Innovation in forage evaluation: digestibility and metabolizable energy modeling with canonical correlation analysis and random forest in Colombia

Innovación en la evaluación de forrajes: modelado de digestibilidad y energía metabolizable con análisis de correlación canónico y random forest en Colombia.

Danny Samuel Martínez Lobo[1], **Elkin Anderson Bravo Rusinque**[1]

Highlights

- Fiber fractions (NDF, ADF, lignin, hemicellulose) showed negative associations with digestibility, while crude protein and starch contributed positively.
- The Random Forest model achieved R2 values above 0.95 with low RMSE and MAE, demonstrating high predictive accuracy for digestibility and metabolizable energy.
- The integration of Canonical Correlation Analysis and Random Forest provides a robust and replicable tool for predicting the nutritional quality of tropical forages

Innovaciencia ISSN: 2346-075X
E- ISSN: 2346-075X
Innovaciencia 2026; 14 (1):e5664
http//dx.doi.org/10.15649/2346075X.5664

ORIGINAL RESEARCH

How to cite this article:

Martínez Lobo DS, Bravo Rusinque EA. Innovation in forage evaluation: digestibility and metabolizable energy modeling with canonical correlation analysis and random forest in Colombia. *Innovaciencia.* 2026;14(1):e5664.
http//dx.doi.org/10.15649/2346075X.5664

Received: 10 September 2025
Accepted: 13 March 2026
Published: 15 May 2026

Keywords:

Tropical Forages, digestibility, metabolizable energy, predictive Model

Palabras clave:

Forrajes tropicales, digestibilidad, energía metabolizable, modelo predictivo.

ABSTRACT

Introduction. Livestock productivity in Colombia largely depends on forage-based feeding systems, where nutritional quality directly influences digestibility and metabolizable energy in ruminants (ME-r). **Objectives.** To analyze the AlimenTro database in order to evaluate the relationship between nutrients, digestibility, and ME of tropical forages using Canonical Correlation Analysis (CCA) and the Random Forest algorithm. **Materials and Methods**. A total of 500 samples of grasses, legumes, and shrubs from the AlimenTro database were evaluated using standardized methodologies, including near-infrared spectroscopy (NIRS), proximate analysis, and in situ techniques. CCA was applied to identify multivariate associations between nutrients and energy parameters, while the Random Forest algorithm was used to assess predictive performance and variable importance, yielding low root mean square error (RMSE) values. **Results**. CCA revealed that fiber fractions, including neutral detergent fiber (NDF), acid detergent fiber (ADF), lignin, and hemicellulose, were negatively associated with digestibility and ME, whereas crude protein and starch showed positive associations. The Random Forest model achieved R^2 values greater than 0.95, with low RMSE and mean absolute error (MAE), confirming its high predictive accuracy and highlighting fiber and protein as key predictors. **Conclusions.** The integration of CCA and Random Forest provides a robust and applicable approach for predicting the nutritional quality of tropical forages. This methodological framework supports the development of more efficient and sustainable feeding strategies in tropical livestock systems.

RESUMEN

Introducción. La productividad ganadera en Colombia depende en gran medida de sistemas de alimentación basados en forrajes, cuya calidad nutricional influye directamente en la digestibilidad y la energía metabolizable en rumiantes (EM-r). **Objetivos.** Analizar la base de datos AlimenTro para evaluar la relación entre nutrientes, digestibilidad y energía metabolizable de forrajes tropicales mediante el Análisis de Correlación Canónica (ACC) y el algoritmo Random Forest. **Materiales y Métodos.** Se analizaron 500 muestras de gramíneas, leguminosas y arbustivas de la base de datos AlimenTro, caracterizadas mediante metodologías estandarizadas, incluyendo espectroscopía de infrarrojo cercano (NIRS), análisis proximal y técnicas in situ. El ACC se aplicó para identificar asociaciones multivariadas entre nutrientes y parámetros energéticos, mientras que el algoritmo Random Forest se utilizó para evaluar el desempeño predictivo y la importancia de las variables, obteniendo valores bajos del error cuadrático medio (RMSE). **Resultados.** El ACC mostró que las fracciones de fibra, incluyendo fibra detergente neutra (FDN), fibra detergente ácida (FDA), lignina y hemicelulosa, se asocian negativamente con la digestibilidad y la energía metabolizable, mientras que la proteína cruda y el almidón presentan asociaciones positivas. El modelo Random Forest alcanzó valores de R^2 superiores a 0.95, con bajos valores de RMSE y error absoluto medio (MAE), confirmando su alta precisión predictiva y destacando la fibra y la proteína como variables clave. **Conclusiones.** La integración del ACC y Random Forest proporciona un enfoque robusto y aplicable para predecir la calidad nutricional de forrajes tropicales. Este marco metodológico contribuye al diseño de estrategias de alimentación más eficientes y sostenibles en sistemas ganaderos tropicales.

1 Maestría en Estadística Aplicada y Ciencia de Datos, Facultad de Ciencias, Universidad El Bosque, Bogotá, Colombia.
*Corresponding Author: dsmartinez@unbosque.edu.co.

INTRODUCTION

In Colombia, livestock farming plays a fundamental role in the economy. In 2023, it contributed 25.25% to the agricultural Gross Domestic Product (GDP) and 1.46% to the national GDP[1]. Additionally, it has a significant impact on food security and the development of the country's rural areas. This activity relies heavily on the use of forages as the main-and in many cases, the only-source of feed for ruminants, due to their availability and adaptability to tropical grazing systems[2]. Beyond their basic function as feed, forages are a vital source of nutrients, providing energy, nitrogen, vitamins, minerals and structural fiber, all of which are essential elements for the nutrition, well-being and performance of the animals[3].

As forages mature, their composition changes, mainly through an increase in fibrous fractions such as neutral detergent fiber (NDF), which reduces the digestibility and energy value of the feed[4]. These changes directly impact feed efficiency, as they affect both the proportion of nutrients the animal can absorb[5]. and the amount of energy available for vital functions. Understanding how these variables interact is key to improving diet formulation and overall system productivity.

Although these parameters have traditionally been assessed using laboratory techniques or regression models, the advancement of analytical technologies such as machine learning offers new opportunities to develop accurate models tailored to local conditions[6, 7]. Furthermore, multivariate analysis allows for the identification of relationships between nutritional components and their combined effect on digestibility and energy[8].

This study proposes an integrated approach that combines Canonical Correlation Analysis (CCA) with the Random Forest algorithm to predict the digestibility and ME-r of tropical forages. This approach seeks to optimize diet formulation by generating practical tools that enhance the use of forage resources, promoting efficient and sustainable livestock systems[9, 10].

MATERIALS AND METHODS

This was an observational, cross-sectional study based on the secondary analysis of data. A total of 500 tropical forage samples (grasses, legumes, and shrubs) were used, obtained from the national AlimenTro database[11]. All samples had complete records, and therefore no missing data were detected; consequently, no data imputation procedures were required. The samples were characterized using standardized methodologies (NIRS, in situ proximate analysis) developed by AGROSAVIA. For the purpose of this analysis, they were grouped according to their botanical origin.

Data processing was performed in RStudio version 4.3.2 (R Foundation for Statistical Computing, Vienna, Austria). A CCA was applied to explore associations between nutritional variables and parameters of digestibility and metabolizable energy[12]. Using the variables selected by the CCA, a predictive Random Forest model was built, configured with 500 trees and K-fold cross-validation (k=10). The model's performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE).

Overall, the model proved to be a robust, adaptable tool to support the nutritional evaluation of forages under tropical conditions, with direct applications in material selection and in the formulation of more efficient and sustainable diets.

The statistical assumptions were evaluated as follows: univariate normality of all nutritional and response variables was examined using the Kolmogorov–Smirnov test. The p-values indicated statistical deviations from normality; however, visual inspection of the Q–Q plots showed that these deviations were minor and not sufficiently severe to affect the multivariate analysis. Multivariate normality was assessed using Mardia's test, which showed no evidence of extreme skewness or kurtosis. Linearity between predictor and response variables was verified using scatterplots. Finally, multicollinearity within each set of variables was evaluated through inspection of the correlation matrices, revealing low correlations ($r < 0.4$), which suggests the absence of relevant multicollinearity.

Canonical Correlation Analysis

This methodology was used in this study because it is designed to examine and quantify the relationship between two sets of variables, one considered dependent on and the other independent[(12)]. This method identified canonical functions, which are linear combinations of variables in each set, optimized to maximize the correlation between the combinations[(13)]. Canonical functions are defined as:

$$U = a_1X_1 + \cdots + a_pX_p$$

$$V = b_1Y_1 + \ldots + b_qY_q$$

Where U and V represented the resulting canonical variables (canonical scores), constructed as linear combinations of the original X_i and Y_j, with coefficients a_i and b_j selected to maximize the correlation between U and V. The correlation between these two functions is known as the Canonical Correlation[(14)].

One of the main strengths of CCA is the ability to handle multiple variables simultaneously, making it a versatile technique for analyzing complex relationships[(15)]. In addition, CCA has been described as a generalization of other multivariate methods such as multiple regression, discriminant analysis and factor analysis, since all of them can be derived as particular cases of the canonical model[(16)].

Random Forest

This methodology, developed by Breiman[(17)], was used in this study due to its effectiveness in identifying nonlinear relationships with multivariate data. This method involves constructing multiple decision trees, each trained on random subsets of the dataset and predictor variables, which helps reduce overfitting and increase the model's generalization capacity. In agricultural and zootechnical applications, Random Forest has proven effective in predicting complex variables, such as the nutritional value of forages, by capturing nonlinear relationships between physicochemical properties and parameters like digestibility or energy value[(18)]. Its robustness to highly correlated variables and flexibility in handling heterogeneous data, such as those obtained from bromatological evaluations, make it a valuable tool, particularly because it requires minimal assumptions regarding data distribution[19]. The model can be conceptually expressed as:

$$\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N} T_i(x)$$

Let $\hat{f}(x)$ xrepresent the final prediction, where $T_i(x)$ denotes the prediction made by tree i and N is the total number of trees constructed. In regression tasks, the final prediction is obtained by averaging the individual of all trees, whereas in classification tasks, majority voting is used. During the training process, Random Forest randomly selects subsets of observations (bootstrapping) and subsets of predictors at each splitting node (feature bagging). This approach introduces diversity among the trees and enhances model stability[18].

An important advantage of Random Forest was its ability to evaluate the relative importance of predictor variables, allowed the identification of the nutritional parameters with the greatest influence on the digestibility or energy value of forages. This characteristic provided not only predictive capacity, but also interpretability, facilitating the design and optimization of diets[20].

Data Availability

The datasets used are from the public database Alimentro: *Alimentos del trópico para alimentación animal* (Tropical Foods for Animal Feed), available at **https://www.datos.gov.co/Agricultura-y-Desarrollo-Rural/Alimentos-del-tr-pico-para-alimentaci-n-animal-Ali/6arb-d547/data_preview** This database, developed by AGROSAVIA, compiles nutritional information on tropical forages using standardized methodologies. The information is publicly accessible and does not contain any sensitive or identifiable information.

RESULTS

The combination of multivariate statistical methods and machine learning algorithms enable an in-depth exploration of the relationship between the forage nutritional composition and its influence on digestibility and ME. The main findings obtained through Canonical Correlation Analysis, and the Random Forest predictive model are presented below, emphasizing the importance of fiber and protein components in the feed efficiency of Colombian livestock systems.

Significance Test for Canonical Equations

A CCA was applied to explore the relationships between forage nutritional variables (X) and parameters of digestibility and metabolizable energy (Y). This multivariate technique identified linear combinations that maximized the correlation between both sets of variables[13,21].

(**Table 1**) summarizes the canonical correlation coefficients for the two extracted functions, as well as their respective p-values. Two canonical functions were identified, with CV_1=0.943 and CV_2=0.134.

Table 1. Results of the CCA between nutritional variables (X) and digestibility and energy parameters (Y).

Canonical Equation	Canonical Correlation	Canonical R^2	P-value
CV1	0.943	0.889	$2.2\,e^{-16}$
CV2	0.134	0.018	$2.2\,e^{-16}$

Canonical correlations represent the strength of the association between linear combinations of forage nutritional variables (X) digestibility and ME-r parameters (Y).

Canonical function 1 (CV1) showed a high correlation (r=0.943), indicating a strong relationship between the input and output variables. This function explained most of the shared variance between sets *X* and *Y* and was found to be statistically significant ($p<2.2e^{-16}$). In contrast, the second canonical function 2 (CV2) exhibited a low correlation (r=0.134) suggesting limited practical relevance despite its statistical significance[22,-24]. These results indicate that CV1 captured the most meaningful relationship between the nutritional composition and the energetic performance of forages, providing a foundation for variable selection in subsequent predictive models.

Results of the Canonical Correlation Analysis

CCA made it possible to identify multivariate patterns between the nutritional components of forage (X) and the parameters of digestibility and ME-r (Y). Two statistically significant canonical functions were identified, of which the first (CV1=0.943) explained nearly all the covariance between the two sets of variables. This finding is consistent with studies that employ multivariate analysis to explore the relationship between nutritional composition and digestive performance in ruminants[25,26]. Furthermore, Leishman et al.[27] noted that this type of analysis offers an advantage over univariate methods by capturing multiple correlations simultaneously.

Canonical coefficients

The canonical coefficients enabled the interpretation of how each individual variable contributed to the construction of the linear combinations in the canonical functions (**Table 2**). In CV1, it was observed that the fibrous components NDF (Neutral Detergent Fiber), ADF (Acid Detergent Fiber), lignin and hemicellulose showed negative associations, while crude protein and total starch presented positive loadings.

Table 2. Canonical Coefficients for the X variables in CV1

Variable	CV1
Crude Protein	0.077
Total Starch	0.067
Ash	0.005
Cut-off age	0.005
NDF	-0.224
Ether Extract	-0.010
Lignin	-0.031
Hemicellulose	-0.231
ADF	-0.259

Positive coefficients reflected a direct contribution to the canonical function, while negative coefficients indicated an inverse association with the corresponding canonical variable.

The distribution of positive and negative coefficients in CV1 shows opposite contributions between structural fiber fractions and non-structural nutrient components within the canonical function.

Canonical cross-loadings

Canonical cross-loadings represent the correlation between the original variables and the canonical functions, enabling the identification of variables that best explain the relationship between sets X (predictors) and Y (responses)[13,16]. The (Table 3) reports the canonical cross-loadings for CV1 for both the predictor set (nutritional variables, X) and the response set (digestibility and energy parameters, Y).

Table 3. Canonical cross-loadings for the First Function (CV1).

Variable	Canonical Cross (CV1)
Crude Protein	0.592
Ether Extract	0.475
NDF	-0.889
ADF	-0.881
Lignin	-0.834
Hemicellulose	-0.670
Total Starch	0.712
Dry Matter Digest	0.999
ME-r	0.999

As shown in (Table 3), the highest absolute cross-loadings within the predictor set corresponded to NDF, ADF, and lignin, whereas Dry Matter Digestibility and ME-r showed values close to 1 in the response set.

Predictive performance of the Random Forest

The developed Random Forest model demonstrated excellent performance in predicting the digestibility of dry matter digestibility (DMD) and ME-r. The results of the goodness-of-fit indicators are summarized in (**Table 4**).

Table 4. Performance Metrics of the Random Forest Model

Variable	RMSE	MAE	R^2
DMD	1.860	0.630	0.956
ME-r	0.076	0.026	0.956

RMSE: Root Mean Square Error; **MAE**: Absolute Error; $\mathbf{R^2}$: Coefficient of Determination. **DMD**: Dry Matter Digestibility, **ME-r**: Metabolizable Energy for ruminants

The high R^2 values (>0.95) for both variables indicate that the model explains more than 95% of the variability observed in the test data, demonstrating an outstanding fit. Moreover, the low RMSE and MAE values further support the model's accuracy and reliability in predictive tasks.

The quality of the model fit was also evaluated using scatter plots of observed versus predicted values (**Figure 1**). In both cases, a strong alignment of the points with the identity line (y=x), is observed, indicating accurate predictions with no apparent bias.

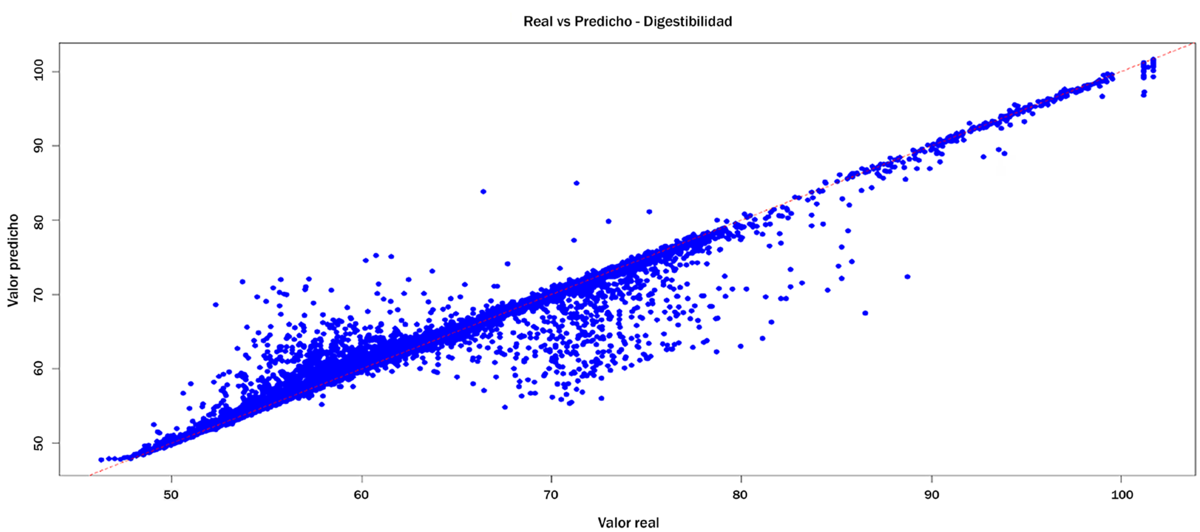

Figure 1. Dry Matter Digestibility.

(**Figure 1 and 2**). Relationship between observed and predicted values for dry matter digestibility (**Figure 1**) and ME-r (**Figure 2**) using the Random Forest model.

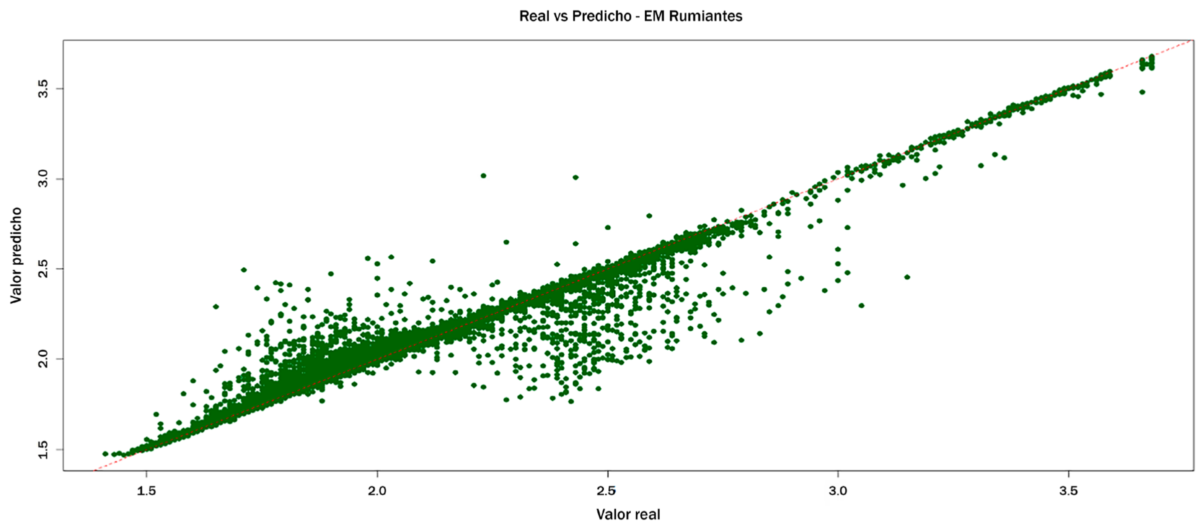

Figure 2. Metabolizable Energy.

The variable importance analysis indicated that NDF (Neutral Detergent Fiber), lignin, and crude protein were the most influential predictors for both response variables. These results are consistent with previous literature[28], where fiber composition has been consistently associated with digestibility and energy efficiency of forages.

The observed versus predicted values shown in (**Figure 1**) exhibit a high concentration of points around the identity line, indicating an accurate fit with no systematic over- or underestimation trends. This visual pattern confirms the robustness of the trained models and their potential applicability for rapid and reliable predictions in forage improvement programs and diet formulation

DISCUSSION

The results obtained confirm that the fiber composition (NDF, ADF, lignin, and hemicellulose) is the main restrictive factor for digestibility and ME-r in tropical forages. The canonical coefficients of CV1 showed negative associations for structural fiber fractions and positive loadings for crude protein and starch, reflecting the multivariate structure identified between nutritional composition and energetic performance.

This finding is consistent with reports by Jung and Allen[28], who indicated that lignin interferes with the hydrolysis of structural polysaccharides, thereby reducing energy availability. Likewise, Gomes et al[29] and Maskalova et al.[30] highlight that lignin accumulation significantly limits the feed utilization efficiency in ruminants. The negative canonical coefficients and high absolute cross-loadings observed for NDF, ADF, and lignin in (**Table 3**) are coherent with this structural limitation effect.

Lourencon et al.[26] and Guo et al.[31] reported similar patterns in tropical forages, where a higher fiber fraction reduced digestibility. In this context, the opposition observed between structural fiber fractions and non-structural nutrients within CV1 represents a gradient of forage quality captured by the canonical structure.

In contrast, the positive association of crude protein and starch with digestibility and energy contribution aligns with recent studies that emphasize their role as determining fractions in the formulation of tropical diets[31,4]. These results reinforce the importance of balancing the protein and energy content of forages as a strategy to counteract the negative effects of indigestible fiber.

Neutral Detergent Fiber (NDF), as a structural component of the plant cell wall, directly influences digestibility due to its low degradability and its relationship with lignification[2]. Shishir et al.[32] reported that higher proportions of indigestible NDF are associated with lower net energy availability. López Jara et al.[4] observed that this effect intensifies with forage maturity. Furthermore, Raffrenato et al.[33] emphasized the variability of the indigestible NDF fraction across species and management systems, reinforcing the need for robust multivariate modeling approaches.

A novel contribution of this study lies in the integration of CCA with the Random Forest algorithm. While the former allowed for the identification of multivariate relationships between nutrients and energy parameters, the latter demonstrated high predictive power (R^2>0.95)[17,18], overcoming the limitations of traditional univariate models. This methodological approach constitutes a relevant contribution to the analysis of nutritional quality of forages, offering an applicable tool for livestock production systems under tropical conditions.

The consistency between the canonical structure and the variable importance identified by Random Forest supports the robustness of the integrated modeling framework. Fiber fractions, particularly NDF and lignin, emerged as dominant predictors in both analyses, confirming their structural role in determining digestibility and ME.

However, it should be noted as a limitation that the results are based on a secondary database (AlimeTRO), without considering edaphoclimatic and management variables that could affect forage quality. Future studies that integrate these contextual variables will allow for an increase in the precision and applicability of the predictive models.

Taken together, the findings highlight the importance of combining multivariate and machine learning approaches to better understand the determinants of digestibility and ME. This approach not only strengthens scientific knowledge in animal nutrition but also provides practical tools for decision making in the formulation of more efficient and sustainable diets in tropical livestock farming.

Finally, it is worth highlighting that the integration of CCA and Random Forest provides a structured framework for modeling complex multivariate relationships in tropical forage systems. This methodological integration reinforces the analytical evaluation of nutritional quality and supports evidence-based decision making in livestock production under tropical conditions.

CONCLUSIONS

The results of this study establish the developed predictive model as a reliable and highly practical tool for estimating forage digestibility and ME-r in tropical livestock systems. Its ability to integrate multiple variables and generate accurate predictions supports more informed decision-making in diet formulation, enabling more efficient use of diverse forage resources.

By combining CCA with Random Forest algorithm, the model not only achieved strong predictive performance but also provided valuable insights into the distinct roles of fiber and protein fractions in forage quality. This methodological integration offers a solid foundation for developing feeding strategies that are better suited to real-world tropical conditions. Furthermore, the proposed approach shows significant potential for replication and adaptation in other tropical regions where production systems share similar limitations and challenges. Future enhancements, such as incorporating edaphoclimatic variables could further improve the model's applicability to more precise characterization of forages within their production contexts.

Overall, this study contributes both a practical modeling tool and a broader, more integrated perspective on animal nutrition where precision, sustainability, and data-driven strategies converge to strengthen the productivity and resilience of tropical livestock systems.

Finally, the proposed methodological integration can be applied to other multivariate prediction contexts in agricultural, biological, or computational sciences, which reinforces its value as a cross-disciplinary contribution to scientific knowledge.

ACKNOWLEDGMENTS

The authors thank El Bosque University and the Master's program in Applied Statistics and Science for the institutional and academic support that enabled the development of this research.

ETHICAL CONSIDERATIONS

This study is classified as a "no-risk" investigation, in accordance with resolution 8430 of 1993 from Colombian Ministry of Health. The analysis was conducted exclusively using secondary information from the AlimeTro database, without involving procedures on human beings or direct experimentation on animals. The data used were managed with criteria of scientific integrity, ensuring the responsible and ethical use of the information.

FUNDING

This work received no external funding for its realization.

DECLARATION OF CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest in relation to the publication of this article.

REFERENCES

1. **Ganadero**. Ganadero Agosto de 2023. Bogotá: Fedegán; 2023. Disponible en: https://bibliotecadigital.ccb.org.co/server/api/core/bitstreams/8084ec30-c67b-4ad8-8ccd-755a0c8e0ac9/content

2. **Da Cruz CH, Santos SA, de Carvalho GGP, Azevedo JAG, Detmann E, Valadares Filho SC, et al**. Estimating digestible nutrients in diets for small ruminants fed with tropical forages. Livest Sci. 2021;249:104532. doi:10.1016/j.livsci.2021.104532
3. **Francy da Costa Backsman C, Monção FP, Aspiazu I, Júnior VRR, Figueiredo Portugal A, et al**. Agronomic traits, fermentation quality, chemical composition, and silage digestibility of different forage sorghum genotypes and biomass in the semi-arid region of Brazil. J Appl Anim Res. 2025;53(1). doi:10.1080/09712119.2025.2462573
4. **López Jara AG, Reta Sánchez DG, Santana OI, Reyes Gonzáles A, Rodríguez Hernández K, et al.** Rendimiento de forraje y valor nutritivo del ensilado de forrajes alternativos y tradicionales de otoño-invierno. Rev Mex Cienc Pecu. 2024;16(1):208-23. doi:10.22319/rmcp.v16i1.6301
5. **Hernández-Arboleda X, Ortiz-Grisales S, Vivas-Arturo WF, Fernández-Romay Y, O-León O, et al.** Nutritional value and in vitro dry matter degradability in Mexican sunflower: Tithonia diversifolia Helms (Gray). Trop Subtrop Agroecosyst. 2024;27(3). doi:10.56369/tsaes.5211
6. **Smith C, Karunaratne S, Badenhorst P, Cogan N, Spangenberg G, Smith K**. Machine learning algorithms to predict forage nutritive value of in situ perennial ryegrass plants using hyperspectral canopy reflectance data. Remote Sens. 2020;12(6):60928. doi:10.3390/rs12060928
7. **Zwick M, Cardoso JA, Gutiérrez-Zapata DM, Cerón-Muñoz M, Gutiérrez JF, et al**. Pixels to pasture: Using machine learning and multispectral remote sensing to predict biomass and nutrient quality in tropical grasslands. Remote Sens Appl Soc Environ. 2024; 36: 101282. doi:10.1016/j.rsase.2024.101282
8. **Gallo A, Moschini M, Cerioli C, Masoero F**. Use of principal component analysis to classify forages and predict their calculated energy content. Animal. 2013;7(6):930-9. doi:10.1017/S1751731112002467
9. **Culqui L, Huaman Pilco AF, Juarez Contreras L, Vigo CN, Goñas M, et al**. Nutritional potential of native shrub species for cattle feeding in northeastern Peru. Rangeland Ecol Manag. 2025;98:600-8. doi:10.1016/j.rama.2024.11.004
10. **Palmonari A, Gallo A, Fustini M, Canestrari G, Masoero F, et al.** Estimation of the indigestible fiber in different forage types. J Anim Sci. 2016;94:248-54.
11. **Ministerio de Agricultura y Desarrollo Rural. Alimentos del trópico para alimentación animal** - AlimenTro. Datos Abiertos Colombia. 2020. Disponible en: https://www.datos.gov.co/Agricultura-y-Desarrollo-Rural/Alimentos-Del-Tr-Pico-Para-Alimentaci-n-Animal-Ali/6arb-D547
12. **Villalpando P, Castillo JA, Cortez K**. Análisis de correlación canónica (ACC) e investigación científica. ResearchGate. 2007. Disponible en: https://www.researchgate.net/publication/315813696
13. **Hair J, Black WC, Babin BJ, Anderson RE**. Multivariate data analysis. 7ª ed. New Jersey: Pearson Education; 2010.
14. **González Ariza A, Navas González FJ, Arando Arbulu A, León Jurado JM, Delgado Bermejo JV, et al**. Discriminant canonical analysis as a tool to determine traces of endangered native hen breed introgression through egg hatchability phenomics. Anim Biosci. 2022;35(7):915-24. doi:10.5713/ab.22.0163
15. **Bock H, Baier AD, Gaul CW, Critchley KF, Vichi MM, et al**. Data Analysis, Machine Learning and Knowledge Discovery. New York. Springer, 2014. doi:10.1007/978-3-319-01595-8

16. **Díaz Monroy LG, Morales Rivera MA**. Análisis estadístico de datos multivariados. 1ª ed. Bogotá: Universidad Nacional de Colombia, 2012.
17. **Breiman L**. Random forests. Mach Learn. 2001;45(1):5-32.
18. **Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, et al**. Random forests for global and regional crop yield predictions. PLoS One. 2016;11(6):e0156571. doi:10.1371/journal.pone.0156571
19. **Pepeta BN, Moyo M, Adejoro FA, Hassen A, Nsahlai IV**. Techniques used to determine botanical composition, intake, and digestibility of forages by ruminants. Agronomy. 2022;12(10):2456. doi:10.3390/agronomy12102456
20. **Lee M, Kim DH, Seo S, Tedeschi LO.** Development of machine learning models for estimating metabolizable protein supply from feed in lactating dairy cows. Animals. 2025;15(5):1068. doi:10.3390/ani15050687
21. **Sherry A, Henson RK**. Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. J Pers Assess. 2005;84(1):37-48. doi:10.1207/s15327752jpa8401
22. **Jiang MZ, Aguet F, Ardlie K, Chen J, Cornell E, et al**. Canonical correlation analysis for multi-omics: Application to cross-cohort analysis. PLoS Genet. 2023;19(5):e1010517. doi:10.1371/journal.pgen.1010517
23. **Martínez Lobo DS**. Análisis de la relación entre las pruebas Saber Pro y los cursos realizados por estudiantes de licenciatura en matemáticas utilizando correlación canónica. Bucaramanga: Universidad Industrial de Santander; 2013.
24. **Quintero E, Martínez D**. Data Science in Secondary Education: Exploring Correlations and Predicting Saber 11 Test Results from the Formative Process. Comunicaciones en Estadística. 2025; 18(2): 33-42. doi.org/10.15332/23393076.11820
25. **Diel MI, Dal'Col Lúcio A, Lambrecht DM, Vinícius M, Pinheiro M, et al**. Canonical correlations in agricultural research: Method of interpretation used leads to greater reliability of results. Int J Innov Educ Res. 2020;7:1-10. Doi:10.31686/ijier.vol8.iss7.2464
26. **Lourencon RV, Patra AK, Ribeiro LPS, Puchala R, Wang W, et al.** Effects of the level and source of dietary physically effective fiber on feed intake, nutrient utilization, heat energy, ruminal fermentation, and milk production by Alpine goats. Anim Nutr. 2024;17:312-24. doi:10.1016/j.aninu.2024.02.002
27. **Leishman EM, Sahar M, Cieslar S, Darani P, Ellis JL**. What the hay: predicting equine voluntary forage intake using a meta-analysis approach. Animal. 2024;18(9):101266. doi:10.1016/j.animal.2024.101266
28. **Jung HG, Allen MS**. Characteristics of plant cell walls affecting intake and digestibility of forages by ruminants. J Anim Sci. 1995 Sep;73(9):2774-90. doi:10.2527/1995.7392774x
29. **Gomes DI, Detmann E, Valadares Filho SC, Fukushima RS, de Souza MA, et al**. Evaluation of lignin contents in tropical forages using different analytical methods and their correlations with degradation of insoluble fiber. Anim Feed Sci Technol. 2011;168(3-4):206-22. doi:10.1016/j.anifeedsci.2011.05.001
30. **Maskaľová I, Vajda V, Timkovičová Lacková P**. Estimation of ruminal digestibility of nutrient and intestinal digestibility of un-degradable proteins at different feedstuffs. Acta Fytotechn Zootechn. 2024;27(1):8-17. doi:10.15414/afz.2024.27.01.8-17.

31. **Guo X, Sun L, Zheng Z, Diao X, He L, et al.** Study on rumen degradability and intestinal digestibility of mutton sheep diets with different concentrate-to-forage ratios and nonfiber carbohydrates/neutral detergent fiber ratios. Animals. 2024;14(19):2816. doi:10.3390/ani14192816
32. **Shishir MSR, Cullen B, Brodie G, Zhong R, Cheng L**. Potential of feeding microwave-treated forage hays to improve sheep intake, digestion, nitrogen partitioning, and metabolism. Anim Feed Sci Technol. 2024;315:116008. doi:10.1016/j.anifeedsci.2024.116008
33. **Raffrenato E, Nicholson CF, Van Amburgh ME.** Development of a mathematical model to predict pool sizes and rates of digestion of 2 pools of digestible neutral detergent fiber and an undigested neutral detergent fiber fraction within various forages. J Dairy Sci. 2019;102(1):351-64. doi:10.3168/jds.2018-15102