RESEARCH ARTICLE

Balancing accuracy, interpretability, and stability in machine-learning models: Live-weight prediction of Andean sheep from morphometric traits

Jordan Ninahuanca[1*] [iD]; Edgar Garcia-Olarte[1] [iD]; Ide Unchupaico Payano[1] [iD]; Vicky Sarapura[2] [iD]
Kevin Zenteno Vera[1] [iD]; Carlos Quispe Eulogio[3] [iD]; Edith Ancco Gomez[3] [iD]; Mohamed Mohamed M. Hadi[4] [iD]
Carolina Miranda-Torpoco[4] [iD]; Wilhelm Guerra Condor[3] [iD]

[1] Facultad de Zootecnia, Universidad Nacional del Centro del Perú, Huancayo, 12006, Av. Mariscal Castilla N° 3909, Junín, Perú.
[2] Facultad de Ciencias Forestales y del Ambiente, Universidad Nacional del Centro del Perú, Huancayo, 12006, Av. Mariscal Castilla N° 3909, Junín, Perú.
[3] Facultad de Ciencias de la Salud, Universidad Peruana Los Andes, Huancayo, 12002, Perú.
[4] Facultad de Ingeniería, Universidad Peruana Los Andes, Huancayo, 12002, Perú.

* Corresponding author: jninahuanca@uncp.edu.pe (J. Ninahuanca).

Abstract

The objective of this research was to predict the live weight of Corriedale lambs using morphological measurements and machine learning algorithms. A total of 291 five-month-old lambs from the Corpacancha Production Unit of SAIS PACHACÚTEC SAC were used. These animals represented a homogeneous group in terms of age, sex, and genetics, as they belonged to the Corriedale breed and were offspring of "Category A" ewes. Morphological measurements recorded included Body Length (BL), Withers Height (WH), Thoracic Girth (TG), Rump Width (RW), Abdominal Girth (AG), Cannon Bone Length (CBL), Chest Depth (CD), and Live Weight (LW). The models evaluated were Multiple Linear Regression, Ridge Regression, Decision Trees, Random Forest, and XGBoost. The comparative analysis of the machine learning models identified ModG and Ridge as the most accurate and stable options, standing out for their low Mean Squared Error (MSE = 0.083) and Root Mean Squared Error (RMSE ≈ 0.287 − 0.288). Additionally, they exhibited the highest coefficients of determination (R^2 = 0.89, R_{Adj}^2 = 0.88), indicating excellent predictive capability and data fit. Their low coefficient of variation (CV%) confirms their stability, establishing them as the best choices for applications where precision is paramount, such as predicting critical values in production processes and high-demand scientific studies. While XGBoost proved to be a robust alternative with an MSE of 0.119, an RMSE of 0.345, and a relative error of 2.22%. These findings confirm that prioritizing models that balance accuracy, interpretability, and stability enable faster, data-driven decision-making in Corriedale sheep production. Such an approach optimizes feed allocation, classifies lambs by market weight, and promptly detects growth deviations, thereby improving overall flock profitability.

Keywords: biometrics; predictive models; mathematical models; young sheep; zoometrical.

1. Introduction

Across the Andean highlands of South America, sheep husbandry remains a cornerstone of rural livelihoods, and ovine meat is increasingly preferred over other animal proteins by local consumers (Ninahuanca Carhuas et al., 2025). The regional sheep sector nevertheless faces persistent challenges: economic losses linked to fluctuating wool prices and declining competitiveness in the global textile market have eroded profitability (Bailey et al., 2021). These pressures have driven producers to rethink their strategies and diversify revenue sources, as noted by Ozen et al. (2024). Within this regional context, Peruvian enterprises, cooperatives and smallholders are now prioritizing the finishing and sale of Corriedale lambs, whose carcasses secure favorable prices and exhibit relatively low market variability (Carhuas et al., 2024). This strategic shift creates opportunities to enhance profitability while demanding management practices that maximize growth performance and preserve carcass quality.

In this new approach, the accurate estimation of live weight in sheep has become a central aspect of decision-making related to feeding, health management, and determining the optimal time for slaughter (Contreras et al., 2024). However, traditional methods,

such as the use of scales, present multiple limitations in large-scale production systems (**Martins et al., 2020**; **Dang et al., 2022**). These include the costs associated with equipment, the logistical challenges of weighing large numbers of animals, and the stress caused by handling, which can negatively affect both animal welfare and productivity (**Jurkovich et al., 2024**). These challenges underscore the need to explore alternative methods that enable rapid, accurate, and non-invasive estimation of live weight. A promising solution lies in the use of morphological measurements, such as body length, withers height, and thoracic circumference, which are closely related to the live weight of sheep (**Gomes et al., 2016**; **Wang et al., 2021**; **Contreras et al., 2024**). These variables can be easily collected in the field and represent an accessible option for overcoming the limitations of traditional methods. However, the relationship between morphological measurements and live weight is neither linear nor uniform, complicating the application of conventional predictive models based on simple regressions or descriptive analyses (**García-Medina & Aguayo-Moreno, 2024**).

Machine learning algorithms, a term coined by **Samuel (2000)**, who categorized them into three types of learning (reinforcement learning, supervised learning, and unsupervised learning) have emerged as innovative mathematical tools with the potential to revolutionize live weight prediction in sheep (**Vlaicu et al., 2024**). Among these, supervised learning algorithms stand out for their ability to learn from labeled datasets to approximate the mapping function between inputs (features) and outputs (target values) (**Dang et al., 2022**). Their capacity to analyze large datasets, capture non-linear relationships and generate accurate predictions in dynamic scenarios has been widely documented, establishing them as high-impact methods in leading indexed journals (**García-Medina & Aguayo-Moreno, 2024**). In animal production, supervised learning enables flock-specific predictive models that enhance efficiency and sustainability (**Peña-Avelino et al., 2021**), and its application already extends beyond weight estimation to areas such as genetic selection and performance evaluation (**Qin et al., 2024**). Empirical studies illustrate this promise: **Ozen et al. (2024)** applied shrinkage regressions and tree-based ensembles to 100 six-month-old Akkaraman lambs, identifying Random Forest as the best performer, while **Kozaklı et al. (2024)** compared nine machine-learning algorithms with multiple linear regression in 25 316 post-weaning Akkaraman lambs and likewise concluded that ensemble models outperformed linear approaches. Although these investigations confirm the utility of machine learning, they are confined to low-altitude

Turkish flocks, overlook Corriedale genetics and focus mainly on raw accuracy-leaving interpretability, numerical stability and practical management implications largely unaddressed.

The objective of this study was to evaluate the most accurate machine learning model for predicting live weight in sheep.

2. Methodology

Animals and distribution

The study used 291 five-month-old male Corriedale lambs raised at the Corpacancha Production Unit (11° 21' 46" S, 76° 13' 11" W) in the Marcapomacocha district, Yauli Province, Junín Region, central Andes of Peru. This single location ensured a homogeneous cohort in age, sex and genetics, as all lambs were offspring of "Category A" ewes. The flock grazed exclusively on natural pastures at 4 149 m above sea level, where mean air temperatures range from –0.6 °C to 11 °C and annual precipitation averages of 700 mm. Animals were maintained under controlled sanitary conditions and routinely dewormed against taeniasis and fascioliasis. The production unit was selected for its well-documented management practices and reliable zootechnical records, guaranteeing high-quality morphometric and live-weight data for predictive analysis.

Data collection

At 6:00 a.m., measurements (cm) were recorded prior to the animal's feed intake. For weight (kg), a livestock scale with a capacity of 150 kg (OMEGA TP model, sensitivity ± 0.01 g) was used. The animals were positioned on a flat surface, standing in a relaxed manner, with their feet firmly placed on the ground (natural body position), following the recommendations of **Lee et al. (2022)**. Body Length (BL) was measured as the distance in centimeters between the base of the tail and the base of the neck (**Karna et al., 2024**). Withers Height (WH) was measured as the distance (cm) from the ground to the highest point of the back (withers) (**Cam et al., 2010**). Thoracic Girth (TG) was measured as the distance (cm) around the chest, just behind the forelimbs (**Contreras et al., 2024**). Rump Width (RW) was measured as the width (cm) of the animal's rear section (**Gonçalves et al., 2025**). Abdominal Girth (AG) was measured as the distance (cm) around the animal's abdomen (**Lee et al., 2022**). Cannon Bone Length (CBL) was measured as the length (cm) from the knee joint to the hoof (**Mahmud et al., 2014**). Chest Depth (CD) was measured as the distance (cm) from the top of the back to the bottom of the chest (**Karna et al., 2005**). The details are shown in **Figure 1**.
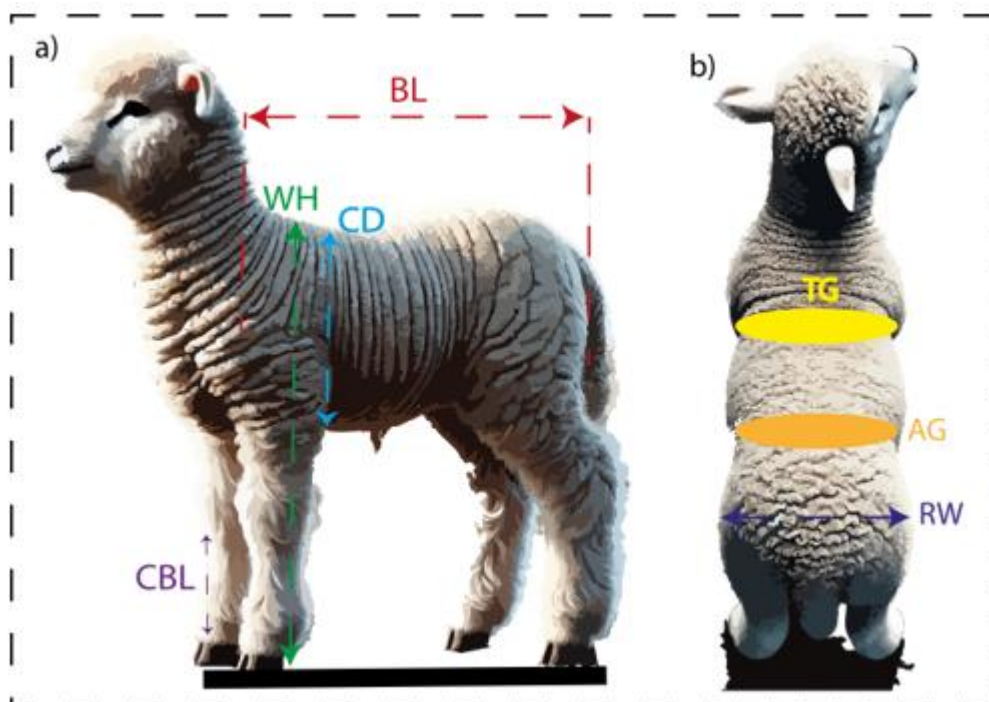
Figure 1. Lamb body measurements.

Predictive machine learning models

Five main models were employed to predict live weight in sheep: Multiple Linear Regression (MLR), Ridge Regression, Decision Trees (DT), Random Forest (RF), and XGBoost. Multiple Linear Regression (MLR) served as the baseline model, enabling the establishment of linear relationships between morphological variables and live weight (**Choque, 2024**). To address potential multicollinearity issues and improve model stability, Ridge Regression was used, a regularization method that penalizes predictor variable coefficients to reduce bias and variance (**Lipovetsky, 2021**). Decision Trees were selected for their ability to split data into homogeneous subsets through hierarchical rules based on the most relevant variables, offering interpretability and ease of use (**Lee et al., 2022**). The Random Forest model, which combines multiple decision trees using a bagging approach, improved model accuracy by reducing overfitting and capturing complex non-linear interactions (**Hu & Szymczak, 2023**). Finally, XGBoost, a boosting-based algorithm, was implemented to iteratively adjust decision trees, optimizing error minimization and enhancing efficiency in large datasets (**Kumar et al., 2023**). Model performance was evaluated using accuracy metrics:

a. Coefficient of Determination (R^2)

$$R^2 = \left[1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}\right] x\ 100$$

b. Adjusted Coefficient of Determination (Adjusted R^2)

$$R_{adj}^2 = \left[1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}\right] x\ 100$$

c. Standard Deviation Ratio (SDR)

$$SD_{RATIO} = \sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

d. Relative Approximation Error (RAE)

$$RAE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i)^2}}$$

e. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}}$$

f. Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

g. Coefficient of Variation (CV)

$$CV(\%) = \sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2}{\bar{Y}}} x\ 100$$

Where Y_i is the observed live weight (kg) of the sheep; $\hat{Y}_i$ is the predicted live weight of the iii-th sheep; $\bar{Y}$ is the mean of the actual live weight values of the sheep; ε_i is the residual value of the iii-th sheep; k is the number of significant independent variables in the model; n is the total number of observations. The residual value for each sheep is expressed as $\varepsilon_i = Y_i - \hat{Y}_i$.

3. Results and discussion

Biometric measurements

The results obtained (Table 1) highlight the relevance of the biometric characteristics analyzed in the population of 5-month-old male lambs. The live weight, with an average of 12.44 kg and a coefficient of variation (CV) of 7.04%, shows low dispersion, suggesting homogeneous management in terms of feeding, health, and environmental conditions. This aligns with previous studies emphasizing the importance of uniform management during early stages of sheep development to ensure optimal and consistent growth (Stewart et al., 2005). On the other hand, body dimensions such as body length (39.99 cm, CV 7.36%) and thoracic girth (39.70 cm, CV 7.08%) exhibit low relative variability, reflecting the uniformity of the population. These variables have been previously identified as key indicators of growth and body condition in sheep, directly influencing productivity (Contreras et al., 2024). Similarly, withers height (35.19 cm, CV 8.60%) and chest depth (17.48 cm, CV 8.28%) show slightly higher variability, potentially associated with genetic differences or environmental factors affecting structural development. In contrast, traits such as rump width (12.52 cm, CV 11.87%) and cannon bone length (12.50 cm, CV 11.39%) exhibit higher coefficients of variation, suggesting significant heterogeneity among individuals. This finding is consistent with research linking these measurements to genetics and productive potential, particularly in extensive management systems where environmental factors have a greater impact on muscular and skeletal development (Peña-Avelino et al., 2021). Abdominal girth (47.50 cm, CV 9.44%) also shows moderate variability, which could be influenced by differences in body condition and nutritional management. This parameter is of particular interest in evaluating digestive capacity and overall condition in sheep, as it is associated with animal welfare and productive efficiency. The results suggest that the studied population presents an adequate level of homogeneity to establish predictive relationships between biometric variables and live weight.

Correlations

The correlation matrix shows the relationship between live weight and the biometric variables measured in the male lambs (Figure 2). Specifically, the following correlations with live weight are observed. Abdominal Girth (AG): This variable exhibits the strongest positive correlation with weight, with a value of 0.71, indicating a strong relationship. This suggests that as abdominal girth increases, weight also tends to rise. This result aligns with the idea that abdominal girth largely reflects body condition and mass accumulation. Thoracic Girth (TG): It shows a moderate correlation with weight, with a value of 0.45. This implies that thoracic girth is also a good indicator of live weight, likely related to muscle development in the thoracic region. Body Length (BL): It has a moderate correlation with weight, with a value of 0.36. This suggests that body length can be an indirect indicator of weight, although its relationship is not as strong as that of abdominal or thoracic girth. Withers Height (WH): This variable shows a low correlation with weight, with a value of 0.30, indicating that while there is some relationship between withers height and weight, it is not as significant compared to other variables. Chest Depth (CD): It has a low correlation with weight, at 0.26. This suggests that chest depth has a limited impact on live weight in this population of lambs. Rump Width (RW): This variable has the weakest correlation with weight, with a value of 0.18, indicating a weak relationship with live weight in the evaluated lambs. Cannon Bone Length (CBL): It shows a very low correlation with weight, with a value of 0.06, suggesting that this variable is not a significant indicator of live weight.

From Table 2, abdominal girth (AG) and thoracic girth (TG) emerge as the most influential variables with the greatest statistical significance in relation to weight. On the other hand, cannon bone length (CBL) does not show a significant correlation, indicating that its inclusion in predictive models may be unnecessary. These observations help prioritize key variables for subsequent analyses.

Table 1
Descriptive analysis of the variables studied

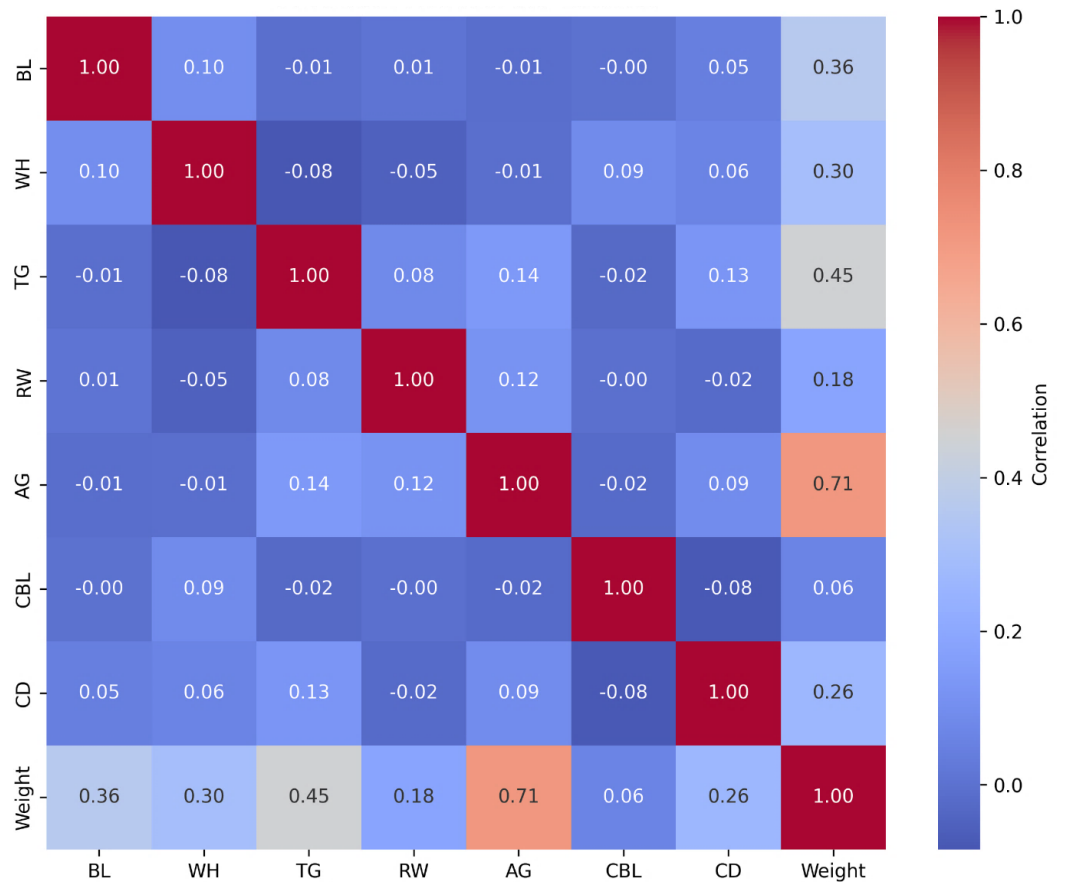| Variable	n	Mean ± sd	CV (%)	Min	Max
Weight (kg)	291	12.44 ± 0.88	7.04	10	15
Body Length (cm)	291	39.99 ± 2.94	7.36	35.10	44.90
Withers Height (cm)	291	35.19 ± 3.03	8.60	30.10	40.00
Thoracic Girth (cm)	291	39.70 ± 2.80	7.08	35.00	45.00
Rump Width (cm)	291	12.52 ± 1.49	11.87	10.10	15.00
Abdominal Girth (cm)	291	47.50 ± 4.48	9.44	40.10	55.00
Cannon Bone Length (cm)	291	12.50 ± 1.42	11.39	10.00	15.00
Chest Depth (cm)	291	17.48 ± 1.45	8.28	15.00	20.00

Figure 2. Correlation Matrix of the Variable. 0 indicates no relationship and 1 indicates a very strong relationship. Body Length (BL), Withers Height (WH), Thoracic Girth (TG), Rump Width (RW), Abdominal Girth (AG), Cannon Bone Length (CBL), Chest Depth (CD).

Table 2
Correlation of variables with weight, most influential variables

	PA	PT	LC	AC	PP	AG	LCA
Peso	0.71	0.45	0.36	0.30	0.26	0.18	0.06
p-value	9.45e-47	3.93e-16	2.12e-10	1.90e-07	0.000005	0.001671	0.30

p-value < 0.05 implies a significant correlation at a confidence level of 59%.
p-value > 0.05 there is insufficient evidence to conclude that the correlation is significant
Body Length (BL), Withers Height (WH), Thoracic Girth (TG), Rump Width (RW), Abdominal Girth (AG), Cannon Bone Length (CBL), Chest Depth (CD).

Multiple Linear Regression

In the evaluation of different multiple linear regression models for predicting the live weight of lambs, six simplified configurations of the general model (ModG) were compared by progressively removing independent variables. The regression models started with the general model, systematically eliminating the least correlated variables with weight to assess which model is ideal, as shown in **Table 3**.

The results obtained from evaluating different multiple linear regression models to predict the live weight of lambs highlight the importance of considering multiple variables to capture the inherent complexity of the phenomenon. The general model (ModG), which includes all independent variables (BL, WH, TG, RW, AG, CBL, CD), emerges as the most robust, with an R^2 of 0.987, an AIC of 78.20, and the lowest relative error (1.844%). This suggests that all included variables contribute significantly to explaining the variability in live weight, consistent with previous research emphasizing the importance of integrating multiple morphometric measurements to improve the accuracy of predictive models (**Courtenay et al., 2019**; **Arabameri et al., 2020**). When analyzing the simplified models, it is observed that the exclusion of CBL (Model 1) does not significantly affect predictive performance, as it maintains an R^2 of 0.986 and an AIC of 81.81. This aligns with studies reporting that cannon bone length has a moderate correlation with live weight in young animals, making it less relevant compared to other measures such as thoracic or abdominal

girth (Salamanca-Carreño et al., 2024). However, when additional variables such as CD and RW are removed (Models 2 and 3), a progressive loss in model fit is observed, with decreases in R^2 to 0.968 and 0.936, respectively. This underscores the relevance of these variables in predicting live weight, as they are associated with key body parameters related to the metabolic and structural capacity of animals (Frizzarin et al., 2021). In the most reduced models, such as Mod5 and Mod6, where only TG (Thoracic Girth) and AG (Abdominal Girth) are included, a significant decrease in predictive capacity is observed (R^2 of 0.655 and 0.428, respectively). However, these models highlight the importance of TG and AG as the most influential variables, aligning with studies identifying these measurements as the best predictors of live weight in sheep production systems (Pannier et al., 2025). This is because these dimensions are directly related to body volume, which is a reliable indicator of live weight. The analysis of the AIC also provides critical insights into the balance between simplicity and accuracy. Although ModG has the lowest AIC (-78.20), positioning it as the most suitable model in terms of overall fit, models like Mod1 represent viable alternatives with a reasonable trade-off between simplicity and predictive power. This approach is consistent with the literature, where the removal of redundant variables is prioritized to avoid issues of multicollinearity and overfitting (Mokri et al., 2025). In terms of stability indicators, the coefficient of variation (CV%) and relative error progressively increase as variables are reduced in the models. This indicates that simplified models lose stability and precision, with CV% reaching 5.231% and relative error rising to 4.189% in Mod6. These results reinforce the need to balance the number of included variables with the explanatory power of the model (García-Medina & Aguayo-Moreno et al., 2024), as highlighted by studies on the prediction of productive parameters in sheep.

Machine Learning

Table 4 provides a comprehensive comparison of the results obtained from five predictive models used to estimate the live weight of lambs: ModG (General Multiple Linear Regression), Ridge Regression, Decision Trees, Random Forest, and XGBoost. Key performance metrics are summarized, including MSE (Mean Squared Error), RMSE (Root Mean Squared Error), coefficients of determination (R^2 and Adjusted R^2), Relative Standard Deviation (SDR), Relative Error (%), and Coefficient of Variation (%).

Tabla 3
Multiple Linear Regression Models

Model	AIC	MSE	RMSE	R^2	R^2A	SDR	E.R (%)	CV(%)
ModG	78.20	0.083	0.287	0.890	0.880	0.925	1.844	6.55
Mod1	81.38	0.089	0.299	0.88	0.872	0.919	1.93	6.51
Mod2	97.956	0.093	0.306	0.876	0.868	0.910	1.997	6.43
Mod3	120.88	0.104	0.3223	0.861	0.855	0.908	2.138	6.428
Mod4	237.76	0.1519	0.389	0.798	0.791	0.869	2.59	6.149
Mod5	312.366	0.309	0.555	0.589	0.579	0.805	3.784	5.654
Mod6	370.330	0.425	0.652	0.435	0.428	0.751	4.189	5.231

AIC = Akaike's Information Criterion, MSE = Mean Squared Error, RMSE = Root Mean Squared Error, R2 = Coefficient of Determination, R2A = Adjusted Coefficient of Determination, SDR = Standard Deviation Ratio, ER = Relative Error, CV = Coefficient of Variance.
ModG: $Y=\mu+\beta_1(LC)+\beta_2(AC)+\beta_3(PT)+\beta_4(AG)+\beta_5(PA)+\beta_6(LCA)+\beta_7(PP)+\epsilon$;
Mod1: $Y=\mu+\beta_1(LC)+\beta_2(AC)+\beta_3(PT)+\beta_4(AG)+\beta_5(PA)+\beta_6(PP)+\epsilon$;
Mod2: $Y=\mu+\beta_1(LC)+\beta_2(AC)+\beta_3(PT)+\beta_4(PA)+\beta_5(PP)+\epsilon$;
Mod3: $Y=\mu+\beta_1(LC)+\beta_2(AC)+\beta_3(PT)+\beta_4(PA)+\epsilon$;
Mod4: $Y=\mu+\beta_1(LC)+\beta_2(PT)+\beta_3(PA)+\epsilon$;
Mod5: $Y=\mu+\beta_1(PT)+\beta_2(PA)+\epsilon$;
Mod6: $Y=\mu+\beta_1(PA)+\epsilon$;

Table 4
Comparison of Machine Learning models

Model	MSE	RMSE	R^2	R^2Adj	Standard Deviation R	Relative Error (%)	CV(%)
ModG	0.083	0.287	0.89	0.880	0.925	1.844	6.55
Ridge	0.083	0.288	0.890	0.880	0.920	1.845	6.52
DT	0.240	0.490	0.681	0.639	0.039	3.20	5.89
RF	0.148	0.384	0.803	0.79	0.031	2.57	5.67
XGBoost	0.119	0.345	0.842	0.828	0.0275	2.19	5.97

DT= Decision Tree, RF= Random Forest. AD: Best Hyperparameters ({'criterion': 'absolute_error', 'max_depth': 10, 'max_features': 'auto', 'min_samples_leaf': 5, 'min_samples_split': 2=}). XG: Best Hyperparameters: {'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 150, 'subsample': 0.8}. RF: Best Hyperparameters: {'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50}.

ModG (General Multiple Linear Regression) and Ridge Regression (**Table 4**) demonstrated the best overall performance, as evidenced by the lowest MSE values (0.083) and RMSE values (0.287 and 0.288, respectively), as well as the highest R^2 (0.89) and Adjusted R^2 (0.88). These metrics indicate that both models exhibit superior predictive capability and effectively capture the relationships present in the data. Furthermore, the relative error below 2% (1.844% and 1.845%) and the low CV% values (6.55% and 6.52%) reinforce their stability and reliability for applications where precision is critical. These findings align with previous studies highlight the performance of penalized linear models such as Ridge Regression in complex datasets with high multicollinearity.

The Decision Trees model demonstrated the most limited performance, with the highest MSE (0.240) and RMSE (0.490) values, as well as the lowest R^2 (0.681) and Adjusted R^2 (0.639) among the evaluated models. Although its residual standard deviation is low (0.039), the relative error (3.20%) and CV% (5.89%) reflect inadequate predictive capacity compared to the other models. This result could be attributed to the model's lack of robustness against data variability and its limited ability to capture complex interactions. Nevertheless, the simplicity and interpretability of Decision Trees may make them useful in specific scenarios where these factors are a priority. The decision tree (**Figure 3**) hierarchically reflects the relationship between the predictor variables and weight, highlighting Abdominal Girth (AG) as the most influential characteristic in the prediction, followed by variables such as Body Length (BL), Thoracic Girth (TG), and Withers Height (WH), which also show significant contributions. The tree's structure, with clear divisions and controlled depth, captures relevant patterns in the data, effectively reducing squared errors at the terminal nodes. However, some nodes with higher errors suggest the presence of noise or the need for additional data to improve accuracy. According to **Long et al. (2025)**, random forest models outperform single decision trees by combining multiple trees and averaging their predictions, which enhances model accuracy and stability. One of the strengths of the decision tree is its visual interpretability, allowing users to observe how variables influence decisions at each level. This aspect is crucial for practical decision-making in the field, where sheep producers can quickly identify which attributes need to be prioritized to optimize the weight of their animals.

The Random Forest (**Table 4**) model also demonstrated acceptable results, with an MSE of 0.148 and an RMSE of 0.384, as well as an R^2 (0.803) and Adjusted R^2 (0.79), reflecting good explanatory power. However, this model exhibited a higher relative error (2.57%) and a slightly lower CV% (5.67%), which could limit its application in contexts where maximum precision is required. Nonetheless, its performance remains competitive and suitable for nonlinear and high-dimensional problems.

Figure 4 provides a comprehensive evaluation of the Random Forest model applied to predicting live weight in lambs through four graphs highlighting different aspects of the model. **Figure 4a** Cross-Validation Curve: Number of Estimators. This graph shows how the average R^2 obtained through cross-validation varies with the number of estimators in the model. A significant improvement in model performance is observed as the number of estimators increases, stabilizing around 100 estimators with an average R^2 close to 0.815, emphasizing the importance of optimizing the number of estimators to balance accuracy and computational efficiency. **Figure 4b** Predictions vs. Actual Values compares the model's predicted values with the actual weight values. The red reference line represents perfect equality between predictions and actual values, and the green points clustered near this line indicate high accuracy in most cases, with slight deviations at some extreme points. **Figure 4c** Distribution of Errors (Residuals) displays a histogram of errors, where most errors are concentrated around zero, indicating generally accurate predictions, although some larger errors suggest potential areas for model improvement. **Figure 4d** Feature Importance (Heatmap) presents the relative importance of each predictor variable in the Random Forest model. The most influential variable is Abdominal Girth (AG), with an important value of 0.55, followed by Thoracic Girth (TG) at 0.14 and Withers Height (WH) at 0.11. Variables such as Rump Width (RW) and Cannon Bone Length (CBL) show lower impact, suggesting they may be less relevant for weight prediction in this model. The application of the Random Forest model for predicting live weight in 5-month-old lambs provides an interesting perspective on the potential and limitations of this approach. This algorithm is widely recognized for its ability to handle nonlinear data and its robustness against overfitting using multiple decision trees (**Shen et al., 2025**; **Jarupunphol et al., 2025**). However, the results obtained in this study suggest that, although the model demonstrates reasonable performance, there are areas for improvement that warrant attention. With a coefficient of determination (R^2) of 0.80 and an adjusted R^2 of 0.79, the model explains approximately 80% of the variability in the data, reflecting a good overall fit.
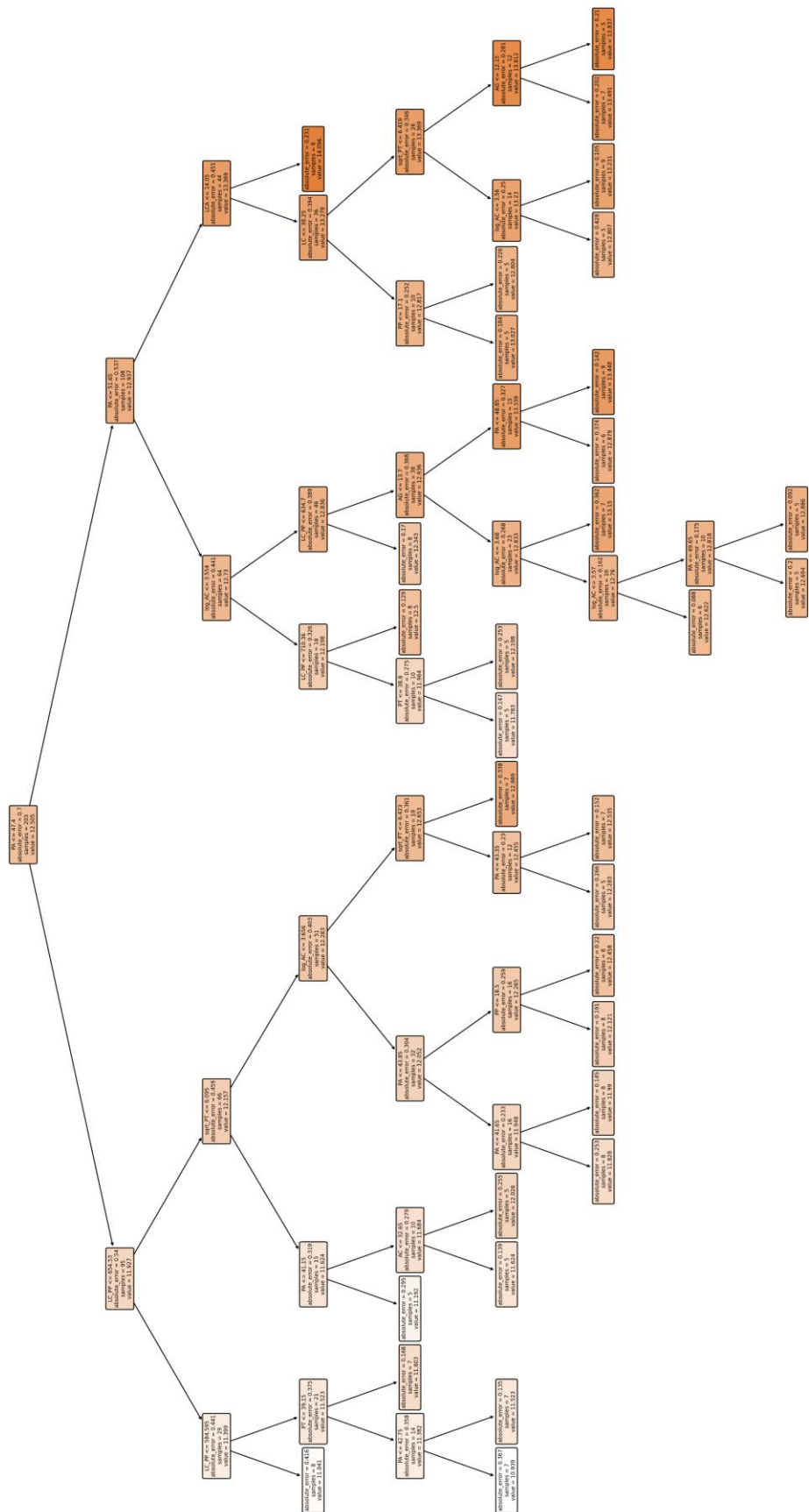
Figure 3. Decision tree generated for weight prediction in morphology analysis (See in high quality in **Supplementary Material**).
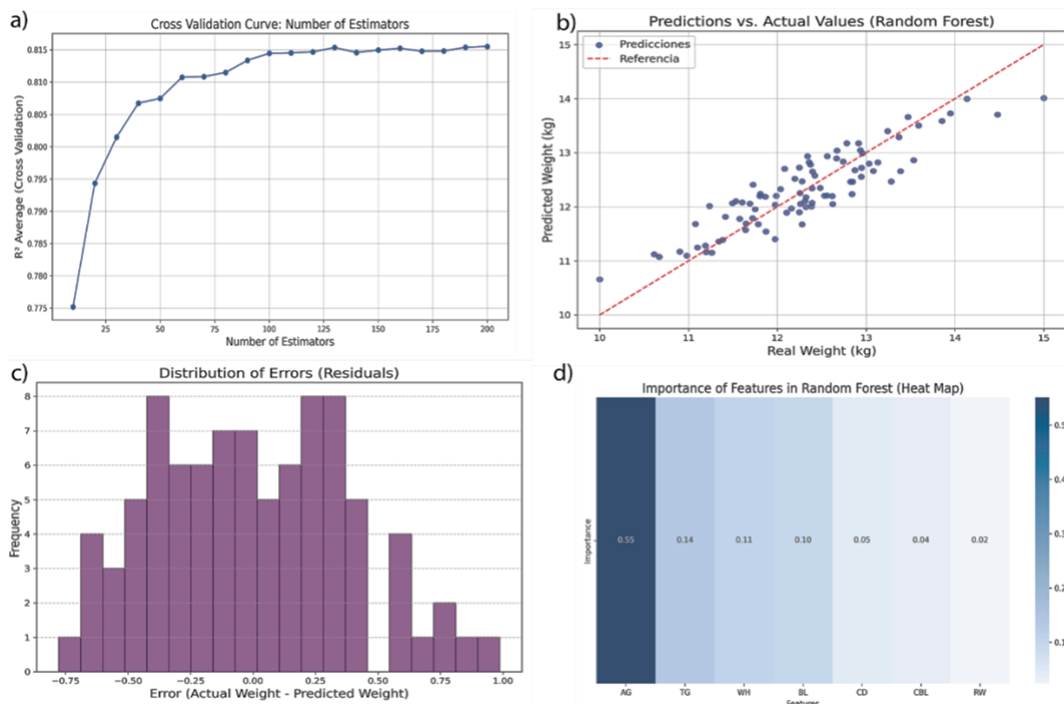
Figure 4. Performance evaluation of the Random Forest Model: Validation curve, predictions, residuals and importance of features.

Nevertheless, this performance is inferior to that achieved with other methods, such as Ridge Regression, which may be attributed to the intrinsic complexity of the interactions between the predictor variables.

XGBoost exhibited outstanding performance, with an MSE of 0.119 and an RMSE of 0.345, accompanied by high R^2 (0.842) and Adjusted R^2 (0.828) values. These metrics position this model as an efficient alternative for prediction tasks where the balance between accuracy and model complexity is crucial. Furthermore, the relative error of 2.19% and CV% of 5.97% indicate adequate robustness for scenarios with moderate variability in the data.

The set of graphs presented in Figure 5 provides a detailed evaluation of the XGBoost model and its performance in predicting live weight in 5-month-old lambs, highlighting various aspects of the model. The Cross-Validation Curve of XGBoost, MSE (Figure 5a), evaluates the model's performance during the cross-validation process. Green points represent the average Mean Squared Error (MSE) values obtained for different hyperparameter configurations, while the red line indicates the final model's MSE. Initial variability between configurations stabilizes as hyperparameters are optimized, suggesting improved performance with specific configurations. The Feature Importance in XGBoost (Figure 5b) shows the relative importance of variables in the model. Abdominal Girth (PA) is identified

as the most influential feature in the model's predictions, followed by Thoracic Girth (PT) and Withers Height (AC), reinforcing the key role of these variables in accurately estimating live weight in lambs. Predictions vs. Actual Values (Figure 5c) compares the model's predictions with actual weight values. Points close to the red reference line indicate a good model fit. While most data points closely follow the reference line, some deviations are visible but remain within acceptable margins, demonstrating reliable model predictions. Distribution of Errors (Residuals) (Figure 5d) presents the distribution of prediction errors (residuals), defined as the difference between actual and predicted weights. Most errors are centered around zero, with an approximately symmetric distribution, indicating no significant biases in the model and consistent predictions. The implementation of the XGBoost model for predicting live weight in lambs showed promising results, excelling in both predictive capacity and the identification of the most influential features. Cross-validation demonstrated the model's ability to optimize effectively by adjusting hyperparameters such as the number of estimators, maximum tree depth, and learning rate, ensuring an appropriate balance between bias and variance (Liang et al., 2025). This is evident in the cross-validation graph, where optimal hyperparameter configurations lead to minimal MSE values, reinforcing the robustness of the tuning process.
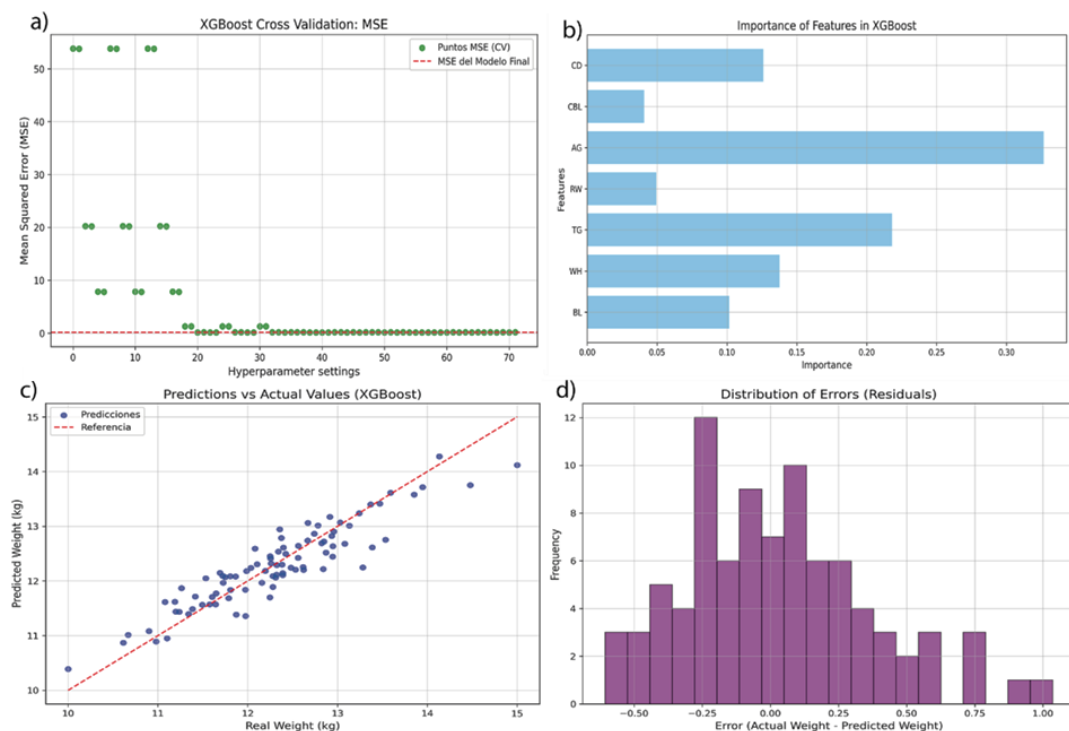
Figure 5. Performance evaluation of the XGBoost: Validation curve, predictions, residuals and importance of features.

Regarding feature importance, Abdominal Girth (PA) emerged as the most significant variable in the model's predictions, followed by Thoracic Girth (PT) and Withers Height (AC). This aligns with previous studies in production animals, where these body measurements are strongly correlated with live weight and are used as reliable indicators in animal management systems (Ergin et al., 2025). The model's ability to identify these key variables underscores its practical utility in the zootechnical field, enabling efficient and accurate assessment of live weight without the need for invasive or expensive tools. The analysis of predictions against actual values revealed significant alignment, with most points closely following the reference line. This indicates that the model not only effectively captures the general trends in weight but also minimizes prediction errors. However, it is important to note that some marginal deviations could be attributed to unmodeled factors such as genetic differences, environmental conditions, or feeding, which affect the animals' physical characteristics (Contreras et al., 2024). On the other hand, the residual distribution showed symmetric dispersion around zero, confirming the absence of systematic biases in the model. This uniform error distribution indicates that the predictions are not influenced by outliers or biased towards a specific range of live weight. This is crucial in practical applications, where the reliability and stability of predictions are determining factors for decision-making in animal production systems. Compared to other machine learning models, XGBoost stood out for its ability to handle datasets with high dimensionality and correlations between variables, contributing to superior performance. Studies such as those conducted by Ahmed et al. (2023) have shown that XGBoost is particularly effective in contexts where the relationship between features and the target variable is nonlinear and complex, as is the case with live weight in animals.

4. Conclusions

The comparative analysis of the Machine Learning models identified ModG and Ridge as the most accurate and stable options, standing out for their low Mean Squared Error (MSE = 0.083) and Root Mean Squared Error (RMSE ≈ 0.287 – 0.288). Additionally, they exhibited the highest coefficients of determination (R^2 =0.89, R_{adj}^2=0.88), indicating excellent predictive capability and data fit. Their low coefficient of variation (CV%) confirms their stability, establishing them as the best choices for applications where precision is paramount, such as the prediction of critical values in production processes and high-demand scientific studies. While XGBoost proved to be a robust alternative with an MSE of 0.119, an RMSE of 0.345, and a relative error of 2.22%.

The results obtained emphasize the importance of selecting predictive models based on the balance between accuracy, interpretability, and stability. While ModG and Ridge are ideal for scenarios where precision is critical, XGBoost emerges as a robust option for problems with high variability. In contrast, Decision Trees, although less accurate, can be useful in applications where the interpretability of decision rules is a key factor. Such an approach optimizes feed allocation, classifies lambs by market weight, and promptly detects growth deviations, thereby improving overall flock profitability.

Further research should validate these findings across multiple flocks that differ in age, sex, breed, and altitude, while simultaneously exploring low-cost 3-D photogrammetry or smartphone imagery to enrich morphometric inputs. Combining inter-pretable ensemble methods with post-hoc explain-ability techniques such as SHAP could also translate predictions into clearer on-farm guidelines, enhan-cing adoption and decision-making throughout diverse sheep-production systems.

Authors contribution

J. Ninahuanca: Writing & formal analysis. E. Garcia-Olarte: review & Conceptualization. I. Unchupaico Payano: Data curation. V. Sarapura: review & editing. K. Zenteno Vera: data collection. C. Quispe Eulogio: Writing & software. E. Ancco Gomez: Investigation. M. Mohamed M. Hadi: mathematical formulas, and mathematical review. C. Miranda-Torpoco: Conceptualization & animal welfare. W. Guerra Condor: data collection & software.

Conflict of interest statement

The authors declare that they have no conflict of interest.

ORCID

J. Ninahuanca https://orcid.org/0000-0002-0137-0631
E. Garcia-Olarte https://orcid.org/0000-0003-1643-288X
I. Unchupaico Payano https://orcid.org/0000-0002-6441-5016
V. Sarapura https://orcid.org/0000-0003-1789-7574
K. Zenteno Vera https://orcid.org/0009-0008-1392-2131
C. Quispe Eulogio https://orcid.org/0000-0002-2316-1646
E. Ancco Gomez https://orcid.org/0000-0002-5119-5202
M. Mohamed M. Hadi https://orcid.org/0000-0003-1940-8383
C. Miranda-Torpoco https://orcid.org/0009-0004-8282-5694
W. Guerra Condor https://orcid.org/0000-0003-1672-1817

References

Ahmed, H., Soliman, H., El-Sappagh, S., Abuhmed, T., & Elmogy, M. (2023). Early Detection of Alzheimer's Disease Based on Laplacian Re-Decomposition and XGBoosting. *Computer Systems Science & Engineering*, *46*(3). https://doi.org/10.32604/csse.2023.036371

Arabameri, A., Tiefenbacher, J. P., Blaschke, T., Pradhan, B., & Tien Bui, D. (2020). Morphometric analysis for soil erosion susceptibility mapping using novel gis-based ensemble model. *Remote Sensing*, *12*(5), 874. https://doi.org/10.3390/rs12050874

Bailey, D.W., Trotter, M.G., Tobin, C., & Thomas, M.G. (2021). Opportunities to apply precision livestock management on rangelands. *Frontiers in Sustainable Food Systems*, *5*, 611915. https://doi.org/10.3389/fsufs.2021.611915

Cam, M. A., Olfaz, M., & Soydan, E. (2010). Body measurements reflect body weights and carcass yields in Karayaka sheep. *Asian Journal of Animal and Veterinary Advances*, *5*(2), 120-127. https://doi.org/10.3923/ajava.2010.120.127

Carhuas, J. N., Capcha, K. B., Garcia-Olarte, E., & Eulogio, C. Q. (2024). Production performance of rejected newborn lambs fed with different concentrations of whey in Perú. *Revista De Ciências Agroveterinárias*, *23*(2), 231–239. https://doi.org/10.5965/223811712322024231

Choque, C. J. B. (2024). Mathematical models of chlorine demand in river waters: a systematic review. *Tecnia*, *34*(1), 26-41. https://doi.org/10.21754/tecnia.v34i1.1635

Contreras, J. P., Cordero, A., Rojas, Y., Carhuas, J., Curasma, J., Mayhua, P., & Salazar, K. (2024). Prediction models for live body weight and body compactness of Criollo sheep in Huancavelica Region, Peru. *The Indian Journal of Animal Sciences*, *94*(7), 637-641. https://doi.org/10.56093/ijans.v94i7.148186

Courtenay, L. A., Yravedra, J., Huguet, R., Aramendi, J., Maté-González, M. Á., González-Aguilera, D., & Arriaza, M. C. (2019). Combining machine learning algorithms and geometric morphometrics: a study of carnivore tooth marks. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *522*, 28-39. https://doi.org/10.1016/j.palaeo.2019.03.007

Dang, C., Choi, T., Lee, S., Lee, S., Alam, M., Park, M., ... & Hoang, D. (2022). Machine learning-based live weight estimation for hanwoo cow. *Sustainability*, *14*(19), 12661. https://doi.org/10.3390/su141912661

Ergin, M., & Koşkan, Ö. (2025). Estimating body weight in Sujiang pigs using artificial neural network, nearest neighbor, and CART algorithms: a comparative study using morphological measurements. *Tropical Animal Health and Production*, *57*(1), 17. https://doi.org/10.1007/s11250-024-04258-7

Frizzarin, M., Gormley, I. C., Berry, D. P., Murphy, T. B., Casa, A., Lynch, A., & McParland, S. (2021). Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. *Journal of Dairy Science*, *104*(7), 7438-7447. https://doi.org/10.3168/jds.2020-19576

García-Medina, A., & Aguayo-Moreno, E. (2024). LSTM–GARCH hybrid model for the prediction of volatility in cryptocurrency portfolios. *Computational Economics*, *63*(4), 1511-1542. https://doi.org/10.1007/s10614-023-10373-8

Gomes, R.A., Monteiro, G.R., Assis, G.J., Busato, K.C., Ladeira, M.M., & Chizzotti, M.L. (2016). Technical note: Estimating body weight and body composition of beef cattle trough digital image analysis. *Journal of Animal Science*, *94*, 5414–5422. https://doi.org/10.2527/jas.2016-0797

Gonçalves, M. A., Castro, M. S. M., Carrara, E. R., Raineri, C., Rennó, L. N., & Schultz, E. B. (2025). Prediction of Weight and Body Condition Score of Dairy Goats Using Random Forest Algorithm and Digital Imaging Data. *Animals*, *15*(10), 1449. https://doi.org/10.3390/ani15101449

Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Briefings in Bioinformatics*, *24*(2), bbad002.

Jarupunphol, P., Buathong, W., Kuptabut, S., & Sudjarid, W. (2025). Assessing decision tree, random forest, and XGBoost models for human capital readiness predictions in low-income areas. *Multidisciplinary Science Journal*, *7*(6), 2025296-2025296. https://doi.org/10.31893/multiscience.2025296

Jurkovich, V., Hejel, P., & Kovács, L. (2024). A review of the effects of stress on dairy cattle behaviour. *Animals*, *14*(14), 2038. https://doi.org/10.3390/ani14142038

Karna, D. K., Mishra, C., Dash, S. K., Acharya, A. P., Panda, S., & Chinnareddyvari, C. S. (2024). Exploring body morphometry and weight prediction in Ganjam goats in India through

principal component analysis. *Tropical Animal Health and Production*, *56*(8), 298. https://doi.org/10.1007/s11250-024-04114-8

Kumar, R., Sharma, D., Dua, A., & Jung, K. H. (2023). A review of different prediction methods for reversible data hiding. *Journal of Information Security and Applications*, *78*, 103572. https://doi.org/10.1016/j.jisa.2023.103572

Kozaklı, Ö., Ceyhan, A., & Noyan, M. (2024). Comparison of machine learning algorithms and multiple linear regression for live weight estimation of Akkaraman lambs. *Tropical Animal Health and Production*, *56*(7), 250. https://doi.org/10.1007/s11250-024-04049-0

Lee, C. S., Cheang, P. Y. S., & Moslehpour, M. (2022). Predictive analytics in business analytics: decision tree. *Advances in Decision Sciences*, *26*(1), 1-29. https://doi.org/10.47654/v26y2022i1p1-29

Liang, Z., Cai, L., Wang, S., & Wang, Q. (2025). K-fold cross-validation based frequentist model averaging for linear models with nonignorable missing responses. *Statistics and Computing*, *35*(1), 18. https://doi.org/10.1007/s11222-024-10554-x

Lipovetsky, S. (2021). Modified ridge and other regularization criteria: A brief review on meaningful regression models. *Model Assisted Statistics and Applications*, *16*(3), 225-227. https://doi.org/10.3233/MAS-210518

Long, K., Guo, D., Deng, L., Shen, H., Zhou, F., & Yang, Y. (2025). Cross-Combination Analyses of Random Forest Feature Selection and Decision Tree Model for Predicting Intraoperative Hypothermia in Total Joint Arthroplasty. *The Journal of Arthroplasty*, *40*(1), 61-69. https://doi.org/10.1016/j.arth.2024.07.007

Mahmud, M. A., Shaba, P., Abdulsalam, W., Yisa, H. Y., Gana, J., Ndagi, S., & Ndagimba, R. (2014). Live body weight estimation using cannon bone length and other body linear measurements in Nigerian breeds of sheep. *Journal of Advanced Veterinary and Animal Research*, *1*(4), 169-176. https://doi.org/10.5455/javar.2014.a29

Martins, B. M., Mendes, A. L., Silva, L. F., Moreira, T. R., Costa, J. H., Rotta, P. P., Chizzotti, M. L., & Marcondes, M. I. (2020). Estimating body weight, body condition score, and type traits in dairy cows using three dimensional cameras and manual body measurements. *Livestock Science*, *236*, 104054. https://doi.org/10.1016/j.livsci.2020.104054

Mokri, M., Safari, M., Kaviani, S., Juneau, D., Cohalan, C., Archambault, L., & Carrier, J. F. (2025). Deep learning-based prediction of later 13N-ammonia myocardial PET image frames from initial frames. *Biomedical Signal Processing and Control*, *100*, 106865.

Ninahuanca Carhuas, J., Cerna, L, A., Unchupaico Payano, I, Garcia-Olarte, E., Mauricio-Ramos, Y., Quispe Eulogio, C., & Hadi Mohamed, Mohamed M. (2025). Counting sheep: human experience vs. Yolo algorithm with drone to determine population. *Veterinary Integrative Sciences*, *23*(2), 1-9. https://doi.org/10.12982/VIS.2025.032

Ozen, H., Ozen, D., Kocakaya, A., & Ozbeyaz, C. (2024). Shrinkage and tree-based regression methods for the prediction of the live weight of Akkaraman sheep using morphological traits. *Tropical Animal Health and Production*, *56*(8), 346. https://doi.org/10.1007/s11250-024-04187-5

Pannier, L., Tarr, G., Pleasants, T., Ball, A., McGilchrist, P., Gardner, G. E., & Pethick, D. W. (2025). The construction of a sheepmeat eating quality prediction model for Australian lamb. *Meat science*, *220*, 109711. https://doi.org/10.1016/j.meatsci.2024.109711

Peña-Avelino, L. Y., Alva-Pérez, J., Ceballos-Olvera, I., Hernández-Contreras, S., & Álvarez-Fuentes, G. (2021). Evaluación de diferentes fórmulas zoométricas para la estimación de peso vivo en cabras criollas de Tamaulipas, México. *Producción Vegetal, 532*. https://doi.org/10.12706/itea.2021.007

Qin, Q., Zhang, C. Y., Liu, Z. C., Wang, Y. C., Kong, D. Q., Zhao, D., ... & Liu, Z. H. (2024). Estimation of the genetic parameters of sheep growth traits based on machine vision acquisition. *Animal*, *18*(7), 101196. https://doi.org/10.1016/j.animal.2024.101196

Salamanca-Carreño, A., Parés-Casanova, P. M., Vélez-Terranova, M., Martínez-Correal, G., & Rangel-Pachón, D. E. (2024). Early Cannon Development in Females of the "Sanmartinero" Creole Bovine Breed. *Animal*, *14*(4), 527. https://doi.org/10.3390/ani14040527

Samuel, A.L. Algunos estudios en aprendizaje automático utilizando el juego de damas. *Revista de investigación y desarrollo de IBM*. 2000, 44206–226.

Shen, Y., Wu, S., Wang, Y., Wang, J., & Yang, Z. (2025). Interpretable model for rockburst intensity prediction based on Shapley values-based Optuna-random forest. *Underground Space*, *21*, 198-214. https://doi.org/10.1016/j.undsp.2024.09.002

Stewart, W. C., Scasta, J. D., Maierle, C., Ates, S., Burke, J. M., & Campbell, B. J. (2025). Vegetation management utilizing sheep grazing within utility-scale solar: Agro-ecological insights and existing knowledge gaps in the United States. *Small Ruminant Research*, *243*, 107439. https://doi.org/10.1016/j.smallrumres.2025.107439

Vlaicu, P. A., Gras, M. A., Untea, A. E., Lefter, N. A., & Rotar, M. C. (2024). Advancing Livestock Technology: Intelligent Systemization for Enhanced Productivity, Welfare, and Sustainability. *AgriEngineering*, *6*(2), 1479-1496. https://doi.org/10.3390/agriengineering6020084

Wang, Z., Shadpour, S., Chan, E., Rotondo, V., Wood, K. M., & Tulpan, D. (2021) Applications of machine learning for livestock body weight prediction from digital images. *Journal of Animal Science, 99*(2), skab022. https://doi.org/10.1093/jas/skab022