scientific reports

Check for updates

OPEN

# Machine learning models for crude protein prediction in Tamani grass pastures

Gabriela Oliveira de Aquino Monteiro[1]✉, Gelson dos Santos Difante[2], Denise Baptaglin Montagner[3], Valéria Pacheco Batista Euclides[3], Marina Castro[4], Jéssica Gomes Rodrigues[2], Marislayne de Gusmão Pereira[2], Juliana Caroline Santos Santana[2,5], Luis Carlos Vinhas Itavo[2], Rafael Torres Nantes[6], Jecelen Adriane Campos[6], Anderson Bessa da Costa[6] & Edson Takashi Matsubara[6]

**Understanding forage quality is essential for meeting animal demands and optimizing production. This study aimed to: (i) test the applicability of machine learning models with tabular data such as climate variables, light interception (LI), nitrogen dose (N dose), interval between grazing (GI), and pre- (HPRE) and post-grazing height (HPOST) to predict leaf crude protein (CP) content of tamani grass pastures; (ii) identify which variables contribute most to CP prediction. A set of 90 instances was used with 80% for training and validation and 20% for testing. The hyperparameters were adjusted with *grid-search* on the training set. We tested Linear Regression (LR), Multilayer Perceptron (MLP), Decision Trees (DT), Random Forest (RF), and XGBoost. The MLP ($r = 0.75$, $R^2 = 44.18\%$, MAE = 1.55), RF ($r = 0.78$, $R^2 = 49.07\%$, MAE = 1.59) and XGBoost ($r = 0.78$, $R^2 = 56.65\%$ MAE = 1.45) models presented the best prediction results ($p < 0.001$). The variables most important in predicting CP content were GI, followed by N dose, HPRE and HPOST. XGBoost outperformed other tested models ($p < 0.001$). Tabular data, including N dose, GI, HPRE, HPOST, LI, and climatic variables, is a viable alternative for predicting CP. In conclusion, the results of this study suggest that management practices may have a greater influence on the chemical composition of Tamani grass than environmental conditions, although further research with larger and more diverse datasets is needed to confirm these findings. Link to the API: https://github.com/GabrielaAquino93/Project-BiomassCalculator.**

**Keywords** Machine learning, *Panicum maximum*, Pasture management, Precision livestock farming

Pastures occupy a substantial portion of global agricultural land, covering approximately 3.2 billion hectares worldwide, which represents a considerable fraction of the Earth's land area (approximately 29% of the planet's surface)[1]. In addition to serving as a primary feed source for livestock, they provide essential ecosystem services, including biomass production for energy and carbon sequestration[2,3]. Using tropical grasses, combined with proper pasture and grazing management, is crucial for optimizing forage resources efficiently. To achieve good results, grazing management must be carried out to promote animal productivity and pasture sustainability, based on the relationship between nutritional value, animal requirements and forage mass[4].

Forage mass refers to the amount of forage available in a given area, expressed in kg of dry matter per hectare. The nutritional value of forage, considering the nutritional needs of ruminants, is determined by its digestibility and chemical composition, which includes the levels of crude protein, structural and non-structural carbohydrates[5]. The chemical composition of forage plants can be influenced by the species or cultivar, soil and climate conditions and management adopted[6].

Based on forage mass, animals select their diet, and daily weight gain is influenced by the quantity and quality of forage consumed[7]. Pastures with low nutritional value are often characterized by high fiber content and reduced protein levels, typical of forage plants at an advanced stage of maturity[8] or those that have not been properly managed. This can lead to changes in intake, compromising the ability to meet the animal's nutritional

[1] Departament of Animal Science, São Paulo State University "Júlio de Mesquita Filho", JABOTICABAL, Brasil. [2]Department of Animal Science, Federal University of Mato Grosso do Sul (UFMS), Campo Grande, MS 79070-900, Brazil. [3]Embrapa Beef Cattle, Campo Grande, MS 79106-550, Brazil. [4]CIMO, LA SusTEC, Polytechnic Institute of Bragança, Alameda de St Apolónia, Bragança 5300-253, Portugal. [5] Graduate Program in Animal Production, Federal University of Rio Grande do Norte, Macaiba, Brasil. [6]Department of Computer, Federal University of Mato Grosso do Sul (UFMS), Campo Grande, MS 79070-900, Brazil. ✉email: gabrielaoliveiraaquino@gmail.com

requirements. Therefore, assessing forage quality is essential for meeting animal needs and improving the efficiency of the production system.

Methods for assessing pasture nutritional value are traditionally destructive, relying on laboratory chemical analyses and near-infrared reflectance spectroscopy (NIRS), which require time, labor, and financial investment[9,10]. Therefore, developing methods and techniques that provide accurate information with lower labor requirements, reduced costs, and faster results is essential for effective decision-making in pasture production systems.

Statistical models are widely used to predict agricultural productivity[11]. However, these models assume that the data are independent and identically distributed (iid) and require prior knowledge about their distribution for adequate modeling. In contrast, machine learning models have been increasingly used to estimate crop yield, as they do not depend on these rigid assumptions and can identify patterns directly from the available data[12]. This flexibility reduces the need for prior knowledge about the data structure, making these models more adaptable to different contexts. Furthermore, studies indicate that machine learning often presents greater accuracy in estimates when compared to traditional statistical methods[13–15].

Therefore, machine learning models have been widely applied in agriculture, including crops such as eucalyptus, sugarcane, wheat and corn[12,16–18]. However, their application is still limited in pasture management, especially for cultivars of *Panicum maximum* (Syn. *Megathyrsus maximus*), which is widely used in Brazil in intensified systems owing to high yield and nutritional value[19].

Recent studies on forage mass and crude protein content prediction use machine learning with data from satellite images, drones, and laboratory analyses, which can make these approaches expensive and difficult to apply on a large scale[20,21]. To date, no studies have been found exploring the prediction of leaf crude protein without using images. In addition, there are no records of studies that use variables such as climate data, nitrogen doses, grazing interval, and pre- and post-grazing heights to estimate crude protein (CP) content in Tamani grass pastures. The present work follows a different approach, proposing a simple and accessible, but also effective, solution for this prediction. An easily obtainable data set is used, eliminating the need for remote sensors and laboratory analyses, in addition to algorithms that do not require high computational power for training and evaluation, making the proposal viable for application in production systems.

The contributions of this work are:

- The proposal for predicting CP using low-cost and easily accessible information, which enhances its applicability and consequently the impact of this research. CP, an important element of the chemical composition of pasture, is obtained from climate data, nitrogen fertilization, interval between grazing and pre- and post-grazing height. None of these characteristics require drone flight or laboratory analysis. The information used can be obtained by recording and collecting information that is usually already part of most experimental field tests.
- The correlation analysis indicates that nitrogen dose (N dose), rainy season, pre-grazing height (HPRE), post-grazing height (HPOST), grazing interval (GI) and precipitation (PREC) have significant correlations (Tukey's test with $p < 0.05$), and these can be used as relevant characteristics for predicting CP.
- Robust experimental evaluation using traditional machine learning algorithms demonstrates that it is possible to obtain promising results with the proposed approach. The best result was obtained with the XGBoost algorithm, presenting a Pearson correlation (r) of 0.78 and a coefficient of determination ($R^2$) of 56.65%.

## Materials and methods

### Data acquisition

The data were collected from an experiment conducted at Embrapa Gado de Corte in Campo Grande, MS (20°27'S and 54°37'W, at an altitude of 530 m), from October 2020 to April 2022. The regional climate follows the Köppen classification as a tropical rainy savanna (Aw subtype), characterized by a seasonal rainfall distribution. Temperature, precipitation, and solar radiation data for the experimental period were obtained from the Embrapa Gado de Corte meteorological station, located approximately 2.4 km from the experimental area. The soil in the experimental area is classified as Nitosol Vermelho Distrófico Latosólico[22], characterized by a clayey texture.

Data were collected from a 0.96 ha area planted with Tamani grass, divided into four blocks, with each block further subdivided into four paddocks (0.06 ha each). Data were collected across the following seasons: 20/21 Rainy (grazing cycles from December 2020 to March 2021), 21/22 Rainy (grazing cycles from December 2021 to March 2022), and the transition periods 21 Rainy-Dry (grazing cycles in April 2021), 21 Dry-Rainy (grazing cycles in November 2021), and 22 Rainy-Dry (grazing cycles in April 2022), ensuring a broad range of variability to support the development of robust models that would be applicable year-round.

Pastures were managed according to the following treatments, which consisted of two light interception (LI) levels (90 and 95%) and two nitrogen (N) rates (80 and 240 kg $ha^{-1}$ $year^{-1}$) resulting in four treatments: 90LI80N, 95LI80N, 90LI240N, and 95LI240N. Light interception (LI) refers to the proportion of incident photosynthetically active radiation (PAR) intercepted by the plant canopy, measured with a canopy analyzer (AccuPAR Linear PAR/LAI Ceptometer, Model PAR-80; Decagon Devices). LI was monitored weekly in each paddock and, as the pre-grazing target approached, measurements were taken daily. LI was assessed at 10 randomly selected points per paddock using a canopy analyzer. At the same time, pre-grazing canopy height was measured at 20 randomly selected points per paddock using a centimeter-graduated ruler. Post-grazing height was recorded following the same procedure after the animals were removed. Fertilization was carried out after the animals left the paddocks, with the 80 kg $ha^{-1}$ $year^{-1}$ dose applied in two equal splits and the 240 kg $ha^{-1}$ $year^{-1}$ dose in four equal splits.

Grazing intensity was maintained at a fixed level, corresponding to 50% of the initial pasture height. The grazing interval (GI) was obtained by summing the days each paddock remained without animals. To determine

the crude protein content of the leaf, pre-grazing forage mass was estimated by cutting three samples at ground level, each within a 1 m × 1 m metal frame, positioned at sites representative of the mean canopy height. These samples were weighed and divided into two subsamples: one for determining dry matter content and the other for separating forage morphological components (leaf, stem, and dead material). Leaf blade samples were subsequently ground to 1 mm and analyzed using a near-infrared reflectance spectroscopy (NIRS) system, following the procedures described by Campos et al.[23].

The NIRS analyses were conducted using an FT-NIR Büchi NIRFlex 500 spectrometer (Büchi, Switzerland), operated with the NIRS Operator software and using standard quartz-bottom minicuvettes (Foss, Denmark; Aquartzo, Brazil). Spectra were collected between 4,000 and 9,000 $cm^{-1}$. Calibration procedures, including preprocessing steps and model development, are detailed in Campos et al.[23], who provide a full description of the calibration strategy, preprocessing routines, and associated prediction error metrics for the set of tropical forage samples used in this study.

Briefly, the spectra were processed and calibrated in the NIRCal® 1.6 software (Büchi, Switzerland) using partial least squares (PLS) regression. The calibration dataset consisted of 1.026 forage samples collected between 2015 and 2020, including multiple seasons and samples from *Brachiaria* and *Panicum* species. Reference values for crude protein were obtained through wet chemistry and used to construct the calibration models. Calibration performance was satisfactory, with Q-values greater than 0.60, meeting the robustness criteria recommended by Büchi[23].

The variables selected as inputs for the tested models were chosen based on ease of acquisition when managing pastures. A minimum of one forage sample was collected from each paddock, within each block, during every grazing season, resulting in an initial total of 4 blocks × 4 paddocks × 5 harvests/seasons = 80 samples. In some treatments (particularly those managed at 90% LI with 240 $kg{\cdot}ha^{-1}{\cdot}year^{-1}$ of nitrogen) multiple grazing cycles occurred within the same season, increasing the total number of physical samples to 90. The dataset used in this study was compiled by integrating leaf crude protein content, management heights, grazing intervals, and climatic variables, comprising 14 variables (one dependent and 13 independent) across 90 instances. Table 1 summarizes the dataset, including each variable's definition, unit of measurement, mean, and standard deviation (for numerical variables only). The nominal categorical variable "season" was converted into a set of binary variables using the One-Hot Encoding technique. The mean values of these binary variables were calculated to characterize their distribution (Table 1).

## Machine learning

Figure 1 presents the flowchart for data collection, processing, training, and validation of machine learning models used to predict crude protein in Tamani grass pasture leaves.

Machine learning models aim to predict Crude Protein (CP) content. The variables selected as input were: season, climate data (mean temperature, solar radiation and precipitation), nitrogen (N) dose (80 and 240 kg/ha/year of N), light interception (LI), grazing interval (GI) and pre- (HPRE) and post-grazing height (HPOST) were included as training data for the models. The target attribute is leaf CP content (Table 1).

The implementation and training of the models were performed using Pandas and Scikit-learn[24]. From a set of 90 samples, 72 samples (80%) were separated for training and validation, and 18 (20%) for testing. Model evaluation was performed using 5-fold cross-validation with 100 iterations, and hyperparameter tuning was conducted using *grid search* on the training set.

Five machine learning models were tested: (i) Linear Regression (LR), (ii) Multilayer Perceptron (MLP), (iii) Decision Trees (DT), (iv) Random Forest (RF), and (v) XGBoost. Table 2 details the evaluated models, including the hyperparameters used and their respective variation ranges.

| Variable (symbol) | Meaning | Unit | Experimental mean | Range | CV (%) |
|---|---|---|---|---|---|
| 20/21 Rainy | Season | Binary var. (1/0) | 0.28 (±0.45) | [0; 1] | – |
| 21 Rainy-Dry | Season | Binary var. (1/0) | 0.13 (±0.0.33) | [0; 1] | – |
| 21 Dry-Rainy | Season | Binary var. (1/0) | 0.15 (±0.36) | [0; 1] | – |
| 21/22 Rainy | Season | Binary var. (1/0) | 0.38 (±0.49) | [0; 1] | – |
| 22 Rainy-Dry | Season | Binary var. (1/0) | 0.06 (±0.23) | [0; 1] | – |
| TEMP | Average temperature | Degrees °C | 27.8 (±1.6) | [22.72; 30.35] | 5.72 |
| RAD | Solar radiation | $KJ/m^2$ | 1589 (±378.5) | [673.79; 2184.90] | 23.82 |
| PREC | Precipitation | mm | 229.8 (±195.3) | [0; 1136.4] | 84.98 |
| N DOSE | Nitrogen dose | $kg\ ha^{-1}\ year^{-1}$ | 181.3 (±77.5) | [80; 240] | 42.75 |
| LI | Light interception | % | 91.7 (±2.4) | [90; 95] | 2.60 |
| GI | Grazing interval | days | 31.9 (±13.5) | [12; 69] | 121.43 |
| HPRE | Pre-grazing height | cm | 34.6 (±6.4) | [24.5; 51] | 18.82 |
| HPOST | Post-grazing height | cm | 17.2 (±3.3) | [11; 26] | 19.15 |
| CP | Crude protein | % of DM | 12.52 (±3.0) | [6.6; 18.51] | 24.06 |

**Table 1**. Description of the variables that make up the collected data set. There are 14 variables, 5 categorical variables and 9 numerical variables. The variable to be predicted is crude protein (CP). *Var.* variable, *CV* coefficient of variation, *DM* dry matter.

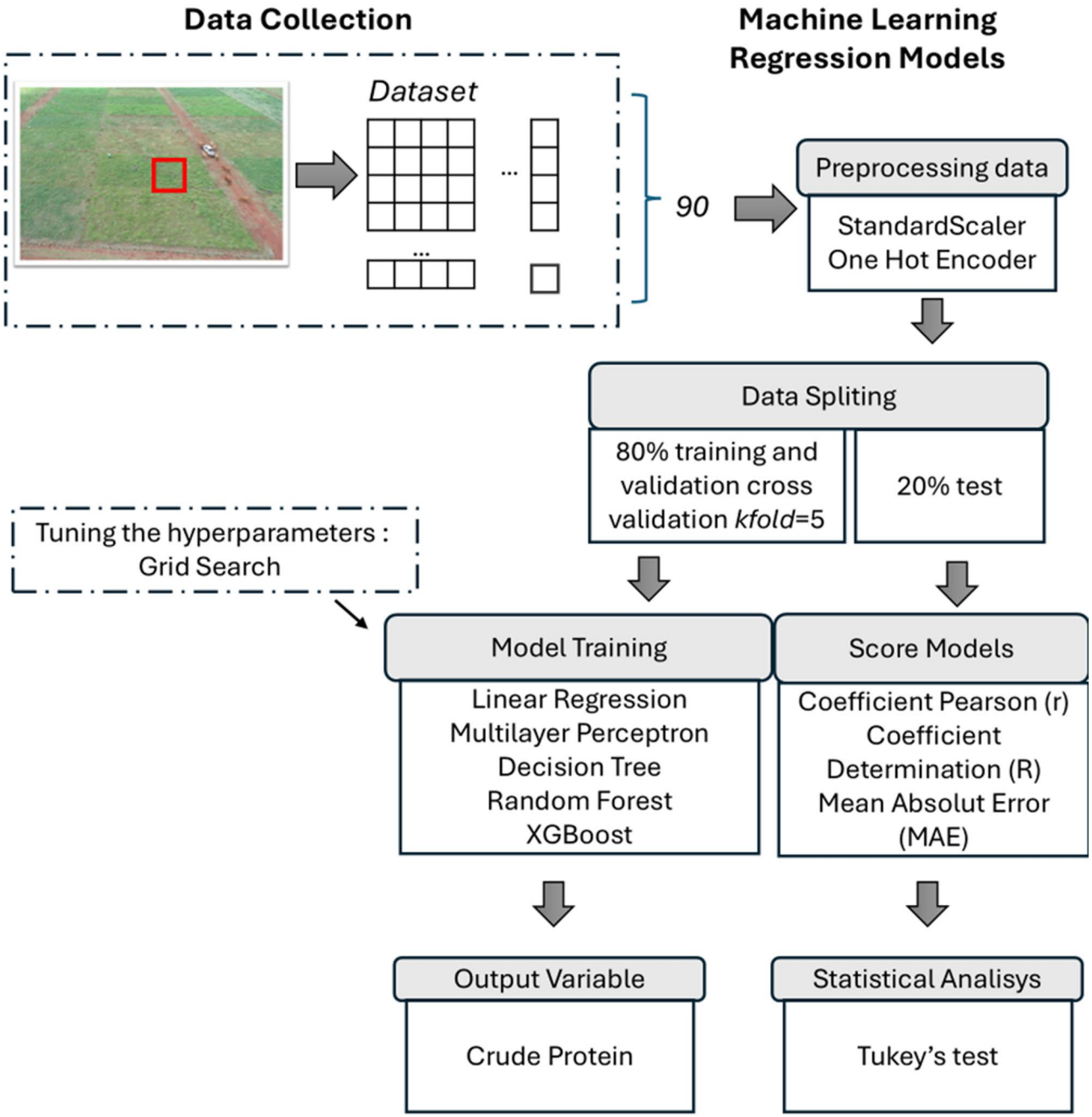


**Fig. 1**. Flowchart of data collection, processing, training, validation and testing of machine learning models for predicting crude leaf protein.

Linear Regression (LR) is a traditional predictive technique used as a baseline for comparing machine learning models. If a linear relationship exists between the independent and dependent variables, good performance can be achieved even with a small dataset. However, when the relationship is nonlinear, performance tends to be unsatisfactory. This model employs the Akaike criterion for model selection and can handle weighted instances (Table 2)[25].

For neural networks, the Multilayer Perceptron (MLP)[26] model was employed as a regressor, utilizing backpropagation to train a multilayer perceptron, for instance, prediction. The MLP configuration included a single hidden layer with 100 neurons, the tanh activation function, and an adaptive learning rate ($\alpha = 0.0001$). The model was trained using Stochastic Gradient Descent (SGD) with a maximum of 5000 iterations (Table 2).

Decision Trees (DT) are fast-learning decision tree models that construct regression/classification trees using information gain and variance reduction, with error pruning through backfitting. DTs classify numeric attributes only once[28]. While these models are simple and interpretable, their performance may be constrained by high variance and a tendency to overfit. The model was configured with a maximum depth of 40, three variables per split, eight samples per leaf, and node splitting occurring when at least four samples were present (Table 2).

| Experiment | Model | Hyperparameters | Interval | References |
|---|---|---|---|---|
| #1 | LR | Copy X: true, fit intercept: true, n jobs: 1, positive: false | Copy X: [True, False], fit_intercept: [True, False], n_jobs: [1], positive: [True, False] | 25 |
| #2 | MLP | Activation: tanh, Alpha: 0.0001, hidden layer sizes: (100,), learning rate: adaptative, max iter: 5000, solver: sgd | Alpha: [0.001, 0.01], hidden layer sizes: [(100,), (100, 50,), (100, 100,),], learning rate: [constant, adaptive], max iter: [10000], solver: [adam] | 26 |
| #3 | RF | Criterion: Absolute error, Max depth: 1000, max features: 2, min samples leaf: 2, min samples split: 2, n estimators: 200 | Max depth: np.arange(2, 100, 40), max features: np.arange (2, 10, 2), min samples leaf: np.arange (2, 10, 2), min samples split: np.arange (2, 8, 2), n estimators: [200, 600, 1000] | 27 |
| #4 | DT | Criterion: Absolute error, Max depth: 40, max features: 3, min samples leaf: 8, min samples split: 4 | Max depth: [10, 40, 80, None], max features: np.arange (1, 14, 2), min samples leaf: [2, 4, 8, 16], min samples split: [2, 4, 8] | 28 |
| #5 | XGBoost | Colsample bytree: 0.6, learning rate: 0.1, max depth: 2, n estimators: 200, reg alpha: 0.6, reg lambda: 2, subsample: 0.6 | Colsample bytree: [0.6, 0.8, 1.0], learning rate: [0.001, 0.01, 0.1], max depth: np.arange(2, 8, 2), n estimators: [200, 600, 1000], reg alpha: [0, 0.2, 0.6], reg lambda': [1, 1.5, 2] | 29 |

**Table 2**. Machine learning models and experimental setups used for prediction. *LR* linear regression, *MLP* multilayer perceptron, *RF* random forest, *DT* decision tree.

Random Forest (RF)[27] is an ensemble method based on Bootstrap Aggregating (bagging), where multiple decision trees are trained using different data samples and subsets of variables. This approach reduces variance and improves model generalization. The final prediction is obtained by averaging tree outputs in regression tasks or by majority voting in classification tasks. RF was configured with 200 trees and a maximum depth of 1000. Each node was considered a maximum of two variables for splitting, with a minimum of two samples required to perform a split and at least two samples required to form a leaf. Feature importance was determined using the mean decrease in impurity, which quantifies the contribution of each variable to node splits across the trees. This metric is based on node impurity, measured using mean squared error in regression tasks.

Despite its robustness, more advanced methods, such as XGBoost, further improve performance by incorporating boosting techniques to optimize model learning. The model uses a decision tree for regression tasks with division of data into smaller categories according to different input feature limits. These divisions create a tree-like structure, where the root represents an initial limit for data division, internal nodes indicate internal divisions, and leaves represent the final output of the model[29]. XGBoost hyperparameter tuning was configured with shallow trees of maximum depth of 2, with 200 iterations and a learning rate of 0.1 with 60% of the data and 60% of the features per tree (Table 2). The importance of the attributes was calculated using XGBoost based on the total gain, which measures the importance of a variable based on the total gain of information it provides across all trees in the model.

## Statistical analysis

A Pearson correlation analysis was performed between the predictor variables and the target variable, CP. The performance of the machine learning models was evaluated using the Pearson correlation coefficient (r), coefficient of determination ($R^2$), and mean absolute error (MAE), calculated by comparing predicted and observed values in the test set. The average values from 100 repetitions (folds) were computed with Random *State = None and Shuffle = True* for each performance metric. The results were then subjected to Tukey's test at a 5% significance level.

The experiments were run on an Intel® Core™ i5 CPU with 8 Gb of RAM, using hyperparameters defined in the scikit-learn library. This information is provided to ensure reproducibility and to indicate that the proposed methodology can be implemented on standard computing resources.

## Results

The experiments were structured into four evaluations aimed at answering the following research questions (Research Questions - RQ):

- *RQ 1: How do the attributes relate linearly?*
- *RQ 2: Which of the classic ML algorithms is the most suitable for predicting crude protein?*
- *RQ 3: What are the most relevant variables in the best-performing models?*

The following sections present the corresponding experiments and evaluations conducted to answer each of these questions.

## Answering RQ1: how do the attributes relate linearly?

To measure linear correlation, we used the Pearson correlation coefficient (r), which quantifies the direction and strength of the linear relationship between two variables[30]. The Pearson coefficient ranges from − 1 to 1, where the sign indicates the direction of the relationship (positive or negative), and the absolute value represents the strength of the correlation. Its formula is as follows:

$$r = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}$$

Where Xi represents the values of variable X in the sample, $\underline{X}$ is the mean of X, Yi represents the values of variable Y in the sample and $\underline{Y}$ is the mean of Y.

Figure 2 presents the Pearson correlation coefficients between the variables in the dataset. CP exhibited a moderate negative correlation with HPRE ($r = -0.49$), HPOST ($r = -0.46$), and the 20/21 Rainy season ($r = -0.43$), a weak negative correlation with GI ($r$=-0.23) and PREC ($r$=-0.33) and a moderate positive correlation with N dose ($r = 0.48$) and 21/22 Rainy ($r = 0.40$) (Fig. 2).

A strong and positive correlation was observed between 21 Dry-Rainy and GI ($r = 0.95$) and PREC ($r = 0.79$), and a negative and weak correlation was observed with 20/21 Rainy ($r$=-0.26) and 21/22 Rainy ($r$=-0.33). The 21 Rainy-Dry season showed a weak and negative correlation with RAD ($r$=-0.29), 20/21 Rainy ($r$=-0.25), and 21/22 Rainy ($r$=-0.31). 21/22 Rainy showed a weak and negative correlation with HPRE ($r = 0.29$), HPOST ($r$=-0.35), GI ($r$=-0.32), PREC ($r$=-0.30), moderate negative with 20/21 Rainy ($r$=-0.49) and weak positive with TEMP ($r = 0.33$) and RAD ($r = 0.31$). A moderate and positive correlation was observed between 20/21 Rainy and HPRE ($r = 0.59$) and HPOST ($r = 0.66$) and a weak negative correlation with GI ($r$=-0.24) and TEMP ($r$=-0.35) (Fig. 2).

RAD showed a strong and positive correlation with TEMP ($r = 0.75$). A strong and positive correlation was observed between PREC and GI ($r = 0.86$). HPOST showed a moderate and positive correlation with LI ($r = 0.53$) and a strong and positive correlation with HPRE ($r = 0.9$). HPRE showed a moderate and positive correlation with LI ($r = 0.64$) (Fig. 2).

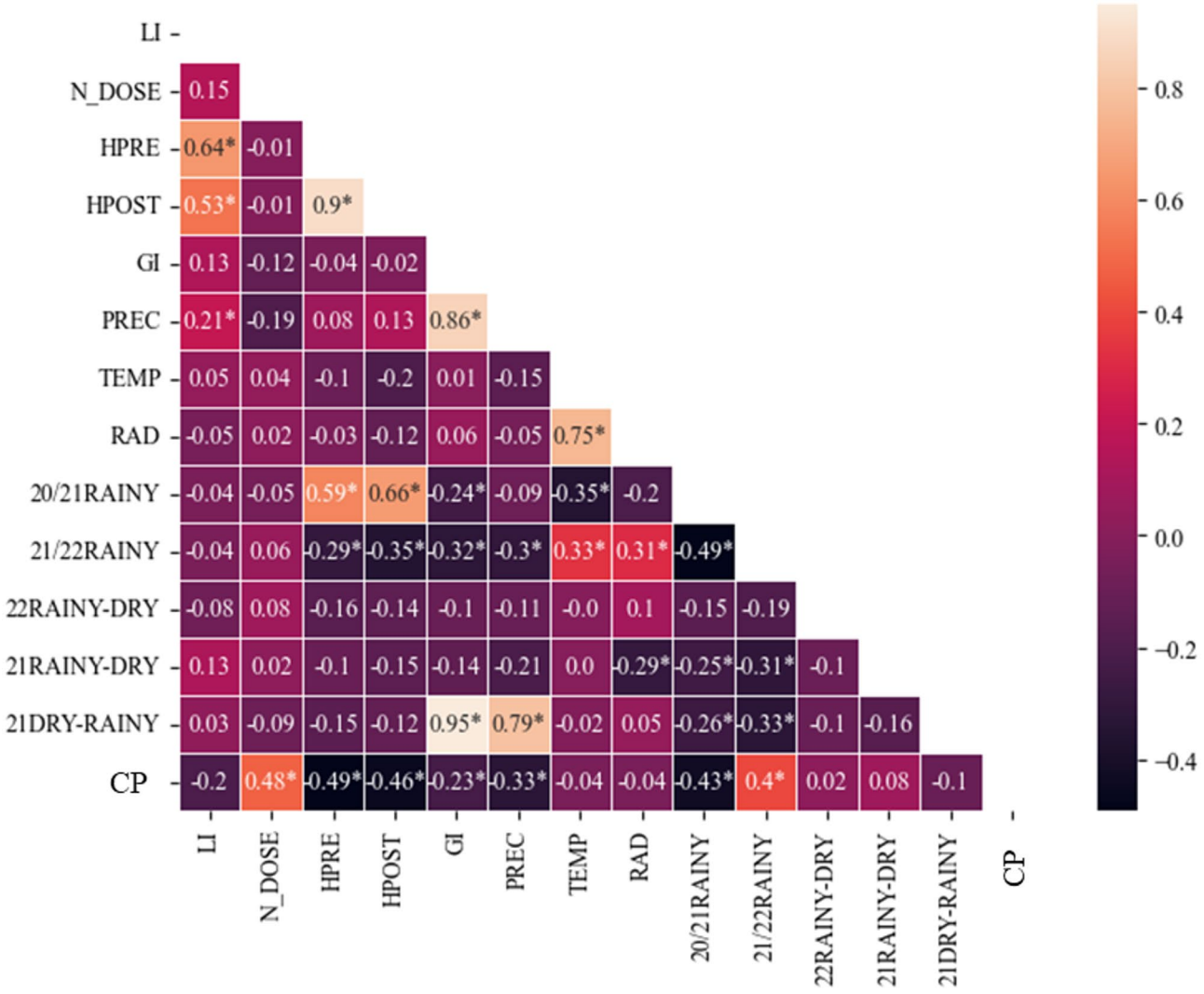


**Fig. 2**. Correlation coefficients of light interception (LI), nitrogen dose (N_DOSE), seasons, pre-grazing height (HPRE), post-grazing height (HPOST), grazing interval (GI), precipitation (PREC), temperature (TEMP), solar radiation (RAD) and crude leaf protein (CP). Correlations with an asterisk are significant by the Tukey test ($p < 0.05$).

## Answering RQ2 - Which of the classic ML algorithms is the most suitable for predicting crude protein?

Figure 3 presents the scatter plots comparing observed and predicted values for each tested model. The LR model showed some good fit; however, it also exhibited overestimated and underestimated values with large error amplitudes and outliers (Fig. 3A). In the DT model, the predicted values were highly dispersed from the actual values (Fig. 3B), confirming the model evaluation results (Fig. 4). When testing the MLP model, an improvement in data distribution was observed compared to LR and DT (Fig. 3C). The comparison between observed and predicted values indicates that the RF and XGBoost models produced predictions that closely align with the actual observations (Fig. 3D and E).

Table 3 presents the correlation coefficients (r), mean absolute error (MAE), and coefficient of determination ($R^2$) of the machine learning models calibrated using Linear Regression (LR), Decision Tree (DT), Multilayer Perceptron (MLP), Random Forest (RF), and XGBoost algorithms. The models were trained using 5-fold cross-validation with 100 repetitions, employing the following predictor variables: light interception (LI), nitrogen dose, climate data, pre-grazing height, post-grazing height, and grazing interval (GI) to predict crude protein (CP) in Tamani grass.

The DT model presented a lower Pearson's r value, when testing LR, the coefficient improved in relation to DT ($p < 0.05$). The MLP ($r = 0.75$), RF ($r = 0.78$) and XGBoost ($r = 0.78$) models presented the highest Pearson's r values ($p < 0.05$), not differing from each other ($p > 0.05$). For the MAE metric, a lower value was observed for XGBoost (MAE = 1.45) ($p < 0.05$). The RF, MLP and LR models presented intermediate MAE values and did not differ from each other ($p < 0.05$). A higher MAE was observed for the DT model (MAE = 2.22) ($p < 0.05$). For the $R^2$ metric, XGBoost stood out with the highest value ($R^2 = 56.65\%$), followed by RF ($R^2 = 49.07\%$) and MLP ($R^2 = 44.18\%$) ($p < 0.05$). Thus, answering RQ2, we can state that the best machine learning model evaluated to predict CP of tamani grass pastures was XGBoost with a performance of $r = 0.78$, MAE = 1.45 and $R^2 = 56.90$ (Table 3).

## Answering RQ 3 - What are the most relevant variables in the best-performing models?

The contribution of individual attributes to estimate the CP of Tamani grass (Fig. 4) was evaluated using the Random Forest (RF) algorithm, considering all input variables to compare their relative importance. In RF, variable importance is estimated based on the mean decrease in impurity, which reflects how much each

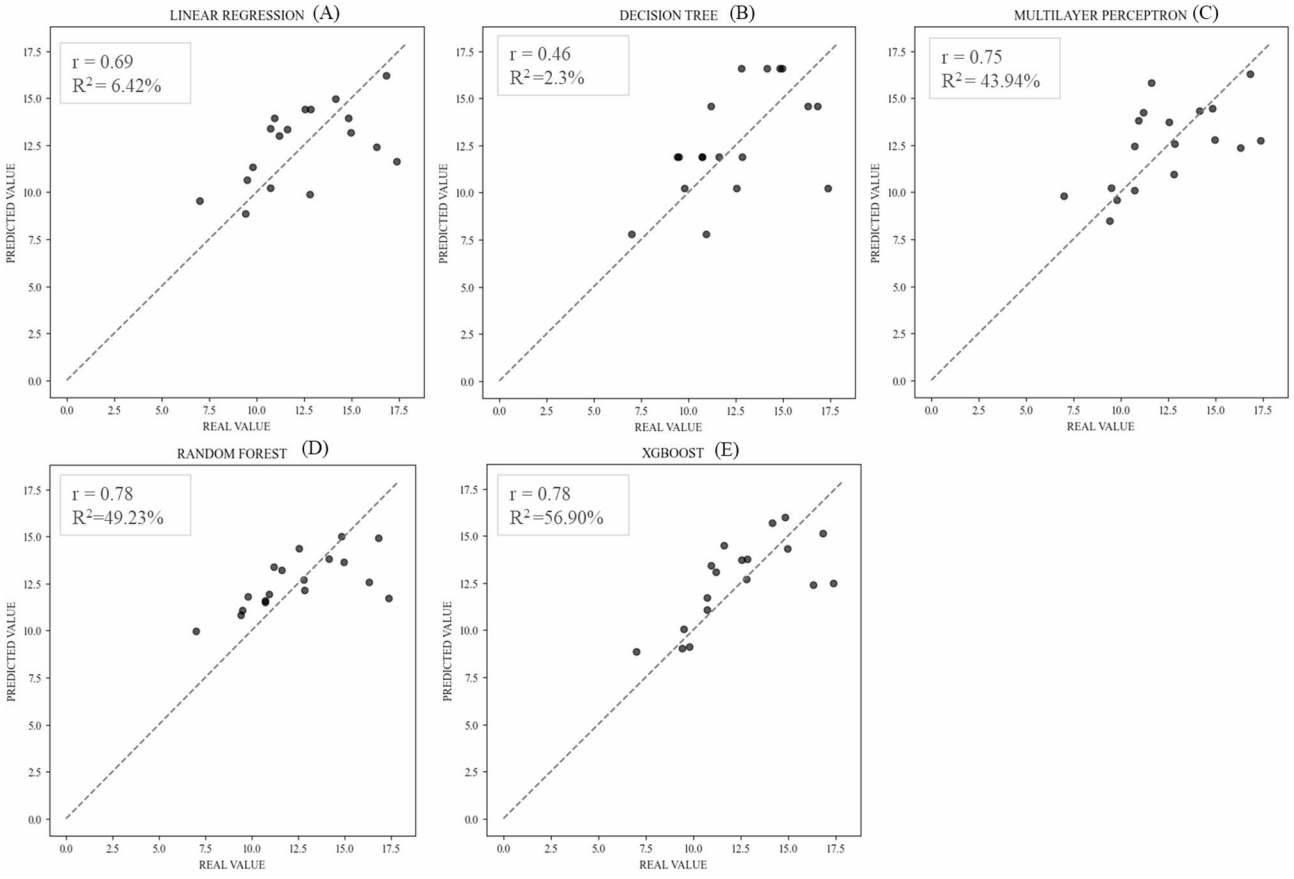


**Fig. 3**. Scatter plots between observed and predicted values of each model tested: Linear Regression (LR), Decision Tree (DT), Multilayer Perceptron (MLP), Random Forest (RF), and XGBoost for predicting crude protein in tamani grass leaves.

| Models | *r* | MAE | $R^2$ |
|---|---|---|---|
| LR | 0.69b | 1.67b | 6.42b |
| DT | 0.46c | 2.23a | 2.30b |
| MLP | 0.75a | 1.56bc | 43.94a |
| RF | 0.78a | 1.60b | 49.23a |
| XGBoost | 0.78a | 1.45c | 56.90a |
| p-value | < 0.001 | < 0.001 | < 0.001 |
| SEM | 0.014 | 0.034 | 5.53 |

**Table 3**. Pearson correlation coefficient (r), mean absolute error (MAE) and coefficient of determination ($R^2$) of each machine learning (ML) model: linear regression (LR), decision tree (DT), multilayer perceptron (MLP), random forest (RF), and XGBoost for prediction of crude protein from Tamani grass leaves. *SEM* standard error of the mean. Means followed by different lowercase letters in the column differ from each other by Tukey's test at 5% probability.

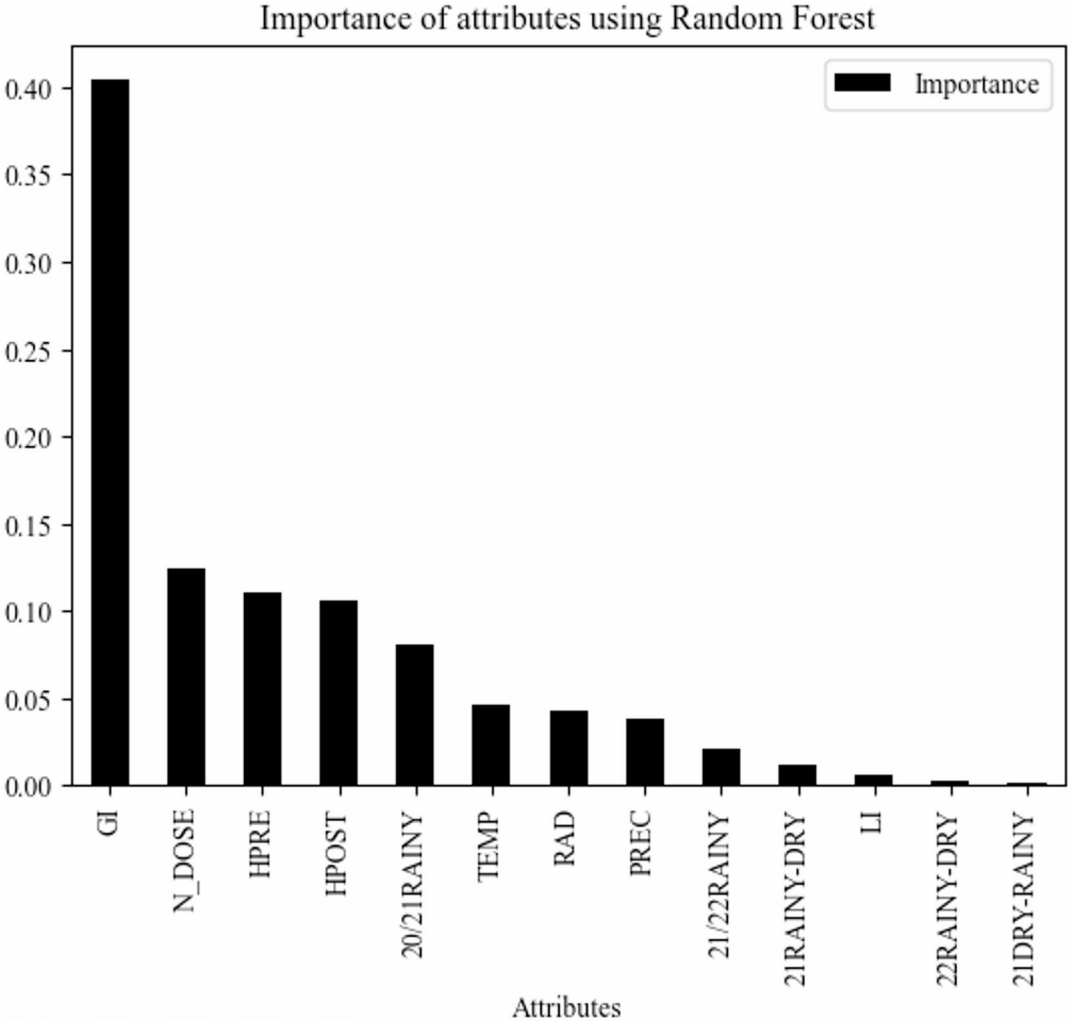


**Fig. 4**. Importance of attributes based on mean impurity decrease (and number of nodes using that attribute) in Random Forest. *GI* grazing interval, *HPRE* pre-grazing height, *HPOST* post-grazing height, *TEMP* temperature, *RAD* solar radiation, *PREC* precipitation, *LI* light interception.

predictor reduces the heterogeneity of the target variable when used to split a node during model training. This reduction is accumulated across all trees and normalized to obtain a relative importance score for each variable. The decrease in impurity is typically measured through the reduction in the residual sum of squares, indicating the extent to which each variable contributes to improving model accuracy[27]. Additionally, tree-based methods inherently handle multicollinearity better than linear models, as they partition the feature space

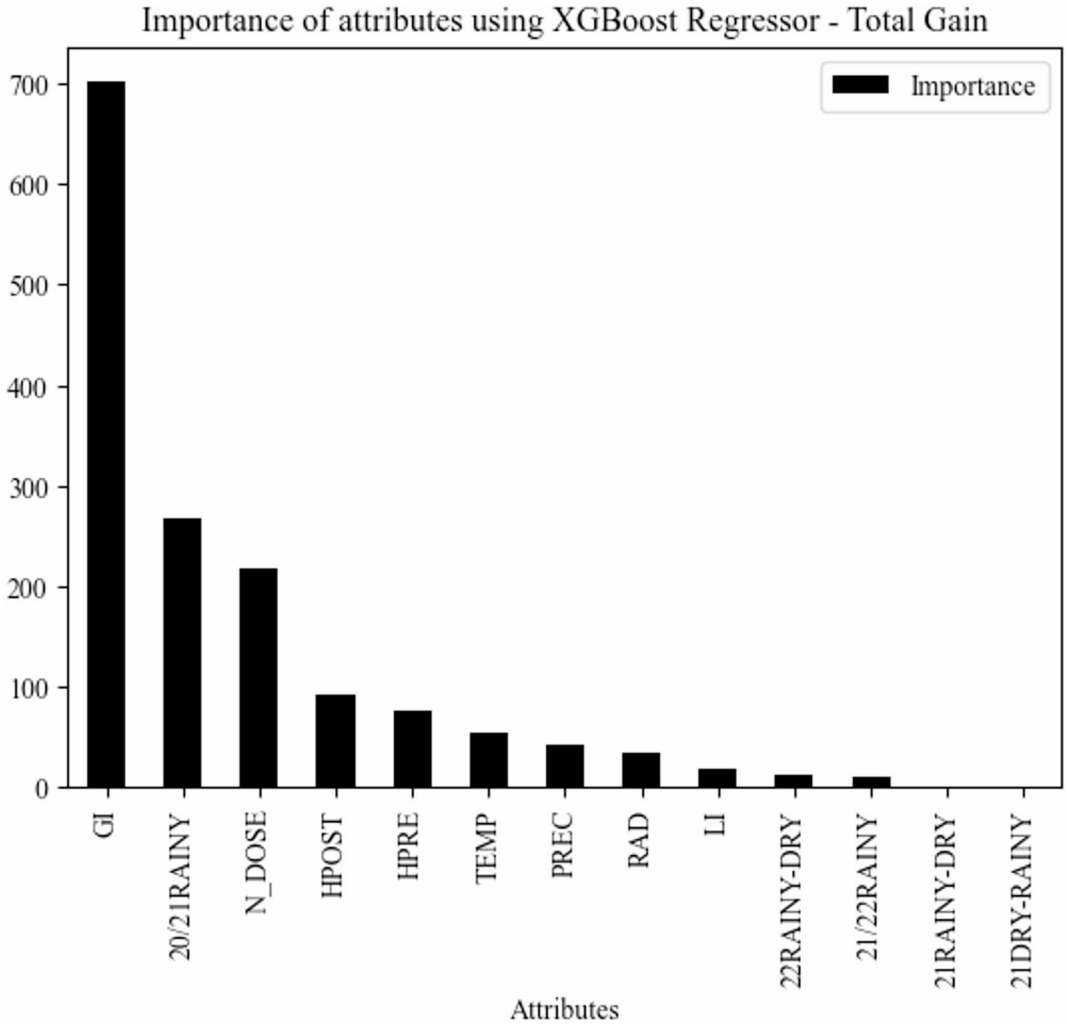


**Fig. 5**. Importance of attributes using XGBoost based on total gain. *GI* interval between grazing, *HPOST* post-grazing height, *HPRE* pre-grazing height, *TEMP* temperature, *RAD* solar radiation, *LI* light interception.

hierarchically and can distribute correlated predictors across different branches[27]. In response to RQ3, the most relevant variables for the RF model is GI, followed by nitrogen dose, pre-grazing height, post-grazing height, rainy season 20/21, precipitation, temperature, radiation, with greater importance in the prediction and finally, the rainy seasons 21/22, rainy-dry transition 21, rainy-dry transition 22, LI and dry-rainy transition 21 with less importance (Fig. 4).

The contribution of individual attributes to estimate the CP of tamani grass (Fig. 5) was based on XGBoost considering all input data to compare the importance of the attributes in XGBoost based on total gain.

The order of importance of the variables changes between the models tested (Figs. 4 and 5). In response to RQ3, the most important variables in the XGBoost model for predicting CP are GI, followed by 20/21 rainy, nitrogen dose, post-grazing height, pre-grazing height, and temperature (Fig. 5). Precipitation, radiation, LI, and the rainy-dry seasons 22 and rainy 21/22 are the attributes with the least importance in contributing to the prediction (Fig. 5).

## Discussion

The significant correlations between CP and N dose ($r = 0.48$), 20/21 Rainy season ($r = -0.43$), 21/22 Rainy season ($r = 0.40$), pre-grazing height (HPRE; $r = -0.49$), post-grazing height (HPOST; $r = -0.46$), grazing interval (GI; $r = -0.23$), and precipitation ($r = -0.33$) highlight the importance of these factors in determining the chemical composition of the forage and, consequently, its quality. It is well known that forage plants respond to abiotic factors such as light, water, temperature, and nutrients[31]. Although no significant correlations were observed between CP and temperature or radiation ($p > 0.05$), correlations were found with the rainy seasons (20/21 and 21/22), which encompass climatic components such as temperature, radiation, humidity, and atmospheric pressure.

Likewise, CP did not correlate with LI, but a correlation was observed with HPRE and HPOST. It is well elucidated in the literature that the height of the forage canopy is related to LI[32,33], being used to control the

structure of the forage canopy. In addition, a strong correlation was observed between LI and HPRE and between HPRE and HPOST. In production systems that use intermittent stocking as a grazing method, the grazing frequency will determine the rest period of the forage plant, which can influence productivity, chemical composition and regrowth capacity[34].

The negative correlation observed between CP and GI demonstrates its influence on the chemical composition of the pasture, as the higher the GI, the older the forage plant, which indicates greater lignification of the tissues and an increase in supporting structures and fibrous components[35]. The concentration of carbon and nitrogen compounds in the plant varies according to the maturity of the plant tissue[36], with changes in the chemical composition of the cell wall.

The positive correlation observed between CP and nitrogen (N) levels can be explained by its fundamental role in plant physiology. This macronutrient constitutes cellular components such as chlorophyll, amino acids, and nucleic acids[37]. The correct use of nitrogen fertilization can result in greater forage production and increases in CP levels[38]. In this sense, Paciullo et al.[39] reported increases in CP in Massai and Tanzania cultivars in response to increasing N rates.

The potential for predicting crude protein in the leaves of Tamani grass pastures was evaluated using machine learning models. A representative database collected over 18 months from Tamani grass pastures with two nitrogen rates (80 and 240 kg N $ha^{-1}$ $year^{-1}$) and two grazing frequencies (90 and 95% LI) was used. The predictor variables were selected for their ease of acquisition and low cost in routine pasture management. CP content was estimated using near-infrared spectroscopy (NIRS), an indirect method that avoids destructive sampling and reduces laboratory costs. While NIRS provides reliable predictions when properly calibrated, its limitations should be considered when interpreting the results. In addition, the visual selection of sampling sites, although conducted following standardized criteria, is inherently susceptible to operator bias, which may influence the representativeness of the harvested forage mass used for NIRS analysis. In this study, sampling sites were chosen based on mean canopy height within each paddock, avoiding water troughs, dung patches, animal trails, and visually heterogeneous areas to minimize such bias and ensure that samples reflected the average canopy structure. Nevertheless, the application of these indirect methods combined with machine learning models offers a practical and cost-effective approach to obtain valuable information on forage quality for pasture-based animal production systems.

Machine learning has the potential to predict leaf CP content based on observed performance, $r = 0.78$ and $R^2 = 49.07\%$ for the RF model and $r = 0.78$ and $R^2 = 56.65\%$ for the XGBoost model. These results were obtained from low-cost and easily accessible information, which enhances their applicability. None of the variables used require the use of drones, satellite images, or laboratory analyses, and can be obtained through records that are usually already part of most experimental field tests. These results demonstrate the importance of the chosen variables and highlight the superior performance of the XGBoost model in relation to the RF. Ramos et al.[18] found that the RF model performs well ($r = 0.72$ to $0.78$) to predict variables for agricultural purposes. Bretas et al.[40] found predictive performance for aboveground fresh biomass and dry matter concentration of ($R^2 = 61\%$ and 69%, respectively) using RF. Reis et al.[20] observed superior performance for XGBoost ($R^2 = 65\%$ and 89%) for aboveground fresh biomass and height, respectively, when compared to RF ($R^2 = 60\%$ and 88%) for aboveground fresh biomass and height, respectively. It is worth mentioning that these authors used images that require the use of specific equipment such as drones, multispectral cameras, or satellites, in addition to favorable environmental conditions for consistent captures[41]. In addition, processing the images demands greater computational capacity and consequently higher costs.

Another important limitation refers to the complexity of preprocessing and data annotation. The use of images often requires steps such as geometric correction, segmentation, and manual labeling, which can be time-consuming and prone to human error[42]. Furthermore, computer vision-based models tend to be less interpretable compared to models based on tabular data, which can make them difficult to adopt by users who require clear explanations for decision-making[43].

The GI stands out as the most influential variable in CP content, however, it is a variable resulting from management and adequate environmental conditions that, when added to nitrogen fertilization, allow shortening the rest period of tropical pastures[44]. This results in younger pastures with higher CP content, while with increasing maturity there is a dilution effect of the protein and an increase in fibrous components[8]. Nitrogen is among the variables that most influences CP content because, when used properly, it enhances the photosynthesis processes for carbon fixation in biomass, causing an increase in cellular constituents, in addition to acting in zones of leaf division and elongation and in meristematic zones of plant tissue growth[37,45,46]. Thus, nitrogen has a direct effect on tissue flow, allowing a reduction in the interval between grazing[47], which influences the harvest time and forage quality.

In addition to the GI factors and N doses being the most important attributes in leaf CP prediction models, pre- and post-grazing heights and 20/21 Rainy also stand out. The definition of grazing goals influences the quality of forage available for animal grazing[48]. When properly established, they allow a significant increase in pasture productivity, since dry matter accumulation is strongly correlated with grazing frequency, which interacts with the intensity of defoliation. This directly affects forage productivity, influencing its components and, consequently, its chemical composition throughout the year[49].

Precipitation, solar radiation, temperature, and seasons have less influence on CP content when compared to management variables. It is worth noting that the seasons of the year are made up of a set of environmental characteristics, namely solar radiation, temperature and water availability. Sunlight increases the ratio of cell content to cell wall, as it increases the formation of non-structural carbohydrates and amino acids, which can result in improved quality of forage plants[50]. Temperature acts directly on biochemical, physical and morphogenic processes, influencing forage quality[51,52]. High temperatures result in greater lignification of the cell wall, which reduces the relationship between cell content and cell wall[44]. Water availability plays a fundamental role in plant

metabolism as it is responsible for the transport of nutrients, absorption pathway and means of dissipating solar energy and excess temperature[50], acting indirectly on the nutritional quality of forage.

This explains why 20/21 Rainy, unlike the other seasons, had a great influence on CP content, as it was the period with the highest rainfall that provided the greatest number of grazing cycles during the experimental period. The pasture ecosystem is dynamic and sensitive to changes in management and climate change, Chen et al.[53] highlighted the importance of meteorological data such as temperature and precipitation as influencing factors in pasture development. The maturity stage of the forage plant is considered a primary factor in changing the components of nutritional value, while environmental variations such as soil moisture and fertility are secondary factors[54]. The maturity stage of the plant can be controlled by management strategies, exerting a greater influence on the chemical composition of the plant than abiotic factors. The results confirm this, since the variables GI, N dose, pre- and post-grazing height are among the attributes that contributed most to the model's prediction. Appropriate management strategies must be adopted when the objective is to produce pastures with good chemical composition regardless of abiotic factors.

The main limitations of applying machine learning (ML) to pasture management include: the scarcity of reliable and representative data for model calibration, due to labor-intensive data collection, limited adoption of new technologies by ranchers, and concerns regarding data privacy and security[56]; the high spatial and temporal variability of grazing lands, which necessitates diverse, multi-year training datasets to enhance model robustness[4]; the substantial morphological variability of forage species, which may require model adjustments or development of species-specific models; and the challenges associated with managing and analyzing large, complex datasets from multiple sensors, further exacerbated by a global shortage of data science professionals[55]. Additionally, the relatively small dataset used in this study should be acknowledged as a limitation. Similar pasture-based studies that rely on field-collected agronomic data often face comparable constraints, as generating sufficiently large datasets is time-consuming, costly, and logistically challenging. Consequently, model performance and generalizability may be affected, particularly when compared to studies in other agricultural domains that employ larger datasets or remote-sensing-based high-throughput phenotyping. These limitations should be carefully considered when developing and applying ML models in pasture systems.

Future work should focus on expanding and diversifying training datasets, developing species-specific or adaptive models, and enhancing data science capacity among pasture managers to improve model accuracy and applicability.

## Conclusion

Information such as climate data, nitrogen fertilization, grazing interval, pre- and post-grazing heights, and light interception are important predictors for estimating leaf crude protein in Tamani grass pastures.

Among these, grazing interval, nitrogen dose, and pre- and post-grazing heights showed the greatest predictive importance, suggesting that management practices may have a stronger influence on the chemical composition of Tamani grass than environmental conditions. However, further research with larger and more diverse datasets is needed to confirm these findings.

The Random Forest and XGBoost machine learning models demonstrated satisfactory predictive performance, with XGBoost performing slightly better. Overall, this study provides a promising approach for estimating leaf crude protein, an important parameter for adjusting nutritional requirements in pasture-based animal production systems.

## Data availability

All data generated or analysed during this study are included in this published article.

Received: 13 May 2025; Accepted: 19 January 2026
Published online: 20 January 2026

## References

1. FAO. Land statistics 2001–2022. Global, regional and country trends. Food and Agriculture Organization of the United Nations. July 4, 2024. Avaiable: https://www.fao.org/statistics/highlights-archive/highlights-detail/land-statistics-2001-2022.-global--regional-and-country-trends/en
2. Reinermann, S. et al. Remote sensing of grassland production and management—A review. Remote Sens. 12, 1949 (2020). https://doi.org/10.3390/rs12121949
3. Wijesingha, J. et al. Predicting forage quality of grasslands using UAV-borne imaging spectroscopy. Remote Sens. 12, 126 (2020).https://doi.org/10.3390/rs12010126
4. Bretas, I. L. et al. Prediction of aboveground biomass and dry-matter content in brachiaria pastures by combining meteorological data and satellite imagery. Grass Forage Sci. 76, 340–352 (2021). https://doi.org/10.1111/gfs.12517
5. Balseca, D. G. et al. Nutritional value of brachiarias and forage legumes in the humid tropics of Ecuador. Cienc. Investig. Agrar. 42, 57–63 (2015). https://doi.org/10.4067/S0718-16202015000100006
6. Lopes, C. M. et al. Massa de forragem, composição morfológica e valor nutritivo de capim-braquiária Submetido a níveis de Sombreamento e fertilização. Arq. Bras. Med. Vet. Zootec. 69, 225–233 (2017). https://doi.org/10.1590/1678-4162-9201
7. Rouquette, F. M. Invited review: the roles of forage management, forage quality, and forage allowance in grazing. Prof. Anim. Sci. 32, 10–18 (2016). https://doi.org/10.15232/pas.2015-01408
8. Zubieta, A. S. et al. Does grazing management provide opportunities to mitigate methane emissions by ruminants in pastoral ecosystems? Sci. Total Environ. 754, 142029 (2021). https://doi.org/10.1016/j.scitotenv.2020.142029
9. Horrocks, R. D. & Vallentine, J. F. Forage quality—the basics. in Harvested Forage 17–47 (Academic Press, 1999).
10. t'Mannetje, L. & Jones, R. Field and Laboratory Methods for Grassland and Animal Production Research. (CABI Publishing, 2000).
11. Gornott, C. & Wechsung, F. Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany. Agric. For. Meteorol. 217, 89–100 (2016). https://doi.org/10.1016/j.agrformet.2015.10.005

12. Jeong, J. H. et al. Random Forests for global and regional crop yield predictions. PLoS One 11, e0156571 (2016). https://doi.org/10.1371/journal.pone.0156571
13. Le Maire, G. et al. MODIS NDVI time-series allow the monitoring of Eucalyptus plantation biomass. Remote Sens. Environ. 115, 2613–2625 (2011). https://doi.org/10.1016/j.rse.2011.05.017
14. Freitas, E. C. S. et al. Modeling of Eucalyptus productivity with artificial neural networks. Ind. Crops Prod. 146, 112149 (2020). https://doi.org/10.1016/j.indcrop.2020.112149
15. Vahedi, A. A. Monitoring soil carbon pool in the hyrcanian coastal plain forest of iran: artificial neural network application in comparison with developing traditional models. Catena 152, 182–189 (2017). https://doi.org/10.1016/j.catena.2017.01.022
16. Vani, S. et al. Prediction of sugar yields during hydrolysis of lignocellulosic biomass using artificial neural network modeling. Bioresour. Technol. 188, 128–135 (2015). https://doi.org/10.1016/j.biortech.2015.01.083
17. Da Silva, A. K. V. et al. Predicting Eucalyptus diameter at breast height and total height with UAV-based spectral indices and machine learning. Forests 12, 582 (2021). https://doi.org/10.3390/f12050582
18. Ramos, A. P. M. et al. A random forest ranking approach to predict yield in maize with UAV-based vegetation spectral indices. Comput. Electron. Agric. 178, 105791 (2020). https://doi.org/10.1016/j.compag.2020.105791
19. Mwendia, S. W. et al. Does harvesting Urochloa and Megathyrsus forages at short intervals confer an advantage on cumulative dry matter yields and quality? J. Sci. Food Agric. 102, 750–756 (2022). https://doi.org/10.1002/jsfa.11407
20. Reis, A. A. et al. Monitoring pasture aboveground biomass and canopy height in an integrated crop–livestock system using textural information from planetscope imagery. Remote Sens. 12, 2534 (2020). https://doi.org/10.3390/rs12162534
21. Rosa, D. et al. Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery. Comput. Electron. Agric. 180, 105880 (2021). https://doi.org/10.1016/j.compag.2020.105880
22. Solos, E. Sistema Brasileiro De classificação De Solos. 3ª ed. (Centro Nacional de Pesquisa de Solos, Rio de Janeiro, 2018).
23. De Campos, R. L. C., De Oliveira, R. H. M., Macedo, M., & Gomes, R. D. C. Predicting the nutritional value of tropical forages by Near Infrared Reflectance Spectroscopy (NIRS) (2024).
24. Pedregosa, F. et al. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830 (2011).
25. Štepanovský, M. et al. Novel age Estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods. Forensic Sci. Int. 279, 72–82 (2017). https://doi.org/10.1016/j.forsciint.2017.08.005
26. Raj, P. & Evangeline, P. The Digital Twin Paradigm for Smarter Systems and Environments: the Industry Use Cases. (Academic Press, 2020).
27. Breiman, L. Random forests. Mach. Learn. 45, 5–32 (2001).
28. Saha, S. et al. Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, reptree: A case study at the Gumani river Basin, India. Sci. Total Environ. 730, 139197 (2020). https://doi.org/10.1016/j.scitotenv.2020.139197
29. Bisong, E. et al. Building machine learning and deep learning models on Google Cloud Platform. (Apress, 2019).
30. Moore, David S. & Mccabe, George. Introduction to the practice of statistics. New York, Freeman (2004).
31. Nabinger, C. Eficiência do Uso de pastagens: disponibilidade e Perdas de forragem. in Simpósio Sobre Manejo Da Pastagem 213–251 (1997).
32. Carnevalli, R. A. et al. Herbage production and grazing losses in panicum maximum cv. Mombaça under four grazing managements. Trop. Grassl. 40, 165–171 (2006).
33. Pedreira, B. C., Pedreira, C. G. S. & Silva, S. C. Acúmulo de forragem Durante a rebrotação de capim-xaraés Submetido a três estratégias de desfolhação. Rev. Bras. Zootec. 38, 618–625 (2009).
34. Costa, N. L. et al. Resposta de Pastagens de Megathyrsus Maximus cv. Zuri à frequência de desfolhação. Res. Soc. Dev. 8, e1185 (2019).
35. Geremia, E. V. et al. Sward structure and herbage intake of brachiaria Brizantha cv. Piatã in a crop-livestock-forestry integration area. Livest. Sci. 212, 83–92 (2018). https://doi.org/10.1016/j.livsci.2018.03.020
36. Hodgson, J. Grazing Management: Science into Practice. (Longman Group UK Ltd., 1990).
37. Taiz, L. et al. Fisiologia E Desenvolvimento Vegetal: Diversidade Vegetal. (Artmed, 2017).
38. Delevatti, L. M. et al. Effect of nitrogen application rate on yield, forage quality, and animal performance in a tropical pasture. Sci. Rep. 9, 7596 (2019). https://doi.org/10.1038/s41598-019-44138-x
39. Paciullo, D. S. C. et al. Morphogenesis, biomass and nutritive value of panicum maximum under different shade levels and fertilizer nitrogen rates. Grass Forage Sci. 72, 590–600 (2017). https://doi.org/10.1111/gfs.12264
40. Bretas, I. L. et al. Canopy height and biomass prediction in Mombaça Guinea grass pastures using satellite imagery and machine learning. Precis. Agric. 1–25 (2023). https://doi.org/10.1007/s11119-023-10013-z
41. Kamilaris, A. & Prenafeta-Boldú, F. X. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90 (2018). https://doi.org/10.1016/j.compag.2018.02.016
42. Lu, D. & Weng, Q. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. 28, 823–870 (2007). https://doi.org/10.1080/01431160600746456
43. Doshi-Velez, F. & Kim, B. Towards a rigorous science of interpretable machine learning. ArXiv Preprint arXiv:1702.08608 (2017).
44. Cunha, A. M. Q. et al. Nitrogen fertilisation as a strategy for intensifying production and improving the quality of Massai grass grown in a humid tropical climate. J. Plant Nutr. 45, 2213–2227 (2022). https://doi.org/10.1080/01904167.2022.2046078
45. Galindo, F. S. et al. Technical and economic viability and nutritional quality of Mombasa Guinea grass silage production. Acta Sci. Agron. 40, e36395 (2018). https://doi.org/10.4025/actasciagron.v40i1.36395
46. Li, P. et al. Elevated carbon dioxide and nitrogen supply affect photosynthesis and nitrogen partitioning of two wheat varieties. J. Plant Nutr. 42, 1290–1300 (2019). https://doi.org/10.1080/01904167.2019.1616758
47. Zanine, A. et al. Effect of season and nitrogen fertilization on the agronomic traits and efficiency of Piatã grass in Brazilian savanna. Agriculture 10, 337 (2020). https://doi.org/10.3390/agriculture10080337
48. Alvarenga, C. A. F. et al. Animal performance and Sward characteristics of Mombaça guineagrass pastures subjected to two grazing frequencies. Trop. Grassl. Forrajes Trop. 8, 1–10 (2020). https://doi.org/10.17138/tgft(8)1-10
49. Da Silva, S. C., Sbrissia, A. F. & Pereira, L. E. T. Ecophysiology of C4 forage grasses—understanding plant growth for optimizing their use and management. Agriculture 5, 598–625 (2015).
50. Van Soest, P. J. Nutritional Ecology of the Ruminant. (Cornell University Press, 1994).
51. Gillet, M. Las gramíneas Forrajeras: Descripción, funcionamiento, Aplicaciones Al Cultivo De La Hierba. (Acribia, 1984).
52. Reis, R. A., Bernardes, T. F. & Siqueira, G. R. Forragicultura: Ciência, Tecnologia E gestão Dos Recursos Forrageiros. (Gráfica Multipress, 2013).
53. Chen, Y. et al. Estimating pasture biomass using Sentinel-2 imagery and machine learning. Remote Sensing 13, 603 (2021). https://doi.org/10.3390/rs13040603
54. Tesk, C. R. M. et al. Impact of grazing management on forage qualitative characteristics: A review. Sci. Electron. Arch. 11, 188–197 (2018).
55. Bretas, I. L., Dubeux Jr, J. C., Cruz, P. J., Oduor, K. T., Queiroz, L. D., Valente, D. S., & Chizzotti, F. H. Precision livestock farming applied to grazingland monitoring and management—a review. *Agron. J.* 116, 1164–1186 (2024). https://doi.org/10.1002/agj2.21346
56. Murphy, D. J., Murphy, M. D., O'Brien, B., & O'Donovan, M. A. Review of precision technologies for optimising pasture measurement on Irish grassland. *Agriculture* 11, 600 (2021). https://doi.org/10.3390/agriculture11120600

## Acknowledgements
AcknowledgementsThe authors thank the Embrapa Beef Cattle, Federal University of Mato Grosso do Sul Foundation, through the Postgraduate Program in Animal Science, the National Council for Scientific and Technological Development (CNPq), the Higher Education Personnel Improvement Coordination (CAPES, Finance Code 001) and the Foundation for the Support of the Development of Education, Science and Technology of the State of Mato Grosso do Sul (FUNDECT).

## Author contributions
G.O.A.M., G.S.D., E.T.M., D.B.M. and V.P.B.E., designed the study. G.O.A.M., J.G.R., M.G.P., J.C.S.S., R.T.N. and J.A.C. performed the experiment and collected data. G.O.A.M., R.T.N., J.A.C., A.B.C., L.C.V.I. and M.M.P.M.F.C. analyzed the data. G.O.A.M. conducted statistical analysis and wrote the manuscript. All authors read and critically revised drafts for intellectual content and provided approval for publication.

## Funding
The authors received no funding for this work.

## Declarations

## Competing interests
The authors declare no competing interests.
The authors declare that they have no conflict of interest.

## Additional information
**Correspondence** and requests for materials should be addressed to G.O.d.A.M.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.