agroforestry (AFS), Robusta coffee with banana, and Robusta coffee full sun.

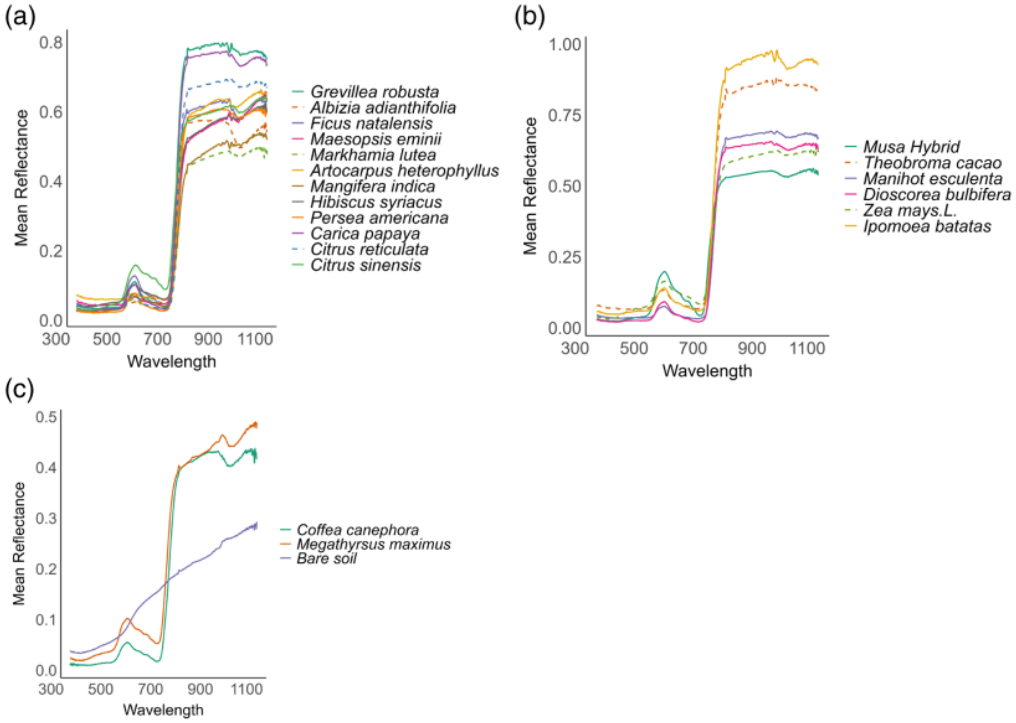

Fig. 4 Mean reflectance of co-existing species across the three Robusta coffee CSs, i.e., Robusta coffee with agroforestry (AFS) (a), Robusta coffee with banana (b), and Robusta coffee full sun (c). The spectral features that indicate a high level of distinguishability for each group of Robusta coffee CS or co-existing species are highlighted.

simple ratio pigment index (SRPI) were the most significant NVIs in discriminating among the three categories of Robusta coffee CS in Fig. 5(b).

3.3 Robusta Coffee Cropping System Discrimination Using the Four Machine Learning Discriminant Algorithms

The results in Table 5 show that the performance of various algorithms using the selected NSB varies depending on the Robusta coffee CS categories. For example, RF and GB exhibited good

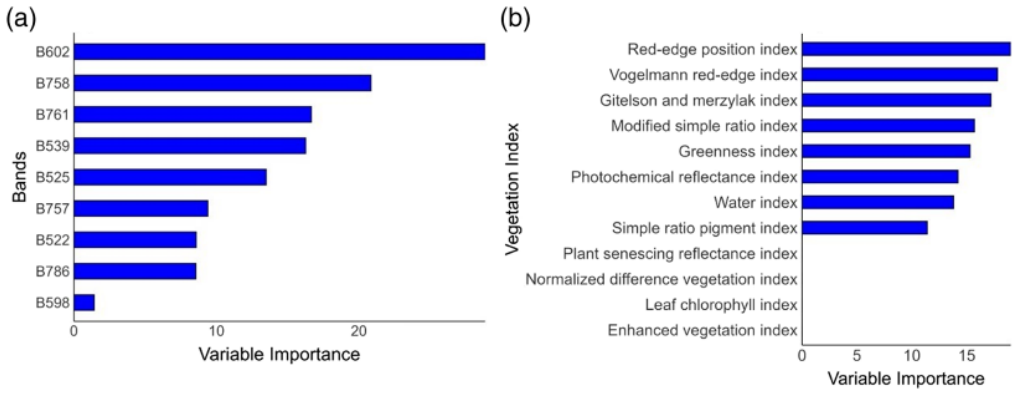

Fig. 5 Selected wavelength spectra (a) and (b) NVIs to discriminate Robusta coffee CSs using GRRF.

Table 5 Classification accuracy matrices for the four ML algorithms using the selected NSB. RF, GB, LDA, and SVM are = random forest, stochastic GB, LDA, and SVM, respectively. AFS = agro-forestry PA is the producer's accuracy, UA is the user's accuracy, and OA is overall accuracy.

	ML algorithm							
	RF		GB		LDA		SVM	
CS	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
Robusta coffee with AFS	94.4	85.0	90.0	88.0	84.0	80.0	66.0	89.2
Robusta coffee with banana	87.5	87.5	87.5	94.7	66.6	62.5	50.0	25.0
Robusta coffee full sun	63.6	100	88.0	80.0	60.0	85.7	100	29.0
OA (%)	87.3	—	83	—	77.8	—	69.8	—

performance in discriminating Robusta coffee with AFS, whereas LDA and SVM had lower PA (%) and UA (%) values for this type of Robusta coffee CS. Similarly, for Robusta coffee with banana, RF and GB had higher PA (87.5% for both RF and GB) and UA (87.5% for RF and 94.7% for GB) than LDA and SVM. For full sun, SVM had the highest PA of 100%. In general, RF and GB performed relatively better across the three Robusta coffee CS categories, whereas LDA and SVM performed better for Robusta coffee with AFS and Robusta coffee full sun CS categories, respectively. In general, RF was the top-performing ML algorithm for discriminating most of the samples of Robusta coffee CSs as indicated by an OA of 87%.

When the selected VIs and the four ML algorithms were employed, the results indicated that Robusta coffee with AFS can be accurately discriminated from other Robusta coffee CSs with a class accuracy that ranged from 80% to 95% in Table 6. For Robusta coffee with banana, RF had the highest PA and UA values, whereas LDA had the lowest. The SVM had relatively higher UA but a lower PA, indicating a potential imbalance between the correctly classified Robusta coffee with banana and the three Robusta coffee CSs. For Robusta coffee full sun, RF, GB, and SVM had perfect PA of 100%, whereas SVM had the lowest UA of 14.2%. Overall, RF and GB were the best algorithms for classifying the three Robusta coffee CSs, whereas SVM had the lowest performance in most cases. On the other hand, LDA had mixed results. Overall, although RF and GB generally perform better overall (especially in OA and high PA/UA in most cases), LDA and SVM showed more variability and lower performance in specific contexts.

The performance of the four ML classification models in discriminating the three Robusta coffee CSs was significantly different ($p \leq 0.05$) from one another in most cases (Table 7). The findings from the pairwise McNemar test (Table 7) illustrate that the performance of RF was significantly ($p \leq 0.05$) different from the performance of the other three classification models

Table 6 Classification accuracy matrices for the four ML algorithms using the selected VIs. RF, GB, LDA, and SVM are = random forest, stochastic GB, LDA, and SVM, respectively. AFS = agroforestry PA is the producer's accuracy, UA is the user's accuracy, and OA is overall accuracy.

	ML algorithm							
	RF		GB		LDA		SVM	
CS	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
Robusta coffee with AFS	92.7	95.0	84.8	93.3	89.7	87.5	80.0	80.0
Robusta coffee with banana	87.5	97.5	90.0	82.0	63.0	75.0	50.0	75.0
Robusta coffee full sun	100	85.0	100	90.0	80.0	57.0	100	14.2
OA (%)	93	—	90.5	—	85.7	—	79.3	—

Table 7 McNemar test for comparing the performance of the four ML discriminant algorithms (RF, GB, LDA, and SVM) in discriminating the three Robusta coffee CS classes using the selected NSB and NVIs.

	McNemar test results			
	Selected NSB		Selected VIs	
Comparison	Chi-square	*p*-Value	Chi-square	*p*-Value
RF versus LDA	8.10	0.004**	2.77	0.006**
RF versus GB	7.23	0.040*	6.77	0.048*
RF versus SVM	4.00	0.046*	1.13	0.289^{NS}
LDA versus GB	9.09	0.003**	3.50	0.061^{NS}
LDA versus SVM	0.50	0.480^{NS}	0.57	0.045*
GB versus SVM	3.27	0.070^{NS}	1.78	0.182^{NS}

*Significant at 95%.CI
**Significant at 99% CI; NS = not significant.

when selected NSB was used. For the other models, there was inconsistency in the significant differences in their performance ($p \leq 0.05$).

4 Discussion

This study examined the efficacy of *in situ* spectroscopic data in differentiating Robusta coffee CSs in Uganda. Using the GRRF algorithm, the study pinpointed specific VIs and NSB as critical predictors for discriminating between diverse Robusta coffee CSs. This analysis, conducted over a single observational period, underscores the value of precise spectral data in agricultural mapping and management. The ability to accurately discriminate coffee CSs offers a comprehensive view of their spatial distribution and characteristics, a key element in promoting sustainable and efficient coffee production. This detailed insight benefits a wide range of stakeholders, including farmers, policymakers, and market participants, by facilitating informed decision-making, optimized resource allocation, and improved market access. Furthermore, it enhances disease and pest monitoring, thereby contributing to healthier crop ecosystems and environmental sustainability. Hyperspectral remote sensing, renowned for its capacity to distinguish between plant species based on their spectral properties, is a pivotal tool in this context.[68] Its effectiveness has been demonstrated in previous studies, which employed both field-based instruments and

hyperspectral satellite imagery to identify and map the presence of alien species amidst native flora.[22,69] This study adds to the growing body of knowledge, showcasing the potential of hyperspectral technology in the nuanced field of agricultural remote sensing, particularly within the realm of coffee production.

Field spectroscopy has been commonly used to evaluate the biophysical and biochemical properties of different plant species.[70,71] In the context of this study, specific VIs and NSB were successfully employed to discriminate the three Robusta coffee CSs. This discriminatory ability can be attributed to the inherent differences in plant characteristics as explained by Aneece and Epstein.[69] These differences include variations in leaf pigments, intercellular spaces, water content, cell wall thickness, cell size, and other structural and biochemical features that are unique to each plant species.

The study's findings highlight the significance of specific hyperspectral wavebands in differentiating among the three Robusta coffee CSs. The most crucial wavebands identified were five in the visible spectrum and four in the NIR region. A key finding was the importance of the red-edge region in the electromagnetic spectrum (EMS) for discriminating plant species. This prominence of the red-edge region in hyperspectral data can be primarily attributed to the unique spectral signatures of chlorophyll, nitrogen, phosphorus, and potassium in Robusta coffee and its coexisting plants. These elements exhibit distinguishable characteristics in the red-edge spectral region,[48,72] making it a reliable indicator for differentiating plant species.

Furthermore, variations in moisture content among plant species are closely linked to the red-edge region of the EMS.[73] Consistent with previous studies, the study corroborates that leaf spectra of plants show the most significant variation in the NIR and red-edge regions.[73,74] These findings enhance the understanding of plant spectral properties and underscore the utility of hyperspectral remote sensing in agricultural applications, particularly in the context of coffee CS.

The study underscored that the ability to identify plant species using spectral signatures depends on their spectral heterogeneity and phenological changes.[75] Key wavelengths in the visible (490, 520, 550, 575, 660, and 675 nm) and red-edge zones are crucial for differentiating tropical tree species, including coffee plants, as they reflect variations in light absorption related to plant biochemistry and structure.[73] This finding is significant for hyperspectral remote sensing in tropical forestry and agriculture, enhancing species identification and monitoring.

To distinguish the three Robusta coffee CSs categories, the study identified the REP, VREI, GMI, and MSRI as key VIs. The REP proved particularly effective, capitalizing on variations in leaf color, intercellular gaps, water contents, cell wall thickness, cell size, and other plant traits.[35] The MSRI is indicative of prolonged chlorophyll stress in the canopy structure.[76] Leaf stress can also be associated with variations in prolonged chlorophyll stress and was also significant in differentiating coffee CSs, reflecting canopy health and developmental stages. These indices demonstrate the utility of hyperspectral data in discerning agricultural systems, aiding sustainable farming management.

The selection of the optimal classifier for a given application involving the use of remote sensing data is contingent upon the choice of an appropriate accuracy measure and the specific objectives of the analysis.[77] In this study, the RF and GB classifiers emerged as the most effective models for classifying Robusta coffee CSs. Their effectiveness was determined based on overall accuracy metrics. These results highlight the suitability and effectiveness of these ML algorithms in accurately classifying complex agricultural systems using hyperspectral remote sensing data.

The relatively lower performance of the SVM classifier in this study could be attributed to the use of a default setting, particularly the linear hyperplane, and standard SVM parameters gamma (γ) and sigma (C). These default parameters might not have been optimal for capturing the complex, nonlinear relationship between the Robusta coffee CSs at varying wavelengths and indices.[59,77] The findings of this study align with previous research that employed hyperspectral data at the leaf or canopy levels, along with one of the classifiers (RF, GB, or SVM) to identify plant characteristics.[52] The two non-linear classification methods, namely GB and RF, demonstrated superior performance outcomes when using the variables determined by the GRRF method. This was observed for the chosen hyperspectral wavebands. This enhanced performance was evident in the context of the selected hyperspectral wavebands, underscoring the

effectiveness of GB and RF in dealing with complex, non-linear data patterns typical in hyperspectral remote sensing applications.

The efficiency of RF, GB, and SVM in this investigation is consistent with prior research findings using leaf- or canopy-level hyperspectral data to detect different plant physiochemical characteristics.[52,78] For both NSB and VIs, the two nonlinear classification methods, RF and GB, performed better when GRRF-chosen variables were used. These two classification algorithms are less susceptible to overfitting and perform accurately on imbalanced datasets.[26] In addition, this study resonates with Mureriwa et al.[34] who employed the GRRF and RF algorithms to identify *Prosopis juliflora* utilizing field spectral measurements data. They observed that accuracy in detection was enhanced when the redundancy of spectral variables was minimized. In the context of discriminating Robusta coffee CSs, this research similarly found that the combination of the RF model and GRRF algorithm yielded the most accurate results. The study evaluated four different discriminant analysis algorithms to ascertain the most effective method for differentiating within the Robusta coffee CSs. Among these, the integration of the RF model with GRRF stood out as the most optimal choice. This combination proved to be highly effective, irrespective of the various dimensions presented by the predictor variables, the number of observations, or the scale of mapping involved. These findings underscore the robustness and versatility of the RF and GRRF models in handling complex remote sensing data for agricultural applications. The success in accurately differentiating various categories within Robusta coffee CSs demonstrates the potential of these models in enhancing precision agriculture, offering valuable insights for sustainable coffee production and land management strategies.

The novelty in this study rests on the assumption that the three forms of Robusta coffee CSs (i.e., Robusta coffee with AFS, Robusta coffee with banana, and Robusta coffee full sun) distinguish themselves based on their species composition and specific biochemical and physical attributes, including pigmentation, nutritional composition, and water content. Moreover, the results of this study could be integrated with satellite-based datasets using multiple endmember spectral mixture analysis (MESMA) to map the Robusta coffee CSs at a landscape scale. Previous studies have integrated Sentinel-2 imagery with *in situ* spectroscopic data and MESMA and demonstrated that such an approach could facilitate large-scale identification and mapping of specific agricultural challenges such as Striga weed infestation in maize farms.[56] This multifaceted approach, leveraging both ground-based and satellite-based remote sensing techniques, offers a comprehensive method for assessing and monitoring agricultural land.

Also, our study approach is unique as we utilized a spectral-rich dataset and a very robust and efficient variable selection method (i.e., GRRF) to reduce the dimensionality of such spectral data by ~98%. We selected nine spectral bands out of 750 and compared the performance of four ML algorithms in discriminating among three unique coffee CSs. Previous studies such as Mosomtai et al.[17] and Moreira et al.[79] have used multispectral data to map coffee as a generic land use/land cover class. It is interesting to note that, unlike other studies, we have separated the full sun coffee into a mono-CS (i.e., Robusta coffee full sun class) and coffee intercropped with fruit crops such as banana (i.e., Robusta coffee with banana class). Studies such as Sabat-Tomala[80] performed parametric classification methods such as the maximum likelihood and the broadband multispectral data to map shaded and full sun coffee. However, such parametric methods can easily overfit and handle only normally distributed datasets. Besides, the broadband data of some remote sensing systems such as Landsat might not be able to distinguish among different coffee AFSs and their surrounding understories, or co-existing plant species.

The differences in Robusta coffee discriminating accuracies and limitations observed in this study can be attributed to several factors. One significant factor is the background (soil) effect[80] and atmospheric noise[81] that might hinder the scalability of our models to different points in time and space. Furthermore, other environmental, climatic, and agronomic conditions such as sunlight intensity and angle, temperature, soil moisture, wind speed and direction, crop age, and health, etc. might also vary from one space to another or from one season to another and affect the replicability of our models to other areas/ regions, CSs, and seasons. Hence, our model results should be interpreted with some caution and tested in different environmental conditions and CSs.

5 Conclusions

This study examined *in situ* hyperspectral data and VIs to distinguish different Robusta coffee CSs in Uganda's Robusta coffee-growing districts. The study results showed that GRRF can be effectively used for variable selection of hyperspectral data, VIs, and multispectral bands. Robusta coffee CSs could be distinguished using five visible and four NIR bands. This work has improved our understanding of the spectral features that best distinguishes Robusta coffee CS categories. The REP, VREI, GMI, and MSRI were better for discriminating the three Robusta coffee CS categories. The study also showed that the RF and GB classifiers were better at discriminating between Robusta coffee CSs utilizing the selected NSB and VIs. Nonetheless, there is a necessity to map the Robusta coffee CSs using high spatial resolution multispectral data. Sentinel-2 data, when combined with MESMA, a technique that discerns spectra within image pixels by identifying the percentage contribution of each CS with more than one end member, could be explored for mapping Robusta coffee CSs on a large scale.

This approach would significantly enhance the mapping of Robusta coffee CSs and contribute to the overall landscape assessment of the coffee status. The outcomes of this research hold paramount importance in effectively distinguishing heterogeneous Robusta coffee CSs in Sub-Saharan Africa. Although the utilization of field hyperspectral data in vegetation studies is not novel, our findings underscore the capabilities and practical applications of such remotely sensed data as a valuable tool for accurately discriminating Robusta coffee CSs. These results open avenues for researchers to employ a similar methodology in precision mapping of Robusta coffee CS using various platforms, including spaceborne multispectral satellite sensors, airborne systems, or unmanned aerial vehicles, which commonly provide broad-band data, for a comprehensive characterization of Robusta coffee CSs classification at localized scales.

Disclosures

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.

Code and Data Availability

The data that support the findings of this study are available in https://dmmg.icipe.org/dataportal/dataset/discriminating-robusta-coffee-coffea-canephora-cropping-systems.

Acknowledgments

Getachew Kebede was supported by the In-Region Postgraduate Scholarship from the German Academic Exchange Service (DAAD). The authors gratefully acknowledge the financial support for this research by the European Union (EU), project "Robusta coffee agroforestry to adapt and mitigate climate change in Uganda" GCCA+-Global Climate Change Alliance & DESIRA (Project/Grant No. FOOD/2021/427-759); the Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); the Australian Centre for International Agricultural Research (ACIAR); the Norwegian Agency for Development Cooperation (Norad); the Federal Democratic Republic of Ethiopia; and the Government of the Republic of Kenya. The project also received some funds from the Centre de Coopération Internationale en Recherche Agronomique pourle Développement (CIRAD). We also acknowledge the support provided by Tony Mugoya and Seklbamu Joseph from Uganda Coffee Farmers Alliance (UCFA) during field surveys and data collection in the main Robusta coffee-growing districts of Uganda.

References

1. M. Oxfam, "Poverty in your coffee cup," *Bost. Oxfam Am.* (2002).
2. D. A. Hunt et al., "Review of remote sensing methods to map coffee production systems," *Remote Sens.* **12**, 1–23 (2020).
3. C. Bunn et al., "Climate-smart coffee in Uganda," *Feed Futur.* 1–24 (2019).
4. H. Bukomeko et al., "Integrating local knowledge with tree diversity analyses to optimize on-farm tree species composition for ecosystem service delivery in coffee agroforestry systems of Uganda," *Agrofor. Syst.* **93**, 755–770 (2019).

5. W. J. Negawo and D. N. Beyene, "The role of coffee based agroforestry system in tree diversity conservation in Eastern Uganda," *J. Landsc. Ecol. Republic* **10**, 5–18 (2017).
6. D. Niyogi et al., "Evapotranspiration climatology of Indiana using *in situ* and remotely sensed products," *J. Appl. Meteorol. Climatol.* **59**, 2093–2111 (2020).
7. Uganda Coffee Development Authority (UCDA). Robusta coffee handbook: a sustainable coffee industry with high stakeholder value for social economic transformation. 2–148 (2019).
8. C. Atzberger, "Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs," *Remote Sens.* **5**, 949–981 (2013).
9. C. Sun et al., "Using multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region," *Sensors*, **19** 1–23 (2019).
10. A. Tridawati et al., "Mapping the distribution of coffee plantations from multi-resolution, multi-temporal, and multi-sensor data using a random forest algorithm," *Remote Sens.* **12**, 3933 (2020).
11. A. Escobar-López et al., "Identifying Coffee agroforestry system types using multitemporal Sentinel-2 data and auxiliary information," *Remote Sens.* **14**, 3847 (2022).
12. M. A. Ortega-Huerta et al., "Mapping coffee plantations with land sat imagery: an example from El Salvador," *Int. J. Remote Sens.* **33**, 220–242 (2012).
13. M. Schmitt-Harsh, "Landscape change in Guatemala: driving forces of forest and coffee agroforest expansion and contraction from 1990 to 2010," *Appl. Geogr.* **40**, 40–50 (2013).
14. R. Hebbar et al. "National level inventory of coffee plantations using high resolution satellite data," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* **XLII-3/W6**, 293–298 (2019).
15. E. A. B. Calderon et al., "Coffee agroforestry systems in Veracruz, Mexico: spatial identification and quantification using GIS, remote sensing and local knowledge," *Terra Latinoam.* **36**, 261–273 (2018).
16. L. C. Kelley et al., "Using Google Earth Engine to map complex shade-grown coffee landscapes in northern Nicaragua," *Remote Sens.* **10**, 952 (2018).
17. G. Mosomtai et al., "Functional land cover scale for three insect pests with contrasting dispersal strategies in a fragmented coffee-based landscape in Central Kenya," *Agric. Ecosyst. Environ.* **319**, 107558 (2021).
18. A. Tariq et al., "Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest," *Geo-Spatial Inf. Sci.* **26**, 302–320 (2023).
19. N. Kussul et al., "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.* **14**, 778–782 (2017).
20. M. Ozdogan et al., "Remote sensing of irrigated agriculture: opportunities and challenges," *Remote Sens.* **2**, 2274–2304 (2010).
21. E. M. Abdel-Rahman et al., "Estimation of thrips (*Fulmekiola serrata* Kobus) density in sugarcane using leaf-level hyperspectral data," *South African J. Plant Soil* **30**, 91–96 (2013).
22. I. M. Iqbal et al., "Identifying the spectral signatures of invasive and native plant species in two protected areas of Pakistan through field spectroscopy," *Remote Sens.* **13**, 4009 (2021).
23. M. Sibanda et al., "Exploring the potential of *in situ* hyperspectral data and multivariate techniques in discriminating different fertilizer treatments in grasslands," *J. Appl. Remote Sens.* **9**, 096033 (2015).
24. K. Jia et al., "Spectral discrimination of opium poppy using field spectrometry," *IEEE Trans. Geosci. Remote Sens.* **49**, 3414–3422 (2011).
25. M. Naghdyzadegan Jahromi et al., "Developing machine learning models for wheat yield prediction using ground-based data, satellite-based actual evapotranspiration and vegetation indices," *Eur. J. Agron.* **146**, 126820 (2023).
26. Y. Chen et al., "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**, 2094–2107 (2014).
27. A. B. Santos et al., "Combining multiple classification methods for hyperspectral data interpretation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6**, 1450–1459 (2013).
28. A. Signoroni et al., "Deep learning meets hyperspectral image analysis: a multidisciplinary review," *J. Imaging* **5**, 52 (2019).
29. V. Vapnik, "Estimation of dependences based on empirical data," Nauk, Moscow (1979). Transl. vol. 27, Springer Verlag, New York, pp. 51–84 (1982).
30. R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.* **7**, 179–188 (1936).
31. J. H. Friedman, "Stochastic gradient boosting," *Comput. Stat. Data Anal.* **38**, 367–378 (2002).
32. M. Reza, S. Miri, and R. Javidan, "A hybrid data mining approach for intrusion detection on imbalanced NSL-KDD dataset," *Int. J. Adv. Comput. Sci. Appl.* **7**, 1–25 (2016).
33. E. Adam et al., "Detecting the early stage of phaeosphaeria leaf spot infestations in maize crop using *in situ* hyperspectral data and guided regularized random forest algorithm," *J. Spectrosc.* **2017**(1), (2017).
34. N. Mureriwa, E. M. I. Adam, and S. Tesfamichael, "Examining the spectral separability of *Prosopis glandulosa* from co-existent species using field spectral measurement and guided regularized random forest," *Remote Sens.* **8**, 144 (2016).

35. H. Deng and G. Runger, "Feature selection via regularized trees," in *Proc. Int. Jt. Conf. Neural Networks* (2012).
36. A. P. Davis et al., A review of the indigenous coffee resources of Uganda and their potential for coffee sector sustainability and development," *Front. Plant Sci.* **13**, 1–16 (2022).
37. "ASD. FieldSpec® HandHeld 2 User Manual," (2010) http://www.geo-informatie.nl/courses/grs60312/material2017/manuals/600860-dHH2Manual.pdf.
38. J. Dai et al., "Mapping understory invasive plant species with field and remotely sensed data in Chitwan, Nepal," *Remote Sens. Environ.* **250**, 112037 (2020).
39. H. Meyer et al., "Hyperspectral data analysis in R: the hsdar package," *J. Statistical Software* **89**, 1–22 (2019).
40. R Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna (2022) https://www.R-project.org.
41. R. C. G. Smith et al., "Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite," *Aust. J. Agric. Res.* **46**, 113–125 (1995).
42. S. J. Jeong, C. I. Ho, and J. H. Jeong, "Increase in vegetation greenness and decrease in springtime warming over east Asia," *Geophys. Res. Lett.* **36**, 1–5 (2009).
43. J. Penuelas, F. Baret, and I. Filella, "Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance," *Photosynthetica* **31**, 221–230 (1995).
44. G. V. G. Baranoski and J. G. Rokne, "A practical approach for estimating the red edge position of plant leaf reflectance," *Int. J. Remote Sens.* **26**, 503–521 (2005).
45. B. Datt, "Visible/near infrared reflectance and chlorophyll content in eucalyptus leaves," *Int. J. Remote Sens.* **20**, 2741–2759 (1999).
46. D. A. Sims and J. A. Gamon, "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 (2002).
47. J. E. Vogelmann, B. N. Rock, and D. M. Moss, "Red edge spectral measurements from sugar maple leaves," *Int. J. Remote Sens.* **14**, 1563–1575 (1993).
48. O. Mutanga and A. K. Skidmore, "Narrow band vegetation indices overcome the saturation problem in biomass estimation," *Int. J. Remote Sens.* **25**, 3999–4014 (2004).
49. A. A. Gitelson and M. N. Merzlyak, "Remote estimation of chlorophyll content in higher plant leaves," *Int. J. Remote Sens.* **18**, 2691–2697 (1997).
50. M. N. Merzlyak et al., "Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening," *Physiol. Plant.* **106**, 135–141 (1999).
51. J. Huang et al., "Meta-analysis of the detection of plant pigment concentrations using hyperspectral remotely sensed data," *PLoS One* **10**, e0137029 (2015).
52. A. Große-stoltenberg et al., "Evaluation of continuous VNIR-SWIR spectra versus narrowband hyperspectral indices to discriminate the invasive *Acacia longifolia* within a Mediterranean dune ecosystem," *Remote Sens.* **8**(4), 334 (2016).
53. E. Izquierdo-Verdiguier and R. Zurita-Milla, "An evaluation of guided regularized random forest for classification and regression tasks in remote sensing," *Int. J. Appl. Earth Obs. Geoinf.* **88**, 102051 (2020).
54. H. Deng, "Guided random forest in the RRF Package," arXiv:1306.0237 pp. 3–5 (2013).
55. H. Deng and G. Runger, "Gene selection with guided regularized random forest," *Pattern Recognit.* **46**, 3483–3489 (2013).
56. B. T. Mudereri et al. "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 (2020).
57. A. Zafari, R. Zurita-Milla, and E. Izquierdo-Verdiguier, "A multiscale random forest kernel for land cover classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **13**, 2842–2852 (2020).
58. T. Dube and O. Mutanga, "Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa," *ISPRS J. Photogramm. Remote Sens.* **101**, 36–46 (2015).
59. E. M. Abdel-Rahman et al., "Detecting *Sirex noctilio* grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers," *ISPRS J. Photogramm. Remote Sens.* **88**, 48–59 (2014).
60. P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.* **27**, 294–300 (2006).
61. T. D. Adugna, A. Ramu, and A. Haldorai, "A review of pattern recognition and machine learning," *J. Machine Comput.* **4** 210–220 (2024).
62. A. Karatzoglou and A. Smola, "kernlab—an S4 package for kernel methods in R," *J. Statist. Software* **11**, 1–9 (2004).
63. J. Holloway and K. Mengersen, "Statistical machine learning methods and remote sensing for sustainable development goals: a review," *Remote Sens.* **10**, 1–21 (2018).
64. A. Liaw and M. Wiener, "Classification and regression by random forest," *R. News* **2**, 18–22 (2002).

65. W. N. Venables and B. D. Ripley, "Statistics and Computing: Modern Applied Statistics with S," Springer-Verlag, New York Inc., New York (2002).
66. A. Max et al., "Package 'Caret' R topics documented," (2023).
67. B. Greenwell et al., "GBM: generalized boosted regression models, R package version 2.1.5," (2019), CRAN Repos, https://cran.r-project.org/.
68. C. M. Chance et al., Invasive shrub mapping in an urban environment from hyperspectral and LiDAR-derived attributes," *Front. Plant. Sci.* **7**, 1528 (2016).
69. I. Aneece and H. Epstein, "Identifying invasive plant species using field spectroscopy in the VNIR region in successional systems of north-central Virginia," *Int. J. Remote Sens.* **38**, 100–122 (2017).
70. K. Prospere, K. McLaren, and B. Wilson, "Plant species discrimination in a tropical wetland using *in situ* hyperspectral data," *Remote Sens.* **6**, 8494–8523 (2014).
71. S. Panigrahy, T. Kumar, and K. R. Manjunath, "Hyperspectral leaf signature as an added dimension for species discrimination: case study of four tropical mangroves," *Wetlands Ecol. Management*, **20**, 101–110 (2012).
72. I. Filella and J. Peñuelas, "The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status," *Int. J. Remote Sens.* **15**, 1459–1470 (1994).
73. P. S. Thenkabail, R. B. Smith, and E. De. Pauw, "Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization," *Photogramm. Eng. Remote Sens.* **68**, 607–621 (2002).
74. A. Thomson, J. Jacobs, and E. Morse-McNabb, "Comparing the predictive ability of Sentinel-2 multispectral imagery and a proximal hyperspectral sensor for the estimation of pasture nutritive characteristics in an intensive rotational grazing system," *Comput. Electron. Agric.* **214**, 108275 (2023).
75. J. Zhang et al., "Intra- and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: implications for species identification using HYDICE imagery," *Remote Sens. Environ.* **105**, 129–141 (2006).
76. S. Ren et al., "Assessing plant senescence reflectance index-retrieved vegetation phenology and its spatio-temporal response to climate change in the Inner Mongolian Grassland," *Int. J. Biometeorol.* **61**, 601–612 (2017).
77. A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: an applied review," *Int. J. Remote Sens.* **39**, 2784–2817 (2018).
78. P. S. Thenkabail et al., "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 (2013).
79. M. A. Moreira, M. Adami, and B. F. T. Rudorff, "Spectral and temporal behavior analysis of coffee crop in Landsat images," *Pesqui. Agropecu. Bras.* **39**, 223–231 (2004).
80. A. Sabat-Tomala, E. Raczko, and B. Zagajewski, "Comparison of support vector machine and random forest algorithms for invasive and expansive species classification using airborne hyperspectral data," *Remote Sens.* **12**, 516 (2020).
81. S. J. Underwood et al., "Atmospheric circulation patterns, cloud-to-ground lightning, and locally intense convective rainfall associated with debris flow initiation in the Dolomite Alps of northeastern Italy," *Nat. Hazards Earth Syst. Sci.* **16**, 509–528 (2016).

Getachew Kebede obtained a BSc degree in natural resources economics and management from Mekelle University and a master of science in geo-information from Bahir Dar University in Ethiopia. He worked for more than 6 years at the Ethiopian Environment and Forest Research Institute as a director and researcher on the GIS and remote sensing unit. Currently, he is pursuing a PhD in environmental science at the University of KwaZulu-Natal in South Africa.

Biographies of the other authors are not available.