Smart Agricultural Technology 14 (2026) 102229

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Prediction of biomass and nutritional quality of tropical pastures using multispectral analysis and machine learning models

Josué Tafur-Culqui [a,*], Nilton Atalaya-Marin [a], Darwin Gómes-Fernandez [a], Victor H. Taboada-Mitma [a], Juancarlos Cruz-Luis [b], Henri Neyra [a], Janella Y. Anchayhua [a], Rosalía Quichua-Baldeon [a], Teiser Sánchez-Fuentes [a], Yadhira M. Olano [a], Mauro Barrazueta [a], Daniel Tineo [a], Malluri Goñas [a]

[a] Centro Experimental Yanayacu, Dirección de Servicios Estratégicos Agrarios (DSEA), Instituto Nacional de Innovación Agraria (INIA), Carretera Jaén San Ignacio KM 23.7, Jaén 06801, Cajamarca, Perú
[b] Dirección de Servicios Estratégicos Agrarios (DSEA), Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Perú

ARTICLE INFO

Keywords:
Crude protein prediction
Dry matter
Extra trees
Random forest
Remote sensing
UAV multispectral imagery

ABSTRACT

Determining pasture productivity and nutritional value through non-destructive approaches aimed at optimizing forage resource management and improving efficiency in livestock systems has become an urgent priority. In this context, the objective of this study was to evaluate the performance of machine learning models in predicting biomass production and the nutritional contribution of different pasture species, as well as to assess the role of vegetation indices (VIs) in these predictions. To this end, a multispectral sensor mounted on a DJI Matrice 350 RTK platform was used, together with agronomic, yield, and nutritional variables. The curated dataset was subsequently analyzed using linear and polynomial models, as well as tree-based algorithms and support vector machines. Model validation was performed using a group-constrained random partitioning scheme (Group Shuffle Split), with species considered as the grouping variable. Model interpretability was addressed through the SHAP (SHapley Additive Explanations) framework. The results indicated better predictive performance for yield-related variables compared to nutritional attributes. In particular, the Extra Trees model achieved the highest coefficients of determination (R^2). SHAP analysis revealed that the Visible Atmospherically Resistant Index (VARI) contributed more strongly to yield-related predictions, whereas the Normalized Difference Red Edge (NDRE) showed a more consistent contribution to nutritional variables. In conclusion, these findings highlight the potential of integrating vegetation indices and machine learning models as effective tools for forage management, supporting informed decision-making in livestock production systems.

1. Introduction

Pastures represent the primary food resource for livestock, supply essential nutrients in livestock production systems [1]. Globally, these ecosystems occupy approximately 70% of arable land [2] and perform key functions such as biodiversity conservation, soil structure improvement, erosion prevention [3,4], water conservation, and carbon sequestration [4]. Likewise, grasslands have the potential to contribute to the achievement of sustainable development goals, reflecting their relevance to the environment, economic growth, and community resilience [5]. In the Peruvian context, in 2025 pastures covered 26,050, 443.71 ha; however, their management remains limited, as only 9% of these areas are managed pastures, which limits productivity and land use efficiency [6]. In addition, there is a critical need for accurate, non-destructive tools that enable timely estimation of both forage biomass and nutritional quality under tropical conditions.

Conventional pasture sampling methods, which rely on in situ assessment of morphophysiological traits, productive variables, and nutritional contribution, are labor-intensive, time-constrained, and insufficient to support precision-based pasture management [7,8]. Consequently, they are inadequate for high-frequency monitoring, particularly in tropical environments where rapid phenological changes,

* Corresponding author.
E-mail address: tafurjosue27@gmail.com (J. Tafur-Culqui).

https://doi.org/10.1016/j.atech.2026.102229
Received 13 January 2026; Received in revised form 30 April 2026; Accepted 17 May 2026
Available online 17 May 2026

structural heterogeneity, and regrowth dynamics compromise the reliability of productivity and nutritive value estimations. On the other hand, the integration of machine learning (ML) into agriculture has enabled more accurate crop yield predictions and data-driven decision-making [9]. Likewise, remote sensing techniques, particularly those based on multispectral imagery acquired through unmanned aerial vehicles (UAVs) and satellite platforms, have emerged as promising alternatives for pasture monitoring [10,11]. Vegetation indices (VIs) derived from multispectral imagery, such as the Normalized Difference Vegetation Index (NDVI), have been widely used to estimate biomass and canopy characteristics; however, NDVI and other VIs tend to saturate at high biomass levels, thereby reducing their sensitivity [12]. In contrast, red-edge-based VIs have demonstrated greater sensitivity to canopy structural changes, offering potential advantages for estimating both biomass and forage quality; nevertheless, their effectiveness in predicting nutritional parameters such as crude protein (CP), crude fiber (CF), and ash remains inconsistent [13], particularly in tropical systems.

Likewise, advances in machine learning (ML) have expanded the capacity to model both linear and nonlinear relationships between spectral predictors and agronomic responses. Tree-based algorithms, such as Random Forest (RF), have demonstrated robustness to multicollinearity and nonlinear effects [14]. However, most studies on pasture monitoring have focused primarily on biomass estimation, limiting the understanding of the comparative performance of different ML algorithms across tropical forage species and multiple phenological stages. Furthermore, although remote sensing represents a valuable tool for optimizing pasture management, its application to decision-making processes such as determining optimal cutting age remains insufficiently explored. Early harvesting generally produces forages with higher protein content and lower fiber concentration, whereas delayed cutting increases accumulated biomass but reduces digestibility and overall nutritional quality [15].

Similarly, the assessment of key nutrients such as protein, ether extract, fiber, and minerals is essential for selecting forage species that more efficiently meet livestock nutritional requirements and enhance animal productivity [1]. In this context, the present study evaluated the performance of machine learning models based on vegetation indices to estimate green forage yield (kg ha^{-1}), dry matter yield (kg ha^{-1}), and nutritional composition (crude protein, ether extract, crude fiber, and ash) in six tropical grass species. By providing a multispecies assessment of UAV-based monitoring, this study contributes to precision livestock management and expands the understanding of the capabilities and limitations of spectral–nutritional modeling in tropical pasture systems.

2. Methodology

2.1. Study area

The study area was located at the geographic coordinates 5°40′35.12″ latitude and 78°46′32.82″ longitude, within the tropical pasture agrostological garden at the Yanayacu Experimental Center of the instituto Nacional de Innovación Agraria (INIA), situated in the district of Jaén, province of Jaén, Cajamarca region, northern Peru, at an altitude of 727 m above sea level (Fig. 1). The evaluated area is characterized by predominantly flat topography and slightly alkaline clay loam soils (pH = 7.8). Soil chemical analysis, performed on composite samples collected at a depth of 0 to 20 cm, revealed a total nitrogen content of 1.05 mg g^{-1}, available potassium of 92.31 mg kg^{-1}, and available phosphorus of 18.31 mg kg^{-1}. Laboratory determinations were

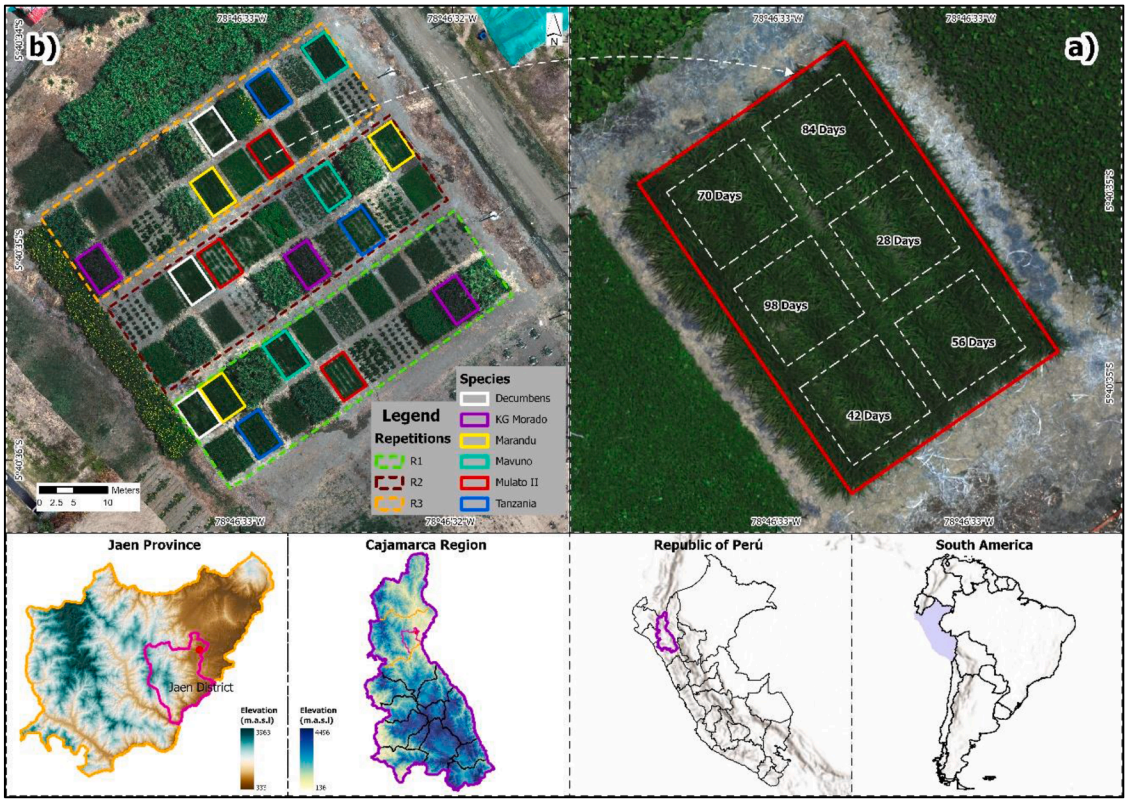

Fig. 1. Map showing the location and spatial distribution of sampling units in the tropical grass agrostological garden at the Yanayacu Experimental Center, belonging to the National Institute for Agricultural Innovation (INIA). a) Experimental plot comprising the sampling units, organized randomly according to the days established for evaluation. b) Spatial distribution of the experimental plots within the study area. The map was produced by the authors using open-access geospatial resources. Provincial and district boundaries were obtained from the Geoportal of the Instituto Geográfico Nacional del Perú (IGN) (https://www.idep.gob.pe/geovisor/VisorDeMapas-3D/) in shapefile format, referenced to the WGS 1984 datum. This map is provided for illustrative purposes only.

conducted at the Red de Laboratorios de Suelos, Aguas y Foliares (LABSAF Yanayacu), accredited under the NTP-ISO/IEC 17,025:2017 standard, thereby ensuring analytical traceability and quality assurance of the results.

The climate of the study area is semi-dry characterized by high relative humidity throughout the year [16]. The study was conducted from October 2024, when the pasture establishment cut was performed, through January 2025. During the evaluation period, temperatures ranged from 20.1 to 37.1 °C, with minimal precipitation between October and mid-November and regular to heavy rainfall from December to January (Fig. 2). These data were obtained from the Jaén meteorological station of the Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), located 200 m from the tropical pasture agrostological garden, at 5°40′35.99″ latitude and 78°46′27.05″ longitude.

A randomized complete block design (RCBD) was implemented to evaluate six forage grass species under six cutting frequencies, with three replicates, resulting in a total of 108 experimental units (6 species × 6 cutting frequencies × 3 replicates). Each experimental unit consisted of a 1 m^2 quadrat demarcated to facilitate the assessment of response variables. The species evaluated were Brachiaria decumbens (Brachiaria decumbens), Brachiaria marandú (Brachiaria brizantha cv. Marandú), Brachiaria mavuno (Brachiaria hybrida cv. Mavuno), Brachiaria Mulato II (Brachiaria hybrida cv. Mulato II), Purple King Grass (Pennisetum purpureum × Pennisetum typhoides), and Tanzania (Panicum maximum cv. Tanzania). All grasses were established by continuous row seeding, with 80 cm spacing between rows. No fertilization or pest management practices were applied; however, drip irrigation was supplied three times per week during the dry season. Six cutting frequencies were evaluated for each species: 28, 42, 56, 70, 84, and 98 days after the establishment cut. At the time of evaluation, pastures were at least six months old after sowing; this criterion was adopted to ensure adequate regrowth following the establishment cut.

The map was produced by the authors using open-access geospatial resources. Provincial and district boundaries were obtained from the Geoportal of the Instituto Geográfico Nacional del Perú (IGN). (https://www.idep.gob.pe/geovisor/VisorDeMapas-3D/) in shapefile format, referenced to the WGS 1984 datum. This map is provided for illustrative purposes only.

2.2. Methodological workflow

The methodological workflow of this study was divided into two phases: data collection and data analysis, which are described below.

2.2.1. Data acquisition

2.2.1.1. Multispectral data acquisition. Multispectral data were acquired using a MicaSense RedEdge-P multispectral sensor mounted on a DJI Matrice 350 RTK unmanned aerial vehicle (UAV). Specifically, this sensor incorporates a reflectance panel for radiometric calibration under field conditions, as well as a downelling light sensor (DLS) designed to record irradiance and solar angle during flights. In addition, it integrates a high-spatial resolution panchromatic band that enables the generation of pan-sharpened imagery. According to the manufacturer's specifications, the system is capable of producing imagery with a spatial resolution of approximately 2 cm at a flight altitude of 60 m. Furthermore, the sensor captures data across five multispectral bands: Blue (475 nm center wavelength, 32 nm bandwidth), Green (560 nm center, 27 nm bandwidth), Red (668 nm center, 14 nm bandwidth), Red Edge (717 nm center, 12 nm bandwidth), and Near-Infrared (842 nm center, 57 nm bandwidth). To ensure geospatial accuracy, four ground control points were surveyed using a South Galaxy G7 GNSS receiver. Flight missions were conducted under cloud-free atmospheric conditions between 10:00 and 14:00 h, at an altitude of 15 m above ground level and a constant flight speed of 2 m s^{-1}, with both lateral and frontal overlap set at 70%. A total of six flights were carried out, corresponding to each cutting interval (28, 42, 56, 70, 84, and 98 days after the establishment cut).

The multispectral images were processed using Agisoft Metashape Professional v.2.2.1 software. This process included photo alignment, adjustment with four ground control points, dense point cloud generation, digital elevation model (DEM) construction, and the creation of a multiband orthomosaic of the evaluation plots with an average pixel size of 0.6 cm. Finally, the orthomosaic was analyzed in ArcGIS Pro v. 3.5 software, which produced vegetation indices (VIs) (Table 1). For each experimental plot, the VIs values were averaged within the evaluated

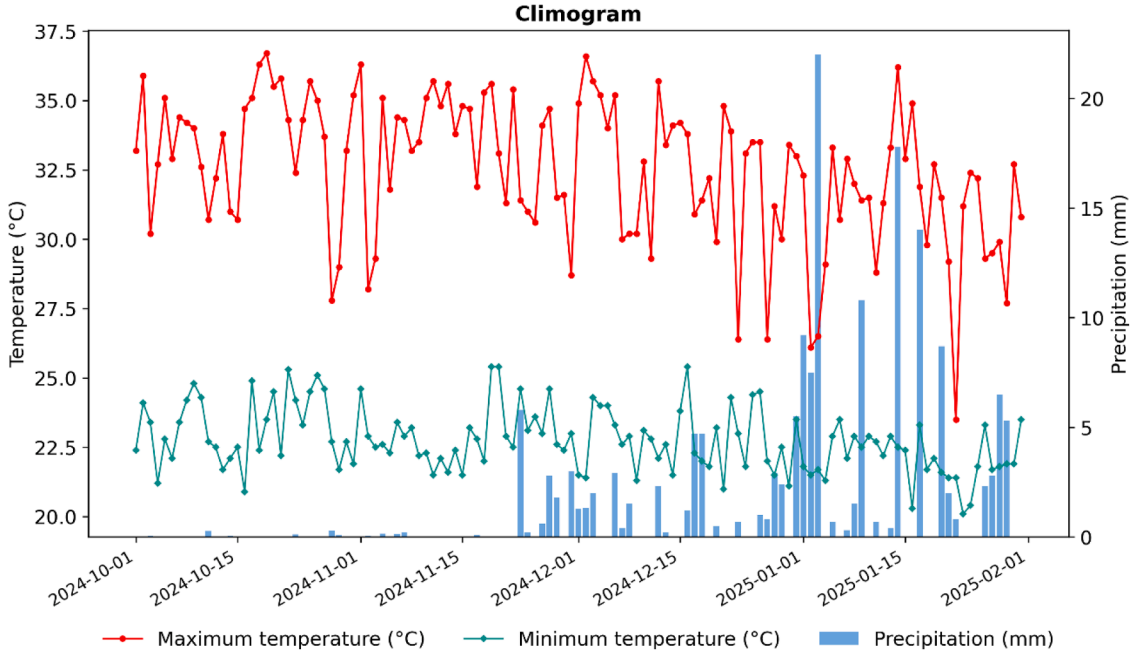

Fig. 2. Climate dynamics of the study area for the period between October 2024 and January 2025. The data were obtained from the Jaén weather station, which belongs to the Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI).

Table 1
Vegetation indices (VIs) obtained with the MicaSense Red-Edge-P coupled to a DJI Matrice 350 RTK from DJI.

Number	Variable (Symbol)	Meaning	Equations	Reference
1	EVI	Enhanced vegetation index	$EVI = 2.5\frac{NIR - RED}{(NIR + 6RED - 7.5BLUE) + 1}$	[17]
2	GNDVI	Green normalized difference vegetation index	$GNDVI = \frac{NIR - GREEN}{NIR + GREEN}$	[18]
3	MSAVI	Modified soil adjusted vegetation index	$MSAVI = 0.5\left(2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - RED)}\right)$	[19]
4	NDRE	Normalized difference red edge index	$NDRE = \frac{NIR - RED_{edge}}{NIR + RED_{edge}}$	[20]
5	NDVI	Normalized difference vegetation index	$NDVI = \frac{NIR - RED}{NIR + RED}$	[21]
6	OSAVI	Optimized soil adjusted vegetation index	$OSAVI = \frac{NIR - RED}{NIR + RED + 0.16}$	[22]
7	SAVI	Soil adjusted vegetation index	$SAVI = 1.5\frac{NIR - RED}{NIR + RED + 0.5}$	[23]
8	VARI	Visible atmospherically resistant index	$VARI = \frac{GREEN - RED}{GREEN + RED - BLUE}$	[24]

Note: The spectral bands used in the calculation of vegetation indices correspond to the BLUE, GREEN, RED, near-infrared (NIR), and red Edge (RED_{edge}) bands.

quadrant to obtain representative spectral metrics. The VIs were selected based on the study by [12], which emphasizes the indicesmost commonly used to calculate biomass and nutritional quality of forages.

2.2.1.2. Agronomic data acquisition. For each cutting frequency, a non-destructive morphometric measurements and relative chlorophyll assessments were conducted within a 1 m^2 quadrat established inside each experimental unit. The variables recorded included plant height (cm), number of clump density (clumps m^{-2}), number of tillers per clump, leaf length (cm), leaf width (cm), and relative chlorophyll content (SPAD units). Within each quadrat, three clumps were selected to represent the average canopy condition while minimizing border effects. Clumps exhibiting mechanical damage or visible stress symptoms were excluded. For each clump, three independent readings per variable were obtained to reduce intra-individual variability and enhance measurement precision. Plant height was measured from ground level to the apical meristem, excluding inflorescences when present. Leaf length and width were determined on fully expanded leaves located in the upper third of the canopy. Relative chlorophyll content was measured using a portable SPAD-502DL Plus meter (Konica Minolta, Japan), with readings taken at the central portion of the same leaf used for morphometric measurements and between 10:00 a.m. and 2:00 pm. to minimize diurnal variability. Relative chlorophyll content was measuredusing a portable SPAD 502DL Plus equipment (Konica Minolta, Japan). Measurements were first averaged at the clump level and subsequently across the three sampled clumps to obtain a representative value per experimental unit and evaluation date. This procedure ensured correspondence with UAV-derived vegetation indices calculated at the quadrat scale.

2.2.1.3. Yield data acquisition. After completing the morphometric evaluation at each phenological stage, destructive sampling was conducted. To determine both fresh forage yield and dry matter yield, all aboveground biomass contained within the 1 m^2 quadrat was cut at a height of 10 cm above ground level. The fresh weight of the harvested biomass was recorded immediately after cutting using a previously calibrated digital balance. Although biomass was collected at the quadrat scale (1 m^2), yield values were initially expressed in kg m^{-2} and subsequently converted to kg ha^{-1} to facilitate agronomic interpretation and comparison with conventional forage productivity metrics. To determine the dry matter (DM) percentage, a representative 500 g subsample was taken from the previously homogenized fresh biomass of each quadrat and replication. The subsamples were dried in a forced-air oven (Memmert UF260 Universal Oven, Memmert GmbH, Germany) at 105 °C for 24 h, or until constant weight was achieved, following the procedures established by the Association of Official Analytical Chemists (AOAC 950.46) [25]. The dry matter percentage was calculated as the ratio between dry weight and fresh weight, as shown in Eq. (1), where DM (%) represents the percentage of dry matter, DW corresponds to dry weight (g), and FW represents fresh weight (g). Finally, dry matter yield (kg ha^{-1}) was estimated by multiplying the fresh forage yield by the corresponding dry matter percentage.

$$DM\ (\%) = \frac{DW}{FW}x100 \quad (1)$$

2.2.1.4. Nutritional data acquisition. For nutritional analysis, an additional subsample independent from that used for dry matter determination was collected from the homogenized fresh biomass of each quadrat. This subsample was dried in a forced-air oven at 65 °C for 48 h until reaching approximately 7% residual moisture, a temperature commonly used in forage analysis to minimize thermal degradation of nitrogenous compounds and other heat-sensitive constituents. The dried material was subsequently ground and passed through a 1 mm sieve to obtain a uniform particle size suitable for laboratory analysis. The nutritional composition of the forage samples was determined by the "Laboratorio de evaluación nutricional de alimentos (LENA)" of the "Universidad Nacional Agraria La Molina", Peru; following protocols established by the Association of Official Analytical Chemists (AOAC) [26]. The procedures used to determine the nutritional value of the pastures are detailed below.

2.2.1.4.1. Crude protein analysis. Crude protein content was determined using the Kjeldahl method (AOAC 984.13) [27]. Prior to analysis, catalysts and indicators were prepared, and the hydrochloric acid solution (0.05 N HCl) was standardized using sodium carbonate (Na_2CO_3) as a primary standard. Samples were oven-dried in a forced-air oven (VENTICELL, 111 L, digital temperature control), then ground and homogenized. Approximately 0.2 g of dry sample was weighed in duplicate using an analytical balance (Elicrom, 0.1 mg precision) with a protective enclosure. Digestion was performed in a protein digestion unit by adding concentrated sulfuric acid (H_2SO_4) and a catalyst mixture (K_2SO_4 and $CuSO_4$), followed by heating at 420 °C for approximately 120 min under a fume hood until a clear solution was obtained. Ammonia distillation was subsequently carried out by adding 50% sodium hydroxide (NaOH) using a manual distillation unit. The released ammonia was captured in a boric acid solution (H_3BO_3) containing mixed indicators (methyl red and bromocresol green). The distillate was then titrated with 0.05 N HCl until the endpoint was reached. Reagents were dispensed using volumetric dispensers to ensure precision. Quality control included acid standardization, reagent blanks to correct for external nitrogen contributions, and duplicate analyses to assess method repeatability.

2.2.1.4.2. Ether extract analysis. Ether extract was quantified using a continuous solvent extraction method with a Soxhlet apparatus (AOAC 2003.05) [28]. The initial weight (tare) of 250 mL round-bottom flasks, previously cleaned and dried, was recorded. Samples were weighed in

duplicate on grease-free filter paper (Whatman No. 1) using an analytical balance (Elicrom, 0.1 mg precision). The samples were placed in the Soxhlet apparatus and extracted with hexane as the organic solvent for approximately 3 h, allowing the solubilization and recovery of lipid compounds. After extraction, the solvent was recovered, and the flasks were dried in a forced-air oven (VENTICELL, 111 L) to completely remove residual hexane. The flasks were then cooled in a desiccator containing silica gel for 20 min to prevent moisture absorption. Ether extract content was determined gravimetrically as the difference between initial and final flask weights. Quality control procedures included duplicate analyses, the use of grease-free filter paper to prevent contamination, verification of complete solvent evaporation, stabilization in a desiccator, and periodic calibration of the analytical balance.

2.2.1.4.3. Crude fiber analysis. Crude fiber was determined by sequential acid and alkaline digestion [29]. Dried and ground samples were first defatted using a Soxhlet apparatus with hexane to remove lipids that could interfere with the analysis. The samples were then digested in a fiber digestion unit using sulfuric acid (H_2SO_4), followed by sodium hydroxide (NaOH), under controlled boiling conditions and within a fume hood. After each digestion step, the residue was filtered and washed until neutral pH was achieved. The resulting material was dried in a forced-air oven (VENTICELL, 111 L) at 105 °C and weighed using an analytical balance (Elicrom, 0.1 mg precision). It was subsequently incinerated in a muffle furnace at 600 °C to remove organic matter. Crude fiber content was calculated as the difference between the dried residue and the ash obtained. Quality control included duplicate analyses to ensure repeatability, strict control of digestion times, reagent concentrations, and boiling conditions, as well as prior defatting and final ashing to ensure that only the structural fiber fraction was quantified.

2.2.1.4.4. Ash analysis. Ash content was determined by dry ashing in a muffle furnace as an indicator of total mineral content [30]. Between 1.5 and 2.0 g of dry sample was weighed in duplicate into pre-calcined and tared porcelain crucibles. Samples were incinerated at 700 °C for 5–7 h until a white or light gray residue was obtained, indicating complete combustion of organic matter. The crucibles were then removed using tongs and placed in a desiccator containing silica gel for approximately 20 min to prevent moisture absorption. Final weights were recorded using an analytical balance, and ash content was calculated gravimetrically. Quality control included duplicate analyses, verification of complete incineration (absence of carbonized residues), constant weight determination to ensure complete removal of organic matter, use of a desiccator to prevent moisture variation, and regular calibration of the analytical balance and thermal stability of the muffle furnace.

2.2.2. Data analysis

Statistical analyses were performed using Python version 3.10 within the Google Colaboratory execution environment, employing the following libraries: NumPy [31], pandas [32], matplotlib [33], scikit-learn [34], seaborn [35], scipy [36], skbio [37] y scikit-posthocs [38]. The dataset, which integrated vegetation indices (Table 1), agronomic variables, yield variables, and nutritional parameters (Table 2), was subjected to the Shapiro-Wilk normality test. For this analysis, a p-value ≥ 0.05 was considered indicative of a normal distribution.

2.2.2.1. Criteria for selecting predictive variables. The selection of predictor variables (VIs) was performed using principal component analysis (PCA), complemented by Spearman's correlation analysis. PCA loadings were used to identify the most influential variables within the first three principal components (PC). Subsequently, to detect potential redundancies and mitigate collinearity, Spearman's correlation coefficients were evaluated. Based on this combined approach, VIs exhibiting the highest loadings across the first three principal components and the lowest degree of intercorrelation were selected.

Table 2
Types of variables, variables, abbreviations, and units of measurement for the study variables.

Variable type	Variable	Abbreviation	Unidad
Agronómics (AV)	Chlorophyll (relative value)	CL	SPAD
	Clumps	Clumps	-
	Leaf length	LL	cm
	Leaf width	LW	cm
	Plant height	pH	cm
	Tillers	Tillers	-
Yield (YA)	Dry matter yield	DMY	kg ha^{-1}
	Green forage yield	GF	kg ha^{-1}
	Percentage of dry matter	DM	%
Nutritional (NV)	Ash	ASH	%
	Crude fiber	CF	%
	Crude protein Ether extract	CP	%
	Ether extract	EE	%

Note: The variables "Clumps" and "Tillers" are defined as the numerical values obtained in the assessment area.

Subsequently, a descriptive analysis of both predictor and response variables was conducted. For this purpose, the non-parametric Kruskal-Wallis test was applied to compare pasture species for each variable as a function of evaluation days (28, 42, 56, 70, 84, and 96 days), using significance thresholds of $p < 0,05$ (*), $p < 0,01$ (**) y $p < 0001$ (***). To complement this analysis, Dunn's post hoc test with Bonferroni correction was performed to identify significant differences between pairs of groups ($p < 0.05$), and boxplots were generated for the VIs. To characterize the behavior of agronomic, yield, and nutritional variables throughout the evaluated period, trend lines were constructed, where each line represents the mean value at each evaluation date, and the adjacent shaded area denotes the standard error of the mean (SEM).

2.2.2.2. Implementation of predictive models. The response variables were classified into two groups. The first group comprised production-related parameters, including dry matter yield (kg ha^{-1}), dry matter percentage, and fresh forage yield (kg ha^{-1}). The second group included nutritional variables such as crude protein (CP), ether extract (EE), crude fiber (CF), and ash content (ASH).

Given that measurements were collected at multiple time points (28, 42, 56, 70, 84, and 98 days), the dataset (predictor and response variables) was reshaped from a wide to a long format, allowing temporal observations to be integrated into a unified structure suitable for longitudinal analysis. Additionally, new categorical variables were generated, including base species and their functional grouping (Brachiaria spp., Tanzania, and King Grass). This approach was motivated by the fact that, in agricultural contexts, the presence of correlated predictors and limited sample sizes can compromise model stability, particularly when data are distributed across multiple categories [39]. In this regard, species within the genus Brachiaria share physiological and morphological similarities, supporting their aggregation into a single group. Accordingly, the Brachiaria spp. group included Brachiaria decumbens, Brachiaria hybrid cv. Mavuno, Brachiaria brizantha cv. Marandú, and Brachiaria hybrid cv. Mulato II. This strategy facilitated the capture of general patterns and reduced species-specific variability, thereby minimizing noise in predictive models.

Data preprocessing included the removal of observations with missing values in both predictor and response variables. Only subsets with at least 20 valid observations were retained to ensure statistical stability during model training. Subsequently, a range of regression algorithms representing different learning paradigms was evaluated: a) linear models, including Ridge regression, Elastic Net, and polynomial regression (degree 2 with Ridge regularization); b) non-parametric distance- and kernel-based models, including K-Nearest Neighbors (KNN) and Support Vector Regression (SVR); c) tree-based methods, such as Random Forest and Extra Trees; and, d) boosting approaches including Extreme Gradient Boosting (XGB). For scale-sensitive models (Ridge,

Elastic Net, K-Nearest Neighbors, Support Vector Regression, and polynomial regression), variables were standardized (zero mean and unit variance) using transformations implemented within machine learning pipelines. The evaluation of these diverse algorithms aimed to compare different predictive frameworks capable of capturing the complex dynamics of agricultural systems [40].

2.2.2.3. Evaluation and selection of predictive model. Model validation was performed using a repeated random partitioning scheme with group constraints (Group Shuffle Split). In this approach, the variable "species" was used as the grouping criterion, ensuring that observations belonging to the same species were not simultaneously included in both training and testing sets. A total of 100 independent random splits were generated, in which 80% of the data were allocated for training and 20% for validation. This strategy reduced the bias associated with a single data partition and provided more stable estimates of predictive performance, while preventing information leakage between datasets [34].

To assess the performance of the predictive models, the coefficient of determination (R^2) Eq. (2), the standard deviation of R^2, the root mean square error (RMSE) Eq. (3), and the normalized root mean square error (NRMSE) based on the data range Eq. (4) were calculated, where Y_{max} represents the maximum observed value, Y_{min} the minimum observed value, and N the sample size [41,42]. The R^2 metric indicates the proportion of variation in the dependent variable explained by the independent variables; a value of $R^2 > 0.5$ was established as indicative of good model performance, meaning that the model explains >50% of the data variability [43,44]. The analysis of R^2 was complemented by its standard deviation, which allowed assessment of model stability across multiple cross-validation splits. Meanwhile, the NRMSE quantifies model error as a percentage of the observed data range; in this study, simulation accuracy was classified as "excellent" when NRMSE ⟨ 0.1, "good" when values ranged between 0.1 and 0.2, "acceptable" when between 0.2 and 0.3, and "poor" when NRMSE ⟩ 0.3 [45].

$$R^2 = \frac{\sum (Y_i - Y_J)^2}{\sum (Y_i - Y_J)^2} \tag{2}$$

$$RMSE = \sqrt{\frac{1 \sum (Y_i - Y_J)^2}{N}} \tag{3}$$

$$NRMSE = \frac{RMSE}{Y_{imax} - Y_{imin}} \tag{4}$$

The best models were selected using a multi-metric approach, prioritizing those with the highest coefficient of determination (R^2) and lowest standard deviation, supplemented by the normalized root mean square error (NRMSE) as an additional performance criterion. To ensure model interpretability, the SHAP (SHapley Additive Explanations) method was used, which breaks down predictions into the individual contributions of each variable. TreeExplainer was used for tree-based models, and model-agnostic approaches were used for other algorithms.

3. Results

3.1. Variable selection

Variable selection was based on the factor loadings derived from principal component analysis (PCA) and Spearman's correlation analysis. PCA enabled the identification of the most influential vegetation indices (VIs) within eachprincipal component (PC). In PC1, the indices OSAVI and SAVI exhibited the highest contributions, in PC2, GNDVI and VARI; whereas in PC3, NDRE was identified as the most influential index (Fig. 3a and b). Overall, the first three PCs explained a cumulative variance of 98.49% (Fig. 3a). Spearman's correlation analysis revealed strong associations among EVI, MSAVI, OSAVI, and SAVI, with coefficients exceeding 0.90, as well as between GNDVI and NDRE (r = 0.96) (Fig. 4). In this context, variables with lower inter-correlation were selected to mitigate collinearity. Accordingly, the indices with the highest factor loadings from each of the first three PCs were retained (Fig. 3a and b). Furthermore, these indices represent distinct spectral regions; for instance, OSAVI (PC1) captures soil-adjusted reflectance, VARI (PC2) corresponds to the visible spectrum, and NDRE (PC3) represents the red-edge region. This selection enabled an integrated characterization of canopy structure, chlorophyll content, and soil background effects.

Asterisks indicate levels of statistical significance: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

3.2. Temporal dynamics of vegetation indices, agronomic, yield, and nutritional variables

The Chi-square test showed that the data were not normally distributed; therefore, nonparametric tests were chosen. To determine significant differences among groups, the Kruskal-Wallis test was used, followed by Dunn's post hoc tests to identify significant differences

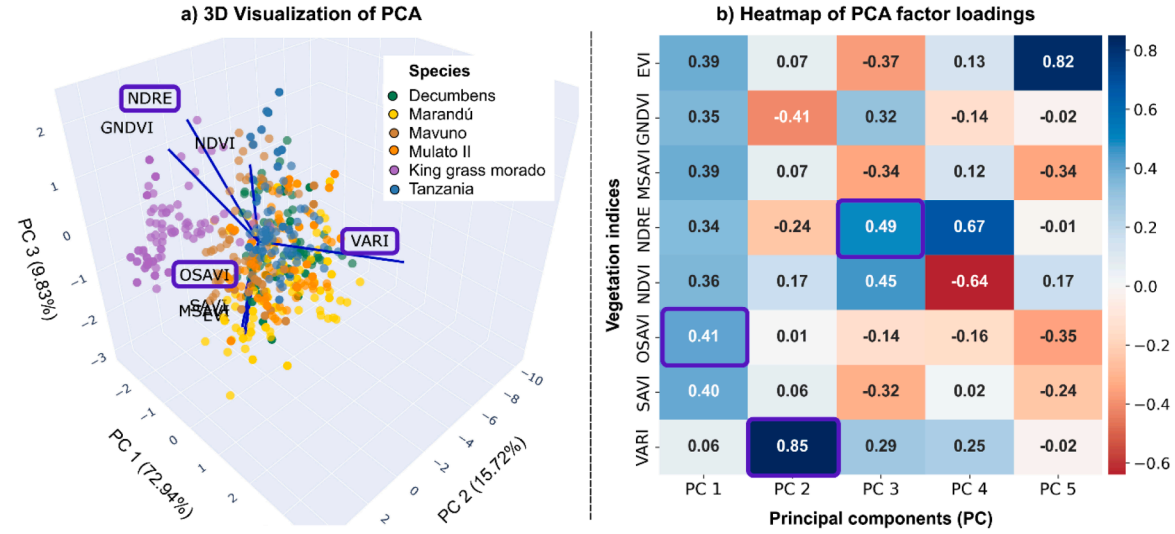

Fig. 3. Principal component analysis (PCA) of vegetation indices (VIs). (a) Biplot representing the distribution of vegetation indices as a function of the first three principal components, indicating the proportion of variance explained by each component. (b) Factor loadings of the vegetation indices corresponding to the first, second and third principal components.

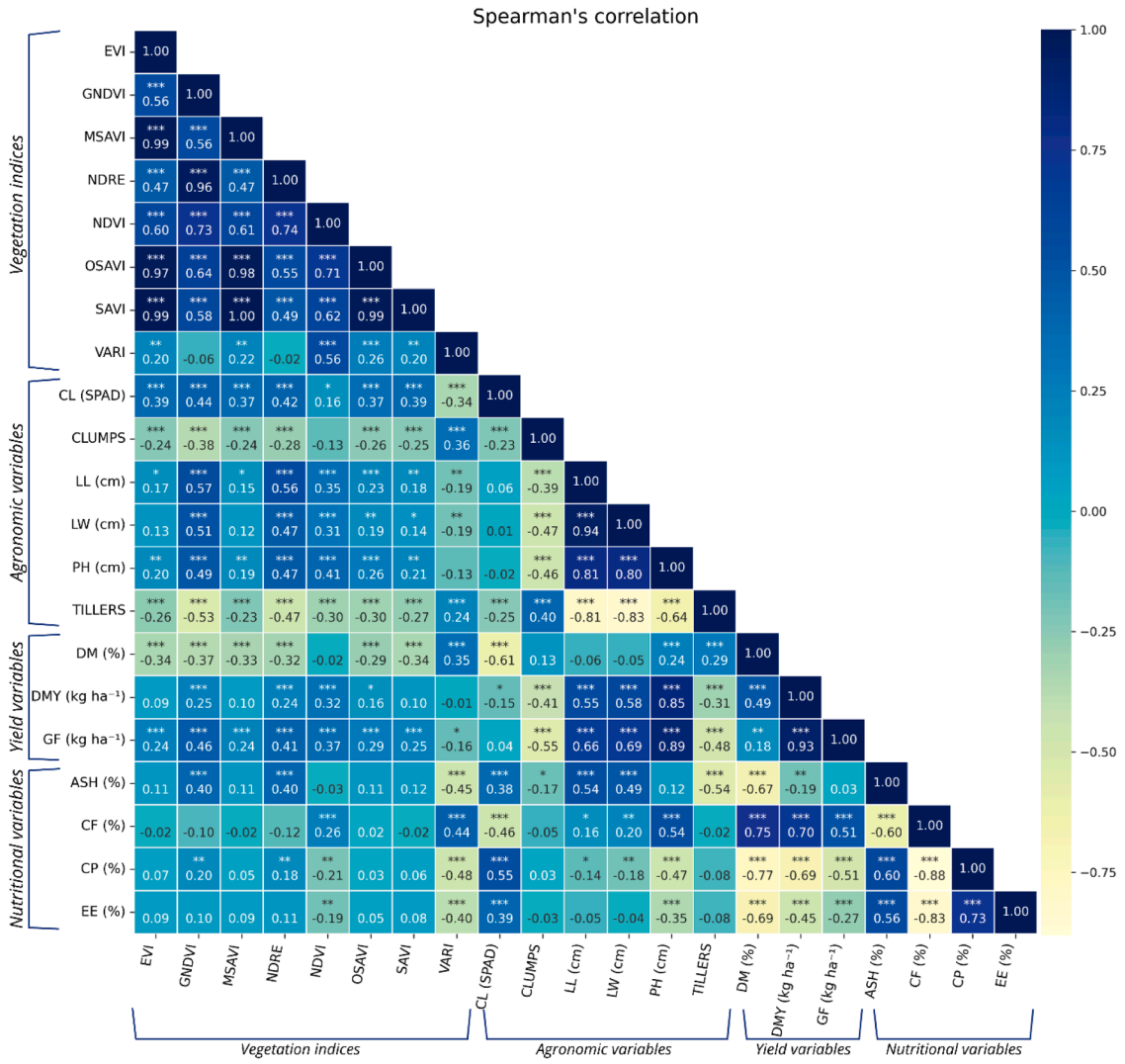

Fig. 4. Spearman correlation matrix showing vegetation indices: EVI, GNDVI, MSAVI, NDRE, NDVI, OSAVI, SAVI, VARI; agronomic variables: relative chlorophyll value (CL), clumps (CLUMPS), leaf length (LL), leaf width (LW), plant height (pH), tillers (TILLERS); yield variables: percentage of dry matter (DM), dry matter yield (DMY), green forage yield (GF); and nutritional variables: ash (ASH), crude fiber (CF), crude protein (CP), ether extract (EE).

considering the following values: $p < 0,05$ (*), $p < 0,01$ (**) y $p < 0001$ (***).

3.2.1. *Temporal dynamics of vegetation indices*

The behavior of vegetation indices for each evaluated species by assessment day is shown in Fig. 6. Regarding NDRE, a progressive increase in normalized values was observed up to day 42, followed by a slightly decreasing trend at later stages of the grass development (Fig. 5a). In the case of OSAVI, significant differences among species were observed, with B. decumbens and B. marandú exhibiting the highest values (Fig. 5b).Finally, VARI showed marked variability among species, with purple King Grass consistently presenting the lowest normalized values throughout the evaluation period (Fig. 5c).

3.2.2. *Agronomic, yield, and nutritional variables*

Regarding agronomic, yield, and nutritional variables, significant differences were observed across all variables and grass species (Supplementary Table 1). Concerning chlorophyll value, significant differences ($p < 0.05$) were found among species throughout the evaluation period; for example, at 28 days after the establishment cut, Brachiaria mulato II and Tanzania showed significant differences compared to the other grass species (Fig. 6a). Purple King grass stood out among the species, exhibiting the highest chlorophyll levels (Fig. 6b). Regarding the number of clumps, a progressive decrease was observed in most of the evaluated species; however, purple King grass maintained relative stability throughout all analyzed periods. Among the Brachiaria species, Brachiaria decumbens had the lowest number of clumps, while Brachiaria mavuno exhibited the highest value (Fig. 6b). For leaf length (Fig. 6c) and leaf width (Fig. 6d), a more pronounced growth pattern was observed in Tanzania and purple King grass, which consistently exceeded the other species at all assessment dates. Similarly, plant size showed a parallel trend, again highlighting Tanzania and purple King grass as the species with the greatest structural development throughout the experiment (Fig. 6e). Finally, the number of tillers per clump remained relatively constant over time (Fig. 6f). This result suggests that, despite the reduction in total clump numbers, there is a tendency for increased tiller production per plant unit, which could partially compensate for the loss of individuals.

Regarding yield, it was observed that older grasses exhibited a higher percentage of dry matter; however, slight decreases were noted in Brachiaria mavuno (day 56), Brachiaria mulato II (day 70), and Tanzania (day 70) when evaluating this parameter (Fig. 7a). Concerning dry matter yield per hectare (Fig. 7b) and green forage yield (Fig. 7c), a progressive increase in production was observed as the grasses aged.

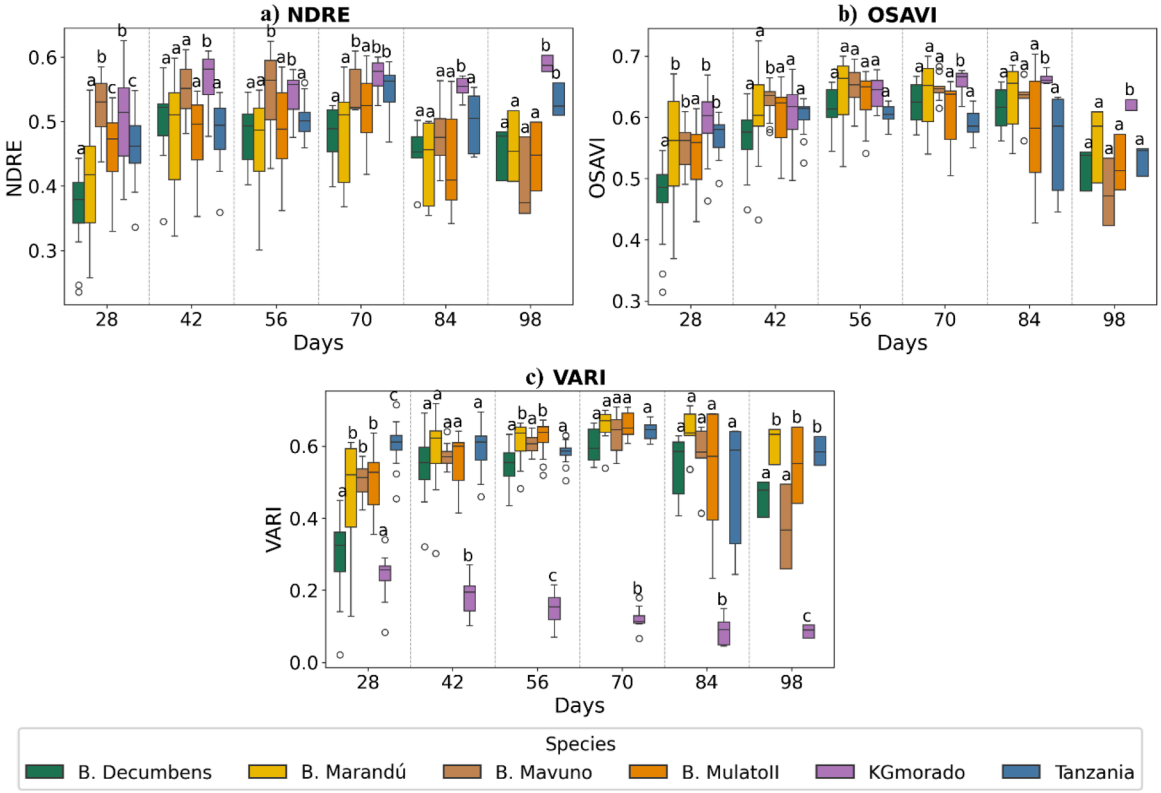

Fig. 5. Temporal dynamics of vegetation indices for each evaluated species. Dunn's post hoc tests with Bonferroni correction were applied to determine differences among grass species by assessment day. The vegetation indices are detailed as follows: (a) NDRE, (b) OSAVI and (c) VARI. Boxplot colors represent grass species: Brachiaria decumbens (green), Brachiaria marandú (yellow), Brachiaria mavuno (brown), Brachiaria mulato II (orange), purple king grass (lilac), and Tanzania (light blue).

Similarly, an increase in crude fiber content was evident (Fig. 7d), accompanied by a decrease in the percentage of ether extract (Fig. 7e) and crude protein (Fig. 7f). However, purple King grass exhibited a particular behavior. However, purple King grass exhibited a particular behavior. Although an increasing trend was observed for most yield parameters (percentage of dry matter, dry matter yield, and green forage yield), the nutritional quality followed a different pattern, with purple King grass showing a slight increase in protein levels (Fig. 7f).

3.3. Selection of predictive models by forage species

The machine learning models exhibited a high predictive capacity, with normalized errors (NRMSE) classified as “excellent” and “good”. In addition, the standard deviation of R^2 was in most cases, below 0.15, indicating good stability and consistency of the predictive models for estimating the target variables.

Tree-based algorithms (Extra Trees, Random Forest, and XGB) consistently exhibited the best performance for yield-related variables, including dry matter percentage, achieving R^2 values above 0.73 and NRMSE below 0.17 across all three species (Fig. 8a). For dry matter yield (kg ha^{-1}), the Extra Trees algorithm achieved the best performance with R^2 values exceeding 0.77 and NRMSE below 0.14 (Fig. 8b). Similarly, for fresh forage yield (kg ha^{-1}), Extra Trees consistently outperformed the other models attaining R^2 values of 0.73 for Brachiaria spp., 0.84 for purple King grass, and 0.76 for Tanzania grass (Fig. 8c).

Regarding the nutritional composition of the grasses, the Extra Trees algorithm again outperformed the other models. For ash content, Extra Trees achieved R^2 values greater than 0.55 and NRMSE below 0.23 for both purple King grass and Tanzania grass. In contrast, Brachiaria spp. showed lower performance, with $R^2 = 0.43$ and NRMSE = 0.17 (Fig. 9a). Differences among species were also observed for crude fiber and crude protein. For crude fiber, purple King grass exhibited the highest predictive performance, reaching $R^2 = 0.79$ and NRMSE = 0.17 (Fig. 9b). In the case of crude protein, Extra Trees achieved its best performance in Tanzania grass, with $R^2 = 0.71$ and NRMSE = 0.15 (Fig. 9c). Regarding ether extract, Extra Trees once again showed consistent results, with R^2 values above 0.54 and NRMSE below 0.19 for both Brachiaria spp. and purple King grass; however, in Tanzania grass, a lower performance was observed ($R^2 = 0.25$) (Fig. 10).

The evaluated metrics included the coefficient of determination (R^2) with standard deviation and the normalized root mean square error (NRMSE), both represented within the bars of the chart. The models assessed were Extra Trees, Random Forest (RF), eXtreme Gradient Boosting (XGB), K-Nearest Neighbors (KNN), Polynomial Regression (PR), Ridge, Elastic Net, and Support Vector Regression (SVR). The best-performing models for each species were highlighted using colored arrows corresponding to the evaluated species: Brachiaria spp. (orange arrow), purple King grass (green), and Panicum maximum cv. Tanzania (blue).

The evaluated metrics included the coefficient of determination (R^2) with standard deviation and the normalized root mean square error (NRMSE), both represented within the bars of the chart. The models assessed were Extra Trees, Random Forest (RF), eXtreme Gradient Boosting (XGB), K-Nearest Neighbors (KNN), Polynomial Regression (PR), Ridge, Elastic Net, and Support Vector Regression (SVR). The best-performing models for each species were highlighted using colored arrows corresponding to the evaluated species: Brachiaria spp. (orange arrow), purple King grass (green), and Panicum maximum cv. Tanzania (blue).

The evaluated metrics included the coefficient of determination (R^2) with standard deviation and the normalized root mean square error (NRMSE), both represented within the bars of the chart. The models

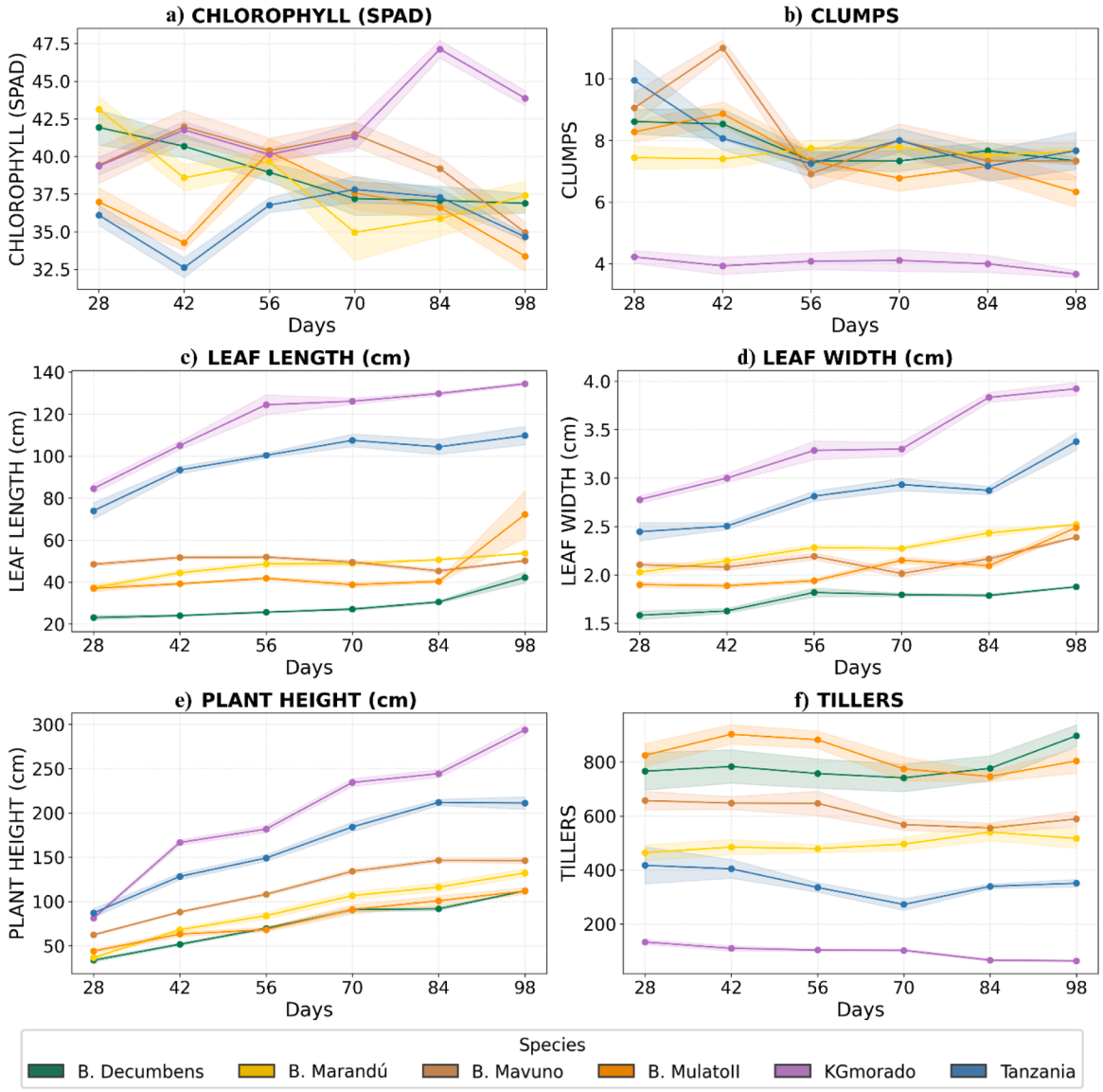

Fig. 6. Temporal dynamics of agronomic variables for each species evaluated. The trend lines represent the mean value of the evaluations performed, while the shaded region adjacent to each line indicates the Standard Error of the Mean (SEM) corresponding to the following variables: agronomic variables: a) relative chlorophyll value (SPAD); b) clumps, c) leaf length (cm) d) leaf width (cm); e) plant height (cm); f) tillers. Each color in the boxplots represents a grass species: Brachiaria decumbens (green), Brachiaria marandú (yellow), Brachiaria mavuno (brown), Brachiaria mulato II (orange), purple king grass (lilac), and Tanzania (light blue).

assessed were Extra Trees, Random Forest (RF), eXtreme Gradient Boosting (XGB), K-Nearest Neighbors (KNN), Polynomial Regression (PR), Ridge, Elastic Net, and Support Vector Regression (SVR). The best-performing models for each species were highlighted using colored arrows corresponding to the evaluated species: Brachiaria spp. (orange arrow), purple King grass (green), and Panicum maximum cv. Tanzania (blue).

3.4. Importance of variables in the model, determined using SHAP analysis

The comparative analysis using SHAP values enabled the assessment of the relative contribution of vegetation indices (NDRE, OSAVI, and VARI) in predicting yield and nutritional variables across the evaluated forage species. For dry matter percentage, NDRE showed the highest relative contribution in Brachiaria spp. (Fig. 11a) and Tanzania grass (Fig. 11c), whereas VARI exhibited higher contributions in purple King Grass (Fig. 11b). Regarding dry matter yield (kg ha^{-1}), VARI presented the greatest contributions in both Brachiaria spp. (Fig. 11d) and purple King Grass (Fig. 11e); in contrast, OSAVI and NDRE showed higher contributions in Tanzania grass (Fig. 11f). For fresh forage yield (kg ha^{-1}), VARI showed the highest contributions in Brachiaria spp. (Fig. 11g) and purple King Grass (Fig. 11h), while NDRE was the most influential index in Tanzania grass (Fig. 11i).

For nutritional variables, vegetation indices exhibited differentiated patterns. For ash content, NDRE showed the highest contributions across Brachiaria spp., purple King Grass, and Tanzania grass (Fig. 12a–c, respectively). In the case of crude fiber, NDRE predominated in Brachiaria spp. and Tanzania grass (Fig. 12d and f), whereas OSAVI was more influential in purple King Grass (Fig. 12e). With respect to crude protein, VARI exhibited the highest SHAP values in Brachiaria spp. (Fig. 12g) and purple King Grass (Fig. 12h), while in Tanzania grass, NDRE showed a broader contribution range (Fig. 12i). For ether extract, NDRE had the highest contributions across all three forage species. In Brachiaria spp., the three indices showed moderate and balanced contributions (Fig. 12j). In purple King Grass, SHAP values were generally

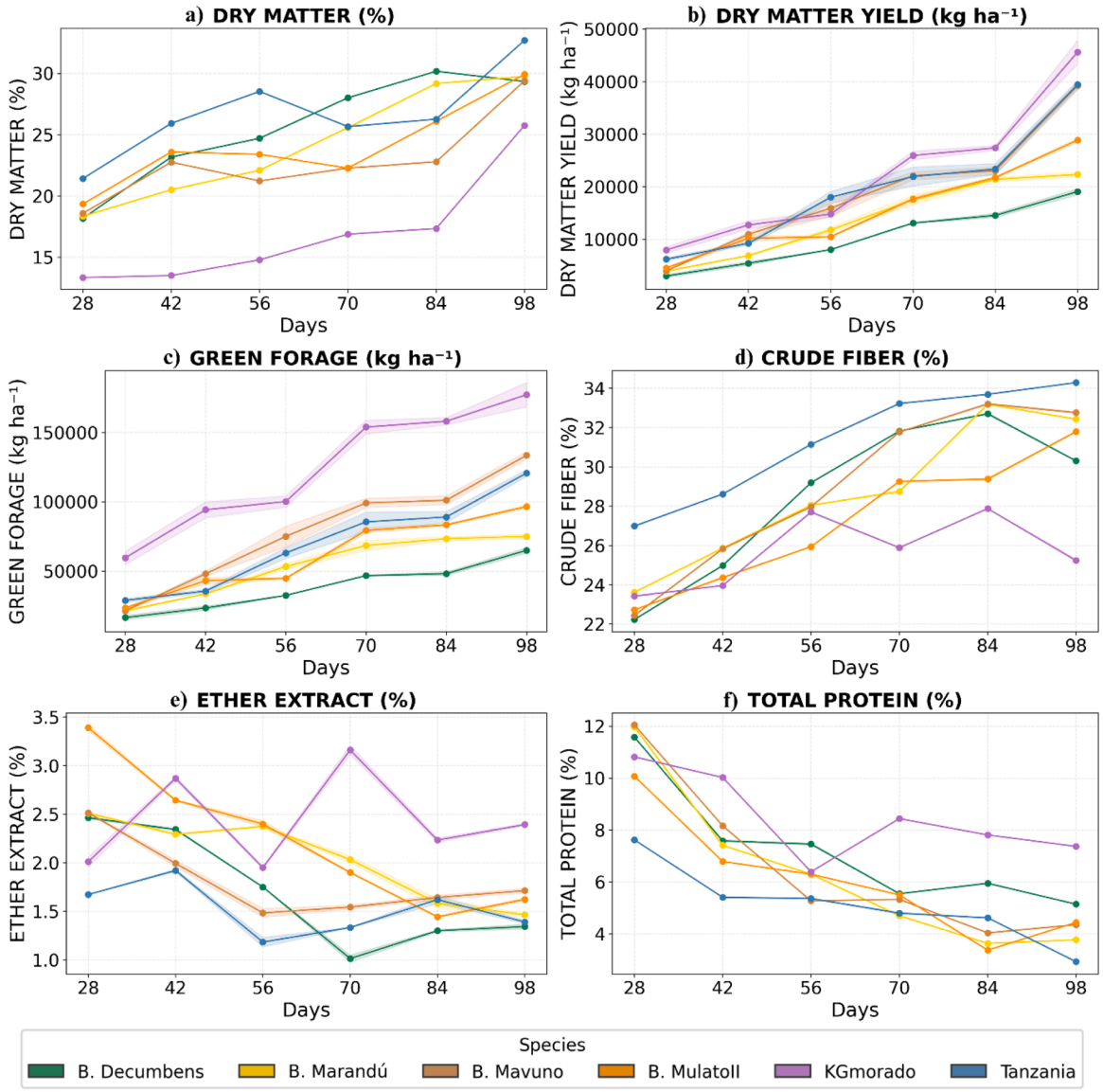

Fig. 7. Temporal dynamics of yield and nutritional variables for each species evaluated. The trend lines represent the mean value of the evaluations performed, while the shaded region adjacent to each line indicates the Standard Error of the Mean (SEM) corresponding to the following variables: yield variables: a) percentage of dry matter, b) dry matter yield (kg ha^{-1}); c) green forage yield (kg ha^{-1}); and nutritional variables: d) crude fiber; e) ether extract; f) crude protein. Each color in the boxplots represents a grass species: Brachiaria decumbens (green), Brachiaria marandú (yellow), Brachiaria mavuno (brown), Brachiaria mulato II (orange), purple king grass (lilac), and Tanzania (light blue).

lower, with a slight predominance of NDRE (Fig. 12k). In contrast, in Tanzania grass, NDRE was the most consistent predictor, although with lower magnitude compared to the other vegetation indices (Fig. 12i).

4. Discussion

4.1. Relationships between vegetation indices, agronomic variables, and nutritional composition

Vegetation indices showed clear associations with key agronomic and productive variables. In particular, indices incorporating the Red-edge region, such as NDRE, exhibited correlations with leaf width, with leaf length, plant size, and green forage yield (kg ha^{-1}). These results are consistent with studies reporting that Red-edge–based indices can explain a significant proportion of the variability in foliar nitrogen concentration and chlorophyll content, both closely linked to the physiological status of the plant [46,47]. This supports the notion that spectral signals integrate both structural and biochemical information. In contrast, indices derived from the visible spectrum, such as VARI, are primarily influenced by canopy structure, vegetation cover, and illumination conditions, whereas indices like NDRE capture physiological traits associated with chlorophyll and nitrogen content [46,48].Agronomic variables such as plant height, leaf length, and leaf width clustered closely with dry matter yield (DMY), reflecting their structural dependence. In contrast, nutritional variables particularly CP and CF, exhibited opposing patterns, highlighting the inverse relationship between forage quality and physiological maturity, which reflects the progressive lignification of foliar tissues over time [49,50].

However, not all pasture species followed this trend. In the case of purple king grass, a slight increase in CP levels and a decrease in CF levels were observed, which may be associated with the emergence of new tillers that introduce younger, protein-rich tissues, thereby buffering the progressive decline in CP [51]. Additionally, a higher number of clumps may limit morphological development, possibly as a result of increased competition for light. In this regard, as noted by [52], greater plant density per unit area tends to reduce individual plant size;

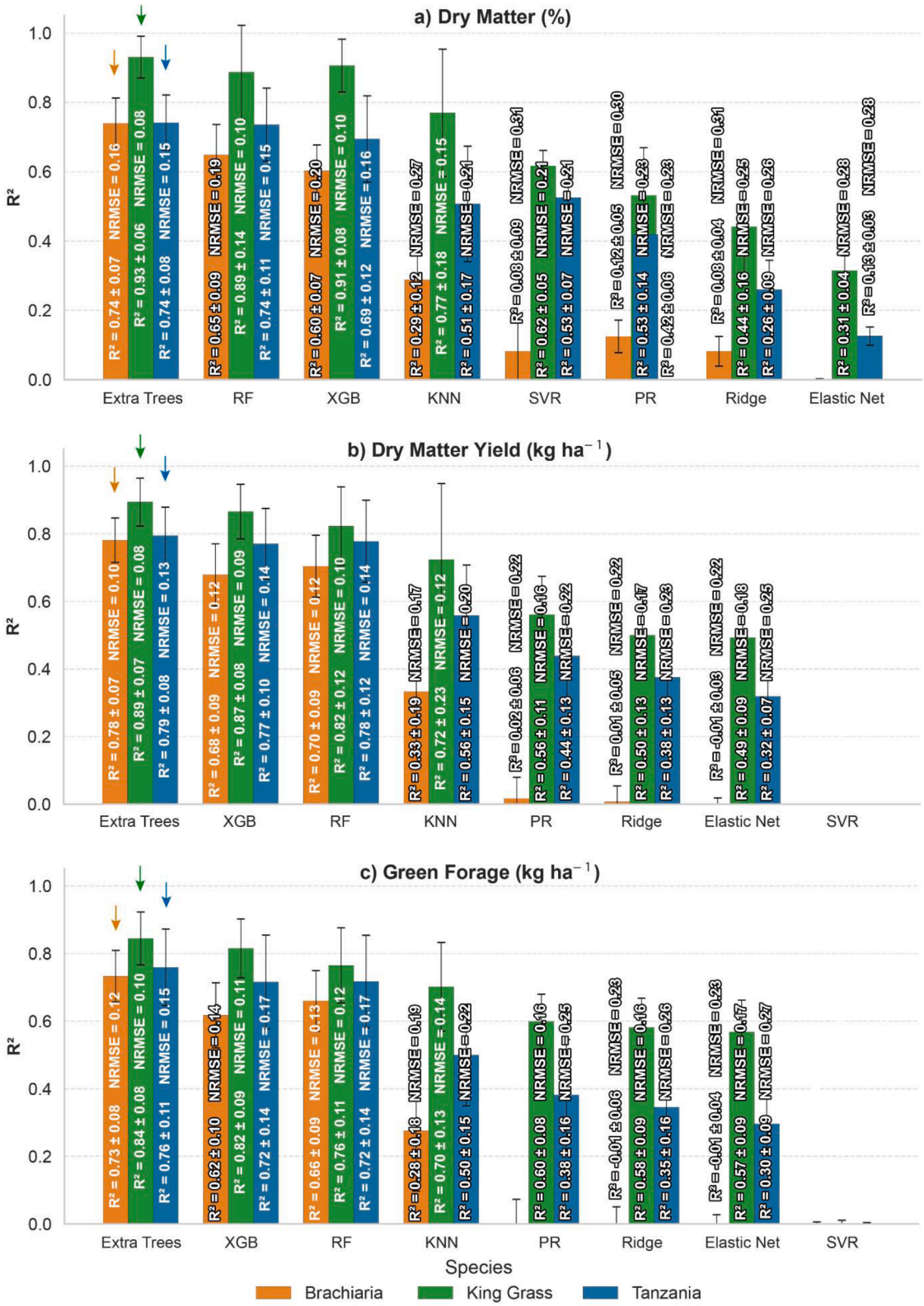

Fig. 8. Performance of machine learning models in predicting yield parameters of Brachiaria spp., purple King grass, and Panicum maximum cv. Tanzania: a) Dry matter percentage, b) Dry matter yield (kg ha^{-1}), c) Fresh forage yield (kg ha^{-1}).

however, the effect of shading on growth is more influential than soil nutrient depletion. Furthermore, although pasture size generally showed an increasing trend, a slight reduction in plant height was observed in Tanzania grass and Brachiaria mavuno at 84 days. This response may be attributed to stem weakening or the action of strong winds or intense rainfall causing lodging [53] which can be considered an indirect indicator of forage maturity.

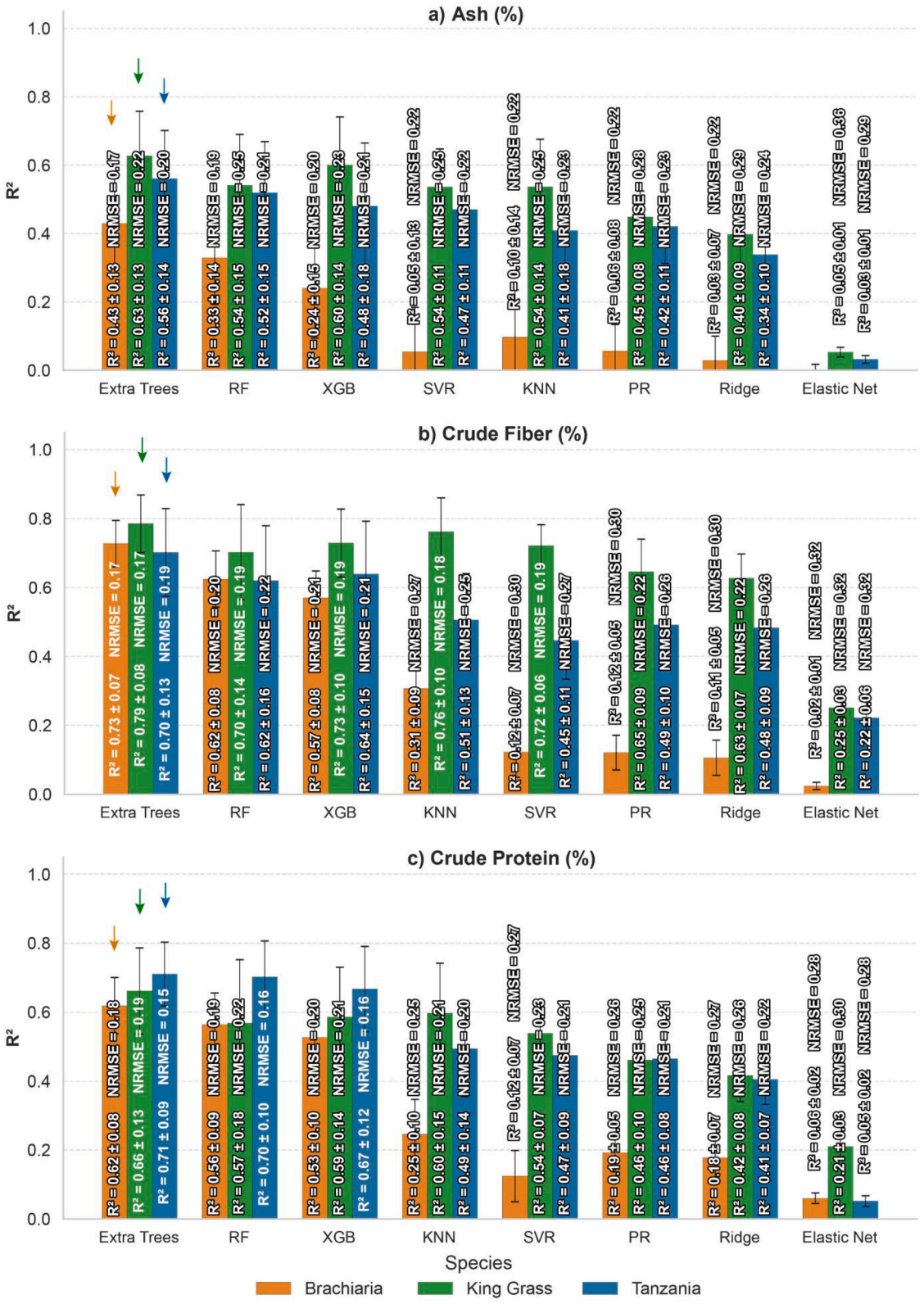

Fig. 9. Performance of machine learning models in predicting nutritional parameters of Brachiaria spp., purple King grass, and Panicum maximum cv. Tanzania: a) Ash, b) Crude fiber, c) Crude protein.

4.2. Performance of machine learning-based predictive models

Machine learning based predictive models demonstrated a high capacity to estimate yield and nutritional quality variables using vegetation indices as the sole predictors. In particular, tree-based algorithms such as Random Forest and Extra Trees showed superior performance, which can be attributed to their ability to model complex nonlinear relationships and handle the collinearity inherent to spectral data [54, 55].

Differences in predictive accuracy among variables reflect the

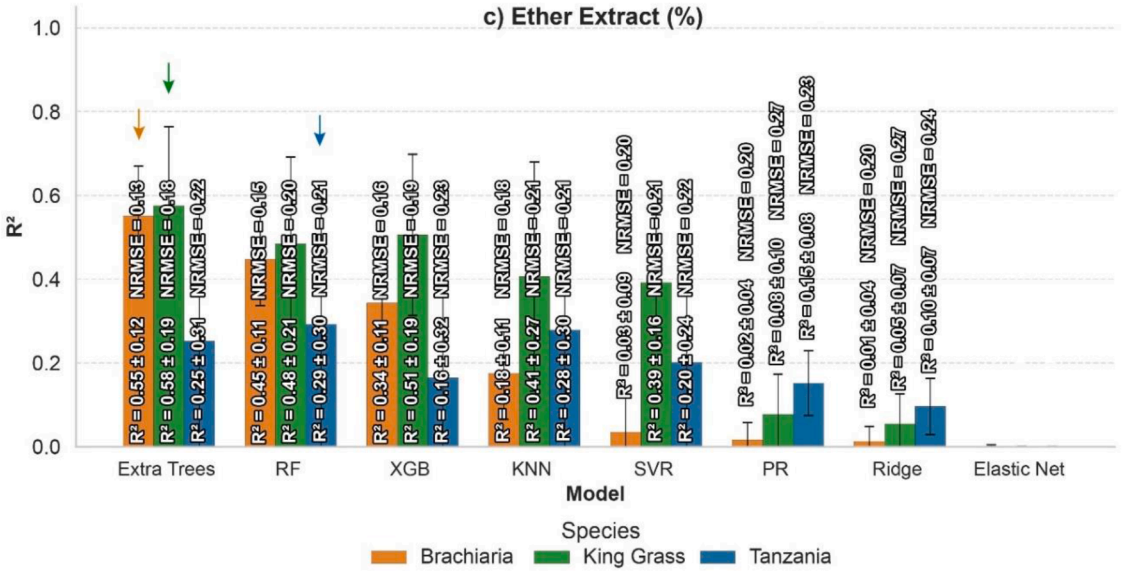

Fig. 10. Performance of machine learning models in predicting ether extract (%) in Brachiaria spp., purple King grass, and Panicum maximum cv. Tanzania.

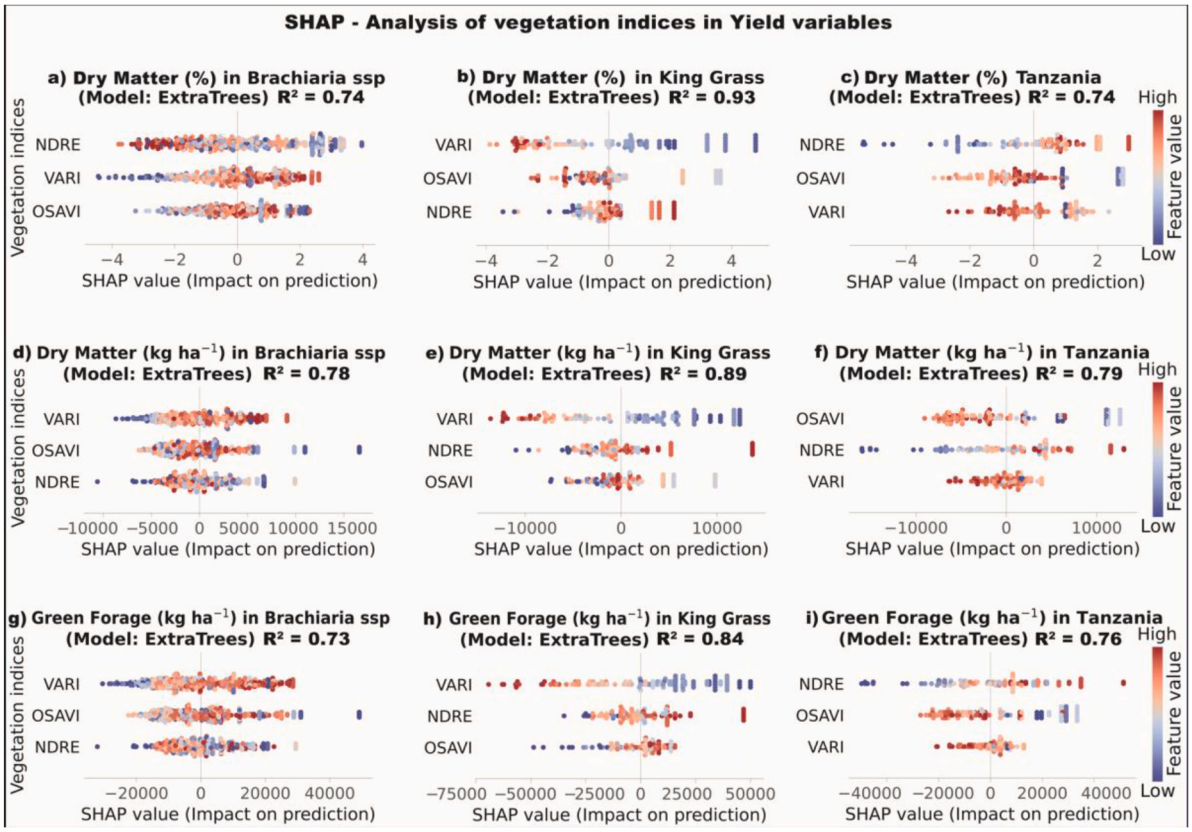

Fig. 11. Comparative analysis of the relative contribution of vegetation indices (NDRE, OSAVI, VARI) using SHapley Additive exPlanations (SHAP values) in the prediction of productive yield variables (dry matter percentage, dry matter yield, and green forage yield).

intrinsic nature of the evaluated attributes. Yield-related variables, being directly associated with canopy structure and biomass accumulation, exhibit more robust and consistent spectral signals likely due to their range of variability. This behavior has been previously reported in forage systems, where variables with greater dispersion tend to be better captured by predictive models [56]. In contrast, variables such as dry matter percentage or specific nutritional components depend on more complex physiological processes and are less directly observable, which limits their detectability through optical sensors.

Furthermore, the effect of phenological stage on predictive model performance highlights the importance of structural stability, as canopy structure determines the quality of the spectral response. At early growth stages, spatial heterogeneity, combined with the influence of exposed soil and limited foliar development, results in less representative signals. Conversely, as the crop advances toward more mature stages, increased vegetation cover and canopy consolidation enhance

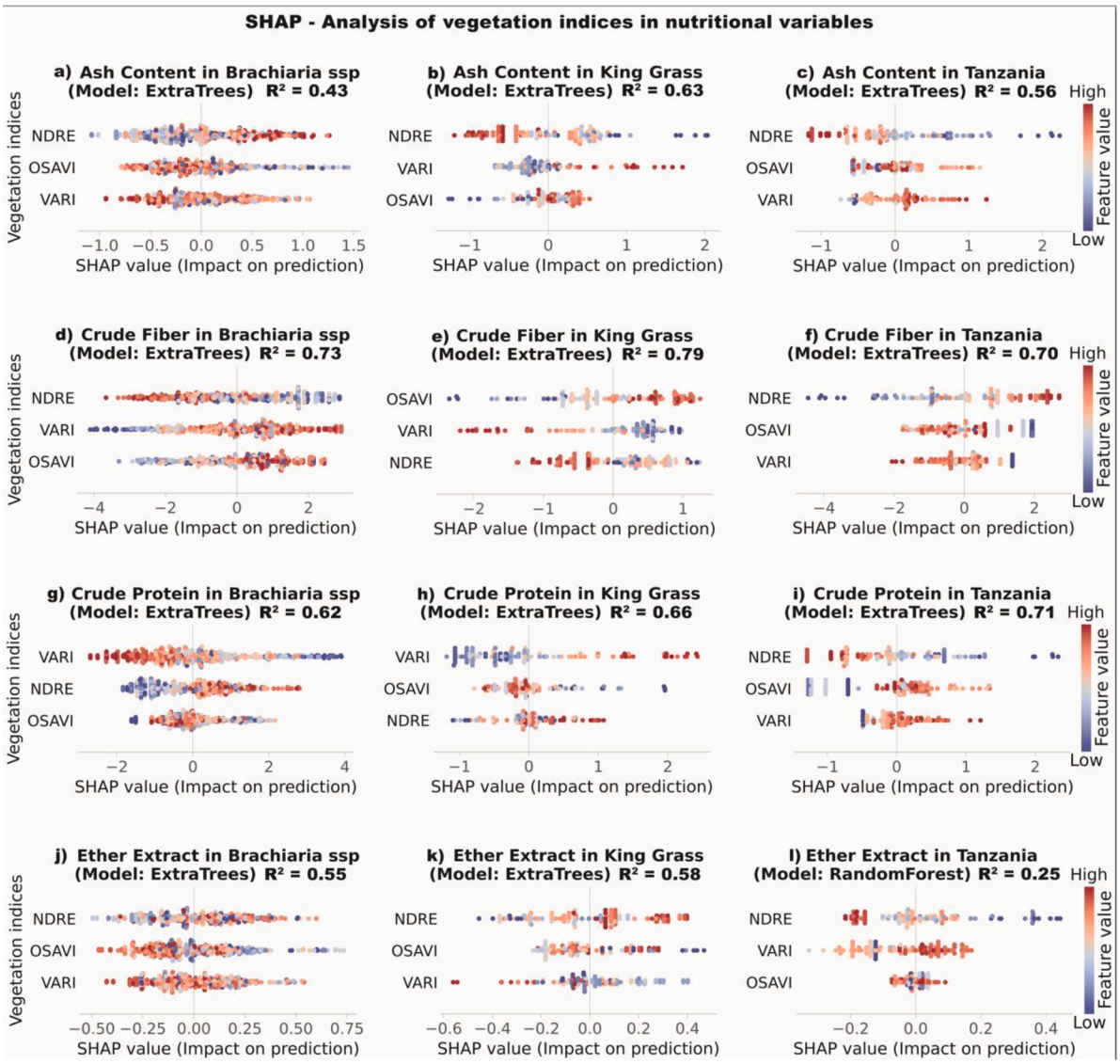

Fig. 12. Comparative analysis of the relative contribution of vegetation indices (NDRE, OSAVI, VARI) using SHapley Additive exPlanations (SHAP values) in the prediction of nutritional variables (ash, crude fiber, crude protein, and ether extract).

spectral signal capture, leading to significant improvements in predictive performance. This pattern has been documented in remote sensing studies applied to agricultural systems [57].

The consistent performance of NDRE underscores the importance of the red-edge region for monitoring crops with high vegetation density. Unlike traditional indices such as NDVI, which tend to saturate under high biomass conditions, NDRE maintains sensitivity to variations in chlorophyll and nitrogen content, making it a more reliable indicator of plant physiological status [24]. In contrast, the behavior of OSAVI highlights the importance of soil background correction, particularly under conditions of partial canopy cover or during intermediate stages of crop development [22,58,59].

4.3. *Interpretation of variable contribution using SHAP values*

The SHAP-based analysis enabled interpretation of the relative contribution of vegetation indices (VIs) for both yield and nutritional variables. Clear differences were identified between structural and nutritional responses. This approach has been widely used to interpret complex models and to understand the individual influence of predictor variables, particularly in nonlinear frameworks. Regarding yield-related variables, VARI generally exhibited higher-magnitude SHAP values, especially in Brachiaria spp. and purple King Grass, suggesting greater sensitivity to canopy structure and biomass accumulation. This behavior is consistent with the nature of the index, which is derived from the visible spectrum and is strongly influenced by vegetation cover, canopy architecture, and soil background reflectance [24]. In this sense, visible-spectrum indices tend to be associated with structural variables due to their sensitivity to the fraction of green cover and the spatial distribution of vegetation.

In contrast, NDRE showed more consistent contributions for nutritional variables such as crude fiber (CF), crude protein (CP), and ash content. This is associated with its sensitivity to the red-edge region, which is closely linked to chlorophyll and foliar nitrogen content [24]. Consequently, the ability of NDRE to capture physiological and nutritional variation is related to the role of nitrogen as a key component of proteins and its involvement in fundamental metabolic processes [60,61].

Although coefficients of determination above 0.61 were achieved indicating that >60% of the variability in the dependent variable was explained by the regression models protein estimation exhibited greater complexity. Notably, variability in the contribution of indices was

observed depending on the pasture species. This suggests that protein content is influenced by multiple interacting factors, including plant physiological status, canopy structure, illumination conditions, and environmental variability [47]. Furthermore, the absence of a dominant index for this variable highlights the need to integrate remote sensing data with additional spectral and non-spectral information sources.

Additionally, the coexistence of positive and negative SHAP values within similar index ranges may indicate the presence of nonlinear relationships and interactions among variables. This behavior is characteristic of complex biological systems, where multiple factors act simultaneously. In this context, tree-based models are particularly suitable, as they can capture such dynamics without assuming predefined linear relationships [55].

4.4. Limitations and future perspectives

Despite the promising results obtained, this study presents several limitations that should be acknowledged. First, the models were developed using data from a single experimental site, which may restrict their applicability under different agroclimatic conditions. Second, although vegetation indices constitute valuable tools for spectral characterization, their exclusive use may limit predictive performance for more complex variables, such as ether extract and protein content. This aspect should not be interpreted as a definitive limitation, but rather as an opportunity to enhance model performance through the integration of additional information sources. Incorporating complementary spectral and non-spectral data could strengthen the estimation of nutritional attributes and improve overall model robustness. In this context, vegetation indices can be regarded as a fundamental component within a broader, integrated framework aimed at the comprehensive characterization of forage systems.

5. Conclusions

This study demonstrated that the use of machine learning models based on vegetation indices constitutes a robust and effective approach for the non-destructive estimation of yield and nutritional quality variables in tropical pastures. Tree-based algorithms, particularly Extra Trees, exhibited superior performance, highlighting their capacity to capture nonlinear relationships inherent to biological systems. It was confirmed that the predictive performance of the models depends not only on the target variables, but also on the grass species and the phenological stage of the crop. Yield-related variables, which are closely associated with canopy structure, showed higher predictive accuracy, whereas nutritional variables especially those lacking a direct spectral signature displayed greater variability. This pattern was supported by SHAP-based analysis, which enabled the identification of functional differences among vegetation indices. Visible-spectrum indices such as VARI were primarily associated with structural and productivity-related variables, whereas NDRE, which is more sensitive to the red-edge region, showed greater consistency in estimating variables related to physiological status, such as protein and fiber content. However, crude protein exhibited a multicausal behavior, with no single index consistently dominating its prediction. Overall, these findings highlight the potential of integrating remote sensing and machine learning as effective tools for pasture monitoring, thereby supporting improved decision-making in livestock production systems. Nevertheless, the extrapolation of these models requires validation across diverse agroecological conditions, as well as the incorporation of additional variables to enhance the robustness and generalizability of the predictions.

Ethical statement

This study focused exclusively on the evaluation of tropical grass species, using agronomic measurements, multispectral remote sensing, and laboratory analysis of plant material. All experimental procedures were carried out in accordance with institutional guidelines for agricultural research and did not require the approval of an ethics committee because they did not involve human participants, biological samples, vertebrate animals, or animal experimentation.

The methodological framework and execution of the study were reviewed and approved under Investment Project CUI N° 2472675: "Mejoramiento de los servicios de investigación y transferencia de tecnología agraria en la estación experimental agraria Baños del Inca en la localidad de Baños del Inca del distrito de Baños del Inca - provincia de Cajamarca - departamento de Cajamarca",” executed by the Instituto Nacional de Innovación Agraria (INIA), Peru.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT to improve the writing and readability of the manuscript, which was then meticulously revised and edited. The authors take full responsibility for the content.

CRediT authorship contribution statement

Josué Tafur-Culqui: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nilton Atalaya-Marin:** Writing – review & editing, Visualization, Validation, Methodology, Data curation. **Darwin Gómes-Fernandez:** Writing – review & editing, Visualization, Validation, Investigation, Data curation. **Victor H. Taboada-Mitma:** Writing – review & editing, Resources, Project administration, Funding acquisition. **Juancarlos Cruz-Luis:** Writing – review & editing, Resources, Project administration, Funding acquisition. **Henri Neyra:** Writing – review & editing, Validation, Methodology. **Janella Y. Anchayhua:** Writing – review & editing, Resources, Funding acquisition. **Rosalía Quichua-Baldeon:** Writing – review & editing, Resources, Funding acquisition. **Teiser Sánchez-Fuentes:** Writing – review & editing, Methodology. **Yadhira M. Olano:** Writing – review & editing, Methodology. **Mauro Barrazueta:** Writing – review & editing, Methodology. **Daniel Tineo:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition. **Malluri Goñas:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition.

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.

Acknowledgment

The authors would like to thank the Instituto Nacional de Innovación Agraria (INIA) through the Investment Project with CUI N° 2472675: "Mejoramiento de los servicios de investigación y transferencia de tecnología agraria en la estación experimental agraria Baños del Inca en la localidad de Baños del Inca del distrito de Baños del Inca - provincia de Cajamarca - departamento de Cajamarca", which financed the execution of the research. They also thank, Gian M. Monteza, Jorge Delgado and Yolmer Dávila for obtaining information for this study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.atech.2026.102229.

Data availability

Upon reasonable request, the corresponding author will provide access to the data used to support the study's findings.

References

[1] P. Jayasinghe, T. Ramilan, D.J. Donaghy, K.G. Pembleton, D.G. Barber, Comparison of nutritive values of tropical pasture species grown in different environments, and implications for livestock methane production: a meta-analysis, Animals 12 (2022) 1806, https://doi.org/10.3390/ani12141806.

[2] S. Reinermann, S. Asam, C. Kuenzer, Remote sensing of grassland production and management—A review, Remote Sens. 12 (2020) 1949, https://doi.org/10.3390/rs12121949.

[3] M.S. Gonçalves, W.R. Ribeiro, D. da Costa Gonçalves, P.A. Guerra Filho, M.M. C. Tanure, J.F.T. do Amaral, A.C. Coser, G. de Oliveira Garcia, E.F. dos Reis, A. R. dos Santos, Tropical grasses adaptability as a sustainable tool for recovering degraded pastures and restricting soil loss: use diagnosis and future scenarios, Environ. Dev. 47 (2023) 100887, https://doi.org/10.1016/j.envdev.2023.100887.

[4] M. Boval, R.M. Dixon, The importance of grasslands for animal production and other functions: a review on management and methodological progress in the tropics, Animal 6 (2012) 748–762, https://doi.org/10.1017/S1751731112000304.

[5] Food and Agriculture Organization of the United Nations (FAO), Las Naciones Unidas declaran 2026 Año Internacional de los Pastizales y los Pastores, FAO (2022). https://www.fao.org/newsroom/detail/un-names-2026-as-the-international-year-of-rangelands-and-pastoralists/es (Accessed March 14, 2025).

[6] Ministerio de Desarrollo Agrario y Riego (MIDAGRI), Perfil productivo pecuario regional, (2025) 2. https://app.powerbi.com/view?r=eyJrIjoiYWM0MDIwYTktNTk3MS00OTc3LThiZTgtZjRmN2ZhMmZlNjVlIiwidCI6IjdmMDg0NjI3LTdmNDAtNDg3OS04OTE3LTk0Yjg2ZmQzNWYzZiJ9&pageName=ReportSection (Accessed December 2, 2025).

[7] G. Defalque, R. Santos, D. Bungenstab, D. Echeverria, A. Dias, C. Defalque, Machine learning models for dry matter and biomass estimates on cattle grazing systems, Comput. Electron. Agric. 216 (2024) 108520, https://doi.org/10.1016/j.compag.2023.108520.

[8] J. Ferner, A. Linstädter, K. Südekum, S. Schmidtlein, International journal of applied earth observation and geoinformation spectral indicators of forage quality in West Africa ' s tropical savannas, Int. J. Appl. Earth Obs. Geoinf. 41 (2015) 99–106, https://doi.org/10.1016/j.jag.2015.04.019.

[9] D. Elavarasan, D. Raj, V. Sharma, A.Y. Zomaya, Forecasting yield by integrating agrarian factors and machine learning models : a survey, Comput. Electron. Agric. 155 (2018) 257–282, https://doi.org/10.1016/j.compag.2018.10.024.

[10] T.B. Shahi, C.Y. Xu, A. Neupane, W. Guo, Machine learning methods for precision agriculture with UAV imagery: a review, Electron. Res. Arch. 30 (2022) 4277–4317, https://doi.org/10.3934/era.2022218.

[11] P. Karmakar, S.W. Teng, M. Murshed, S. Pang, Y. Li, H. Lin, Crop monitoring by multimodal remote sensing: a review, Remote Sens. Appl. Soc. Environ. 33 (2024) 101093, https://doi.org/10.1016/j.rsase.2023.101093.

[12] N. Furnitto, J.M. Ramírez-Cuesta, D.S. Intrigliolo, G. Todde, S. Failla, Remote sensing for pasture biomass quantity and quality assessment: challenges and future prospects, Smart Agric. Technol. 12 (2025) 101057, https://doi.org/10.1016/j.atech.2025.101057.

[13] M.H.M. da Rocha Fernandes, J. de Souza Fernandes Jr, J.M. Adams, M. Lee, R. A. Reis, L.O. Tedeschi, Using sentinel - 2 satellite images and machine learning algorithms to predict tropical pasture forage mass, crude protein, and fiber content, Sci. Rep. 14 (2024) 8704, https://doi.org/10.1038/s41598-024-59160-x.

[14] D.J. Stekhoven, P. Bühlmann, MissForest—Non-parametric missing value imputation for mixed-type data, Bioinformatics 28 (2012) 112–118, https://doi.org/10.1093/bioinformatics/btr597.

[15] K. Rupay T, G. Ampuero T, C. Vela G, C. Angulo V, M. Mathios F, R. Torres S, Evaluación agronómica y nutricional de pastos tropicales de corte sometidos a diferentes frecuencias de corte en Yurimaguas, Amazonia peruana, Rev. Investig. Vet. Del Perú 34 (2023) e24461, https://doi.org/10.15381/rivep.v34i5.24461.

[16] Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), Mapa climático del Perú, (2025). https://www.senamhi.gob.pe/?p=mapa-climatico-del-peru (Accessed November 24, 2025).

[17] A. Huete, K. Didan, W. Van Leeuwen, A. Jacobson, R. Solanos, T. Laing, MODIS Vegetation index (MOD 13) algorithm theoretical basis document, 1999. https://www.semanticscholar.org/paper/MODIS-VEGETATION-INDEX-(-MOD-13-)-ALGORITHM-BASIS-3-Huete-Didan/2204b55a9ad69e8b69d19e88ad1f0e1f81a5d72b (Accessed December 12, 2025).

[18] A.A. Gitelson, Y.J. Kaufman, M.N. Merzlyak, Use of a green channel in remote sensing of global vegetation from EOS- MODIS, Remote Sens. Environ. 58 (1996) 289–298, https://doi.org/10.1016/S0034-4257(96)00072-7.

[19] J. Qi, A. Chehbouni, A.R. Huete, Y.H. Kerr, S. Sorooshian, A modified soil adjusted vegetation index, Remote Sens. Environ. 48 (1994) 119–126, https://doi.org/10.1016/0034-4257(94)90134-1.

[20] E.M. Barnes, T.R. Clarke, S.E. Richards, P.D. Colaizzi, J. Haberland, M. Kostrzewski, P. Waller, R.E. Choi C., T. Thompson, R.J. Lascano, H. Li, M.S. Moran, Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data, Proc. Fifth Int. Conf. Precis. Agric. (2000). https://www.researchgate.net/publication/43256762_Coincident_detection_of_crop_water_stress_nitrogen_status_and_canopy_density_using_ground_based_multispectral_data (Accessed March 12, 2026).

[21] J.W. Rouse Jr., R.H. Haas, J.A. Schell, D.W. Deering, J.C. Harlan, Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation, 1974. https://ntrs.nasa.gov/citations/19740022555 (Accessed November 14, 2025).

[22] C.S.T. Daughtry, C.L. Walthall, M.S. Kim, E. Broun de Clostoun, J.E. McMurtrey III, Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance, Remote Sens. Environ. 74 (2000) 229–239, https://doi.org/10.1016/S0034-4257(00)00113-9.

[23] A.R. Huete, A soil-adjusted vegetation index (SAVI), Remote Sens. Environ. 25 (1988) 295–309, https://doi.org/10.1016/0034-4257(88)90106-X.

[24] A.A. Gitelson, Y.J. Kaufman, R. Stark, D. Rundquist, Novel algorithms for remote estimation of vegetation fraction, Remote Sens. Environ. 80 (2002) 76–87, https://doi.org/10.1016/S0034-4257(01)00289-9.

[25] Association of Official Analytical Chemists (AOAC),AOAC Official Method 950.46, AOAC Official Method 950.46, Loss on drying (moisture) in meat, in: G. W. Latimer Jr. (Ed.), Official Methods of Analysis of AOAC International, Oxford University Press, New York, 2023. 10.1093/9780197610145.003.3431.

[26] Association of Official Analytical Chemists (AOAC), I3-1Method number index. Official Methods of Analysis of AOAC INTERNATIONAL, Oxford University Press, 2023. 10.1093/9780197610145.001.0001.

[27] Association of Official Analytical Chemists (AOAC), AOAC Official Method 984.13, Protein (Crude) in animal feed and pet food: copper catalyst kjeldahl method, in: G. W. Latimer Jr. (Ed.), Official Methods of Analysis of AOAC International, 22nd, Oxford University Press, 2023. 10.1093/9780197610145.003.1401.

[28] Association of Official Analytical Chemists (AOAC), AOAC Official Method 2003.05, Crude fat in feeds, cereal grains, and forages: randall/Soxtec/Diethylether Extraction-submersion method, in: G.W. Latimer Jr. (Ed.), Official Methods of Analysis of AOAC International, 22nd, Oxford University Press, 2023. 10.1093/9780197610145.003.1414.

[29] Association of Official Analytical Chemists (AOAC), AOAC Official Method 962.09, Fiber (Crude) in animal feed and pet food: ceramic fiber filter method, in: G. W. Latimer Jr. (Ed.), Official Methods of Analysis of AOAC International, 22nd, Oxford University Press, 2023. 10.1093/9780197610145.003.1416.

[30] Association of Official Analytical Chemists (AOAC), AOAC Official Method 942.05 Ash of Animal Feed, in: G.W. Latimer Jr. (Ed.), Official Methods of Analysis of AOAC International, 22nd, Oxford University Press, New York, 2023. 10.1093/9780197610145.003.1389.

[31] C.R. Harris, K.J. Millman, S.J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. Van Kerkwijk, Array programming with NumPy, Nature 585 (2020) 357–362, https://doi.org/10.1038/s41586-020-2649-2.

[32] W. McKinney, Data structures for statistical computing in Python, in: Proc. 9th Python Sci. Conf., 2010, pp. 56–61, https://doi.org/10.25080/majora-92bf1922-00a.

[33] J.D. Hunter, Matplotlib: a 2D graphics environment, Comput. Sci. Eng. 9 (2007) 90–95, https://doi.org/10.1109/MCSE.2007.55.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf. Accessed December 8, 2025.

[35] M.L. Waskom, seaborn: statistical data visualization, J. Open Source Softw. 6 (2021) 3021, https://doi.org/10.21105/joss.03021.

[36] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C.N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D.A. Nicholson, D.R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G.A. Price, G.L. Ingold, G.E. Allen, G.R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J.T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J.V. de Miranda Cardoso, J. Reimer, J. Harrington, J.L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N.J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P.A. Brodtkorb, P. Lee, R.T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T.J. Pingel, T.P. Robitaille, T. Spura, T.R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y.O. Halchenko, Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods 17 (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.

[37] M. Aton, D. McDonald, J. Cañardo Alastuey, R. Azom, P. Batra, V. Bezshapkin, E. Bolyen, A. Cagle, J.G. Caporaso, J.W. Debelius, K. Gorlick, N. Hamsanipally, L. Hunger, A. Keluskar, D. Liao, Y.Y. Lu, J.A. Navas-Molina, A. Pitman, J. R. Rideout, A. Sazonov, B. Sathappan, K. Schwarzberg Lipson, I. Sfiligoi, C. Tapo, Y. Vázquez-Baeza, Z. Wu, Z.Z. Xu, M.S. Ye, J. Zhao, R. Knight, J.T. Morton, Q. Zhu, Scikit-bio: a fundamental Python library for biological omic data analysis, Nat. Methods 23 (2026) 274–276, https://doi.org/10.1038/s41592-025-02981-z.

[38] M. Terpilowski, scikit-posthocs: pairwise multiple comparison tests in Python, J. Open Source Softw. 4 (2019) 1169, https://doi.org/10.21105/joss.01169.

[39] T. Weitkamp, P. Karimi, Evaluating the effect of training data size and composition on the accuracy of smallholder irrigated agriculture mapping in Mozambique using

remote sensing and machine learning algorithms, Remote Sens. 15 (2023) 3017, https://doi.org/10.3390/rs15123017.
[40] K.G. Liakos, P. Busato, D. Moshou, S. Pearson, Machine learning in agriculture: a review, Sensors 18 (2018) 2674, https://doi.org/10.3390/s18082674.
[41] J. Wijesingha, T. Astor, D. Schulze-Brüninghoff, M. Wengert, M. Wachendorf, Predicting forage quality of grasslands using UAV-borne imaging spectroscopy, Remote Sens. 12 (2020) 126, https://doi.org/10.3390/rs12010126.
[42] H. Wen, Y. Zhang, X. Wang, R. Wang, W. Wu, Inversion study of the meadow steppe above-ground biomass based on ground and airborne hyperspectral data, Geocarto Int. 39 (2024), https://doi.org/10.1080/10106049.2024.2370304.
[43] C. Santhi, J.G. Arnold, J.R. Williams, W.A. Dugas, R. Srinivasan, L.M. Hauck, Validation of the swat model on a large rwer basin with point and nonpoint sources, J. Am. Water Resour. Assoc. 37 (2001) 1169–1188, https://doi.org/10.1111/j.1752-1688.2001.tb03630.x.
[44] M.W. Van Liew, J.G. Arnold, J.D. Garbrecht, Hydrologic simulation on agricultural watersheds: choosing between two models, Trans. Am. Soc. Agric. Eng. 46 (2003) 1539–1551, https://doi.org/10.13031/2013.15643.
[45] P.D. Jamieson, J.R. Porter, D.R. Wilson, A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand, Field Crops Res. 27 (1991) 337–350, https://doi.org/10.1016/0378-4290(91)90040-3.
[46] D.W. Lamb, M. Steyn-Ross, P. Schaare, M.M. Hanna, W. Silvester, A. Steyn-Ross, Estimating leaf nitrogen concentration in ryegrass (Lolium spp.) pasture using the chlorophyll red-edge: theoretical modelling and experimental observations, Int. J. Remote Sens. 23 (2002) 3619–3648, https://doi.org/10.1080/01431160110114529.
[47] J. Serrano, S. Shahidian, J. Marques da Silva, Monitoring seasonal pasture quality degradation in the Mediterranean Montado Ecosystem: proximal versus remote sensing, Water 10 (2018) 1422, https://doi.org/10.3390/w10101422.
[48] M.D.L. Bezerra, E.M. Bonfim-Silva, T.J.A. da Silva, A.P.F. Ferraz, A.P.A. B. Damasceno, Phytometric characteristics and chlorophyll index of “paiaguás” grass (Urochloa brizantha) as a function of wood ash doses and soil water stress, Aust. J. Crop Sci. 13 (2019) 1883–1891, https://doi.org/10.21475/ajcs.19.13.11.p2100.
[49] A. El Mouttaqi, I. Mnaouer, A. Nilahyane, D.S. Ashilenje, E. Amombo, M. Belcaid, M. Ibourki, K. Lazaar, A. Soulaimani, K.P. Devkota, L. Kouisni, A. Hirich, Influence of cutting time interval and season on productivity, nutrient partitioning, and forage quality of blue panicgrass (Panicum antidotale Retz.) under saline irrigation in southern region of Morocco, Front. Plant Sci. 14 (2023) 1186036, https://doi.org/10.3389/fpls.2023.1186036.
[50] Y. Wekgari, F. Dereba, N. Gamachu, Determination of cutting frequency for optimum herbage yield and nutritive value of Desho grass (Pennisetum glaucifolium) in western Oromia, Ethiopia, Heliyon 10 (2024) e38132, https://doi.org/10.1016/j.heliyon.2024.e38132.
[51] M. Luciano de Melo, J.A. Martuscello, D.M. da Fonseca, C. Mistura, R.V. de Morais, J.I. Ribeiro Jr, Perfilhamento, acúmulo de forragem e composição bromatológica do capim-braquiária adubado com nitrogênio, Rev. Bras. Zootec. 38 (2009) 1675–1684, https://doi.org/10.1590/S1516-35982009000900006.
[52] J.A. Postma, V.L. Hecht, K. Hikosaka, E.A. Nord, T.L. Pons, H. Poorter, Dividing the pie: a quantitative review on plant density responses, Plant. Cell Environ. 44 (2021) 1072–1094, https://doi.org/10.1111/pce.13968.
[53] S. Tan, A.K. Mortensen, X. Ma, B. Boelt, R. Gislum, Assessment of grass lodging using texture and canopy height distribution features derived from UAV visual-band images, Agric. For. Meteorol. 308-309 (2021) 108541, https://doi.org/10.1016/j.agrformet.2021.108541.
[54] S. Arjasakusuma, S.S. Kusuma, S. Phinn, Evaluating variable selection and machine learning algorithms for estimating forest heights by combining lidar and hyperspectral data, ISPRS Int. J. Geo-Inf. 9 (2020) 507, https://doi.org/10.3390/ijgi9090507.
[55] L. Breiman, Randon Forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1007/978-3-030-62008-0_35.
[56] B.K. Kisambo, O.V. Wasonga, O.K. Kipchirchir, G.N. Karuru, E.C. Kirwa, Forage yields and quality of Cenchrus ciliaris and Panicum maximum ecotypes under varied harvest intervals in a semi-arid environment in Kenya, Int. J. Trop. Drylands 7 (2023) 102–111, https://doi.org/10.13057/tropdrylands/t070205.
[57] E. Prudnikova, I. Savin, G. Vindeker, P. Grubina, E. Shishkonakova, D. Sharychev, Influence of soil background on spectral reflectance of winter wheat crop canopy, Remote Sens. Artic. 11 (2019) 1932, https://doi.org/10.3390/rs11161932.
[58] M.H.M. da Rocha Fernandes, J. de Souza Fernandes Jr, J.M. Adams, M. Lee, R. A. Reis, L.O. Tedeschi, Using sentinel-2 satellite images and machine learning algorithms to predict tropical pasture forage mass, crude protein, and fiber content, Sci. Rep. 14 (2024) 8704, https://doi.org/10.1038/s41598-024-59160-x.
[59] I.L. Bretas, D.S.M. Valente, F.F. Silva, M.L. Chizzotti, M.F. Paulino, A.P. D'Áurea, D. S.C. Paciullo, B.C. Pedreira, F.H.M. Chizzotti, Prediction of aboveground biomass and dry-matter content in brachiaria pastures by combining meteorological data and satellite imagery, Grass Forage Sci. 76 (2021) 340–352, https://doi.org/10.1111/gfs.12517.
[60] N.R. Kabange, S. Lee, D. Shin, J. Lee, Y. Kwon, J. Kang, J. Cha, H. Park, S. Alibu, J. Lee, Multiple facets of nitrogen : from atmospheric gas to indispensable agricultural input, Life 12 (2022) 1272, https://doi.org/10.3390/life12081272.
[61] Q. Wang, S. Li, J. Li, D. Huang, The utilization and roles of nitrogen in plants, Forests 15 (2024) 1191, https://doi.org/10.3390/f15071191.