R. Bras. Zootec., 55:e20250244, 2026
https://doi.org/10.37496/rbz5520250244

Animal production systems and agribusiness
Full-length research article

Revista Brasileira de Zootecnia

Brazilian Journal of Animal Science
e-ISSN 1806-9290
www.rbz.org.br

Developmental factors and genetic groups evaluated by computational modelling for pregnancy prediction when mating at 14 months of age

Ricardo Zambarda Vaz[1]*, Guyomar Ubyrathan Ramos[1], Rangel Fernandes Pacheco[2], Rafael Espigolan[1], Cristian Faturi[1], Mariana Assis Borges[1], Bibiana Teixeira Rodrigues[3], João Restle[4]

[1] Universidade Federal de Santa Maria, Departamento de Zootecnia e Ciências Biológicas, Palmeira das Missões, RS, Brasil.
[2] Instituto Federal Farroupilha, Frederico Westphalen, RS, Brasil.
[3] Universidade Federal de Santa Maria, Departamento de Zootecnia, Santa Maria, RS, Brasil.
[4] Universidade Federal de Goiás, Departamento de Zootecnia, Goiânia, GO, Brasil.

***Corresponding author:**
rzvaz@terra.com.br

Received: November 11, 2025
Accepted: March 5, 2026

How to cite: Vaz, R. Z.; Ramos, G. U.; Pacheco, R. F.; Espigolan, R.; Faturi, C.; Borges, M. A.; Rodrigues, B. T. and Restle, J. 2026. Developmental factors and genetic groups evaluated by computational modelling for pregnancy prediction when mating at 14 months of age. Revista Brasileira de Zootecnia 55:e20250244.
https://doi.org/10.37496/rbz5520250244

Editors:
Mateus Pies Gionbelli
Luan Sousa dos Santos

ABSTRACT - The aim of this study was to predict pregnancy in beef heifers, identifying the most impactful variables using computational modelling. Developmental data of 98 heifers up to 480 days (16 months) of age (Charolais, Nellore, and their crossbreeds) were used. At the beginning and end of the mating season initial body weight (IBW), initial body condition score (IBS), final body weight (FBW), and final body condition score (FBS) were assessed. The age at the start of the mating season (AGE), and the genetic composition of the heifers were also evaluated. Principal Component Analysis (PCA) was carried out, where the first two components (PC1 and PC2) explained 60.7% of the total variance in the data. The PCA biplot revealed a separating trend, with PC1 (44.6%) associated with weight and age, while PC2 (16.1%) was related to the body condition score and genetics. A random forest (RF) model was then developed and trained to classify pregnancy status, achieving an area under the ROC curve (AUC) of 0.80, showing the model to have discriminatory power. The variables with the greatest predictive importance were IBW (30.1%), EBW (19.9%), IBS (14.4%), and EBS (12.4%). Nellore genetics is the main factor for pregnancy, followed by variables related to development, nutritional status and age. The combination of PCA and RF proved to be a reliable approach for identifying critical factors and predicting reproductive success in beef heifers.

Keywords: beef heifers, machine learning, pregnancy, principal component analysis

1. Introduction

In the context of geopolitical instability, climate change and population growth, the global supply of animal protein has become a strategic pillar for food security and the Brazilian economy, internationally recognized for its production of high-quality beef (Farias et al., 2023). In southern Brazil, forage seasonality is common, in addition to increasing pressure for agricultural land to be allocated to grain production (Jaurena et al., 2021).

To remain competitive and environmentally sustainable, beef production systems must increase productivity without expanding the production area. Breeding herds aim to increase the pregnancy rate by reducing the age at first calving. These indicators impact calf production (Pravia et al., 2014) and improve its carbon footprint (Cullen et al., 2016). Optimizing these parameters maximise calf production, reduces the maintenance costs of unproductive animals, accelerates the return on investment in genetics, and increases protein production per unit of carbon emitted (Cullen et al., 2016).

Pregnancy in heifers is a complex process influenced by both genetic and environmental factors (Restle et al., 1999; Vaz et al., 2012; Kgari et al., 2020). Choosing the appropriate genetic group and individual for a given production environment is crucial for reproductive efficiency. However, no single genetic group is fully adapted to all production environments. Therefore, evaluating genetic groups and their crossbreeding combinations to take advantage of heterosis is a viable strategy for improving the efficiency of production systems (Restle et al., 1999). In addition, an environment suited to the animal’s requirements promotes adequate development through proper management and nutrition, factors directly associated with improved reproductive performance (Eloy et al., 2022). Since pregnancy is a multifactorial and complex trait, new approaches for identifying and predicting key variables are important for improving efficiency and intensifying production systems.

Research on more efficient tools for predicting production variables, such as pregnancy, has gained increasing attention (Brand et al., 2021; Rabaglino et al., 2023). Machine learning has emerged as a promising technology due to its ability to analyze large and complex datasets and identify patterns that may not be detected through conventional statistical approaches (Schleder and Fazzio, 2021). The application of machine learning algorithms enables the development of predictive models that can support strategic decision-making, ranging from the selection of animals with greater production potential to the optimization of management practices (Zhang et al., 2021). In this context, predicting pregnancy in beef heifers using machine learning allows the early identification of conception likelihood, making it possible to implement more targeted and efficient management interventions (Marques et al., 2024). Among the various machine learning algorithms, decision trees, particularly random forest, stand out because of their robustness and ability to process complex datasets by averaging multiple decision trees while minimizing overfitting (Breiman, 2001).

Random forest is an ensemble learning method that constructs multiple decision trees during the training process and predicts the final outcome based on the majority vote for classification tasks or the average prediction for regression tasks (Breiman, 2001). The ability of random forest to identify the relative importance of predictor variables makes it a valuable tool for identifying the main factors influencing pregnancy in beef heifers. Therefore, the aim of this study was to predict pregnancy in beef heifers by identifying and quantifying the effects of phenotypic variables and genetic groups using machine learning based on the random forest algorithm.

2. Material and methods

All animal handling and procedures were approved by the Ethics Committee for the Use of Animals (CEUA) of the Universidade Federal de Santa Maria under process number 2388280122.

2.1. Location

The experiment was conducted at the Animal Science Department of the Universidade Federal de Santa Maria, located in the Central Depression of the state of Rio Grande do Sul (29°43' S, 53°42' W, altitude 95 m). According to the Köppen classification, the climate is subtropical (Alvares et al., 2013).

2.2. Animals

A total of 98 heifers were used, with an average initial age of 420 days (fourteen months) and an average initial weight of 255 kg, from the Charolais (CH), Nellore (NE), 1/2 CH 1/2 NE, 1/2 NE 1/2 CH, 3/4 CH 1/4 NE, 3/4 NE 1/4 CH, 5/8 CH 3/8 NE, and 5/8 NE 3/8 CH genetic groups. The sires of the purebred Charolais and Nellore heifers were the same as those of the crossbred heifers.

2.3. Nutritional management

Prior to the experiment, the heifers were kept on natural pasture with their mothers until weaned at an average age of 210 days (seven months). From 210 days (seven) to 420 days (fourteen months), the calves remained on cultivated winter pasture consisting of Triticale (*Tritico secale*) and Italian ryegrass (*Lolium multiflorum*) at an average stocking rate of 5.7 heifers/ha or 1,200 kg live weight/ha.

During the mating season (420 days to 480 days), the heifers were kept on natural pasture (average stocking rate of 3.0 heifers/ha or 810 kg body weight/ha), receiving concentrated supplement at 0.42% of body weight. The paddock of natural pasture had been earlier fertilised over the winter period, with the introduction of oats via direct sowing. Once the oats were harvested, the remaining pasture was mowed and deferred for 45 days prior to the mating season.

The heifers gradually adapted to the management and supplement over a period of 19 days. The supplement was a commercial pelleted concentrate consisting of soybean hulls, wheat bran and molasses, with a total of 12.5% crude protein (CP) and 67.3% total digestible nutrients (TDN).

Supplementation lasted 90 days and was always offered at 07:00 in troughs accessed from both sides, giving a total length of 1.0 linear metre per animal. After weaning, the batch of heifers was maintained as a single group in the same area of pasture and was always provided with good-quality water and suitable minerals for the category.

2.4. Data collection

The animals were weighed at weaning, and at the beginning and end of the reproductive period after minimum fasting period of eight hours. In addition, body weight was recorded every 21 days to monitor weight gain and adjust the stocking rate, and during the reproductive period to adjust their supplement intake. The average daily gain (ADG) was determined as the difference between the initial and final weights of the heifers divided by the number of days in the period under evaluation. Subjective body condition score assessments were also performed at the time of weighing according to Lalman and Stein (2024) were also carried out when weighing. These assessments were made visually assigning scores from 1 to 5, in which 1 = very lean, 2 = lean, 3 = average, 4 = fat, and 5 = very fat.

2.5. Health and reproductive management

The mating season lasted 90 days, and artificial insemination was used as the reproductive method without the use of natural mating. Estrus detection was performed twice daily, in the morning and afternoon. Heifers detected in estrus during the morning were inseminated in the afternoon, whereas those detected in estrus during the afternoon were inseminated the following morning. Semen from a single Red Aberdeen Angus bull was purchased in a single batch, and all insemination doses originated from the same collection lot. Pregnancy diagnosis was performed by ultrasonography 30 days after the end of the mating season. At the end of the reproductive period, 52 heifers were diagnosed as pregnant (53.06%), whereas 46 heifers were classified as non-pregnant (46.94%).

Health management consisted of vaccinations that were administered in accordance with the schedule of the Rio Grande do Sul Department of Agriculture. Endo- and ectoparasites were strategically controlled using specific products whenever necessary, based on the degree of infestation.

2.6. Statistical analysis

The data were submitted to principal component analysis (PCA), which was carried out in the Python programming environment using the scikit-learn package, to investigate the relationships between performance characteristics and to evaluate the grouping of heifers based on the pregnancy response variable (0 = non-pregnant; 1 = pregnant).

The variables included in the PCA were initial body weight (IBW), final body weight (FBW), initial body condition score (IBS), final body condition score (FBS), age (AGE), and genetic dominance, coded as indicator variables for Zebu and synthetic (crossbred) groups, with taurine animals used as the reference category. The coding procedure was performed using dummy variables through the get_dummies function of the pandas library, with drop_first = True to ensure reference category coding.

The random forest (RF) procedure, implemented with the scikit-learn package RandomForestClassifier, was used to identify the most relevant variables associated with pregnancy. Although PCA was used to visualise the data structure, variable selection in the RF model was carried out independently based on the mean reduction in the Gini coefficient. The RF classifier operates by generating a set of decision trees trained on random subsets of samples (bootstrap = false) and random subsets of variables at each split, improving the robustness of the model and reducing overfitting.

The explanatory variables submitted to RF were categorised into four groups: (1) body weight (IBW and FBW), (2) body condition score (IBS and FBS), (3) genetic dominance (Breed Zebu and Breed Synt) with Taurine as the reference value, and (4) heifer age in days (AGE). The occurrence of pregnancy was used as the response variable. The dataset was randomly divided into two subsets, with 70% of the observations used for training and the remaining 30% for testing, using random_state=42 to ensure reproducibility. This internal validation approach (holdout) was adopted as an alternative to external validation. The random forest model was trained with 1,000 trees (n_estimators=1000) and a maximum depth of four levels (max_depth=4) to prioritise more-robust and global splits in line with field knowledge, aiming for greater interpretability.

The statistical procedures employed for Random Forest modelling followed recommended practices for supervised learning in biological datasets, as described by Breiman (2001) and Hastie et al. (2009). All analytical procedures were implemented in Python (version 3.11). The following library versions were used to ensure reproducibility: pandas (v2.2.3), numpy (v1.24.0), scikit-learn (v1.4.2), matplotlib (v3.7.5), and seaborn (v0.11.2). The full set of hyperparameters adopted in the Random Forest classifier was explicitly defined as follows: n_estimators = 1000, max_depth = 4, criterion = “gini”, bootstrap = False, max_features = “auto” (default), min_samples_split = 2, min_samples_leaf = 1, class_weight = None, and random_state = 42. The same random seed (random_state = 42) was applied to both the training/testing split and model construction to ensure deterministic and fully reproducible outputs.

Preprocessing steps consisted exclusively of dummy-encoding the categorical variable “Breed” using the pandas.get_dummies(drop_first=True) function, with Taurine as the reference category. No scaling procedures were applied, as all predictors were expressed in compatible numerical units. The dataset contained no missing values, which was verified before model training. All predictors entered the models exactly as recorded in the original field dataset, without imputation or modification.

To evaluate model performance, the confusion matrix, accuracy, precision, recall, specificity, receiver operating characteristic (ROC) curve, and the area under the ROC curve (AUC) were all calculated. Although the pregnancy outcomes in the dataset were practically balanced (53% pregnant vs. 47% not pregnant), we calculated the balanced accuracy (0.768) and the area under the precision-recall curve (AUC-PR = 0.700). No class weighting strategy was necessary, as both classes contributed similarly to the model training process. The importance of each variable was determined based on its average contribution to reducing Gini impurity at the tree nodes. In addition, a representative decision tree was extracted from the RF model (composed of 1,000 trees). This tree was exported for visualisation using the export_graphviz function. In a complementary step, a full decision tree (DecisionTreeClassifier) was also trained, but with no depth restriction (max_depth = None), using the same training data. The full decision tree was used to evaluate the maximum splitting potential for each variable, with the aim of providing a complete overview of the modelling process.

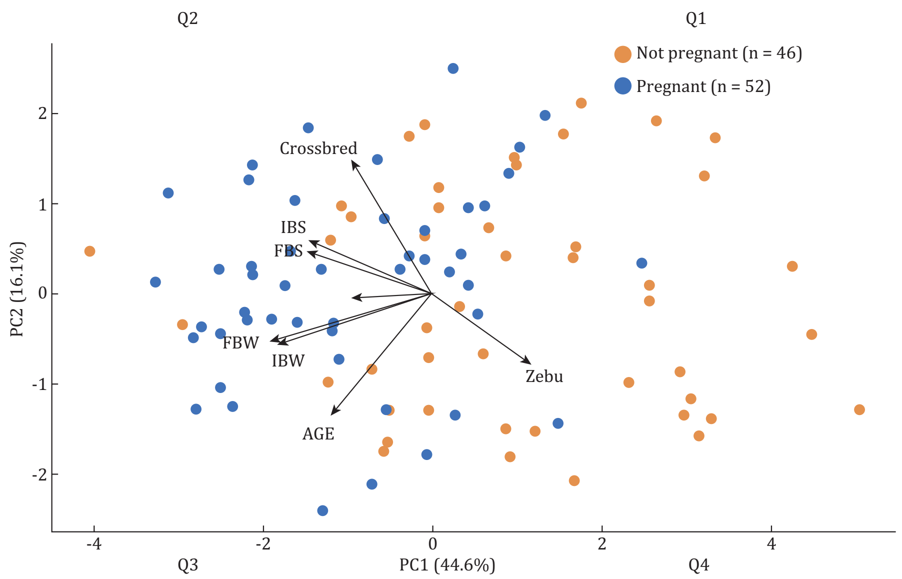
3. Results

The PCA was used to reduce data dimensionality, simplifying their interpretation by transforming correlated variables into uncorrelated principal components, preserving the main sources of variation. The resulting generated orthogonal matrix allowed underlying patterns to be identified and facilitated visualization of the data structure. After running the PCA, the first two principal components were plotted on a biplot, with colors indicating pregnancy status.

The biplot was constructed based on the developmental and genetic variables of the heifers (Figure 1), with the colour of the points representing their gestational status: pregnant (blue) and non-pregnant (orange); the multivariate variation of the data allowed for natural separation of the groups of pregnant and non-pregnant heifers. The first two principal components (PC1 and PC2) explained a considerable proportion of the total variance, with PC1 = 44.6% and PC2 = 16.1%, giving a total of 60.7%.

The PC1 axis was composed of body weight at the beginning and end of the study, as well as age, indicating a positive correlation among these variables. PC2 was composed of body condition score at the beginning and end of the mating season and the genetic variable representing crossbred animals. In contrast, pregnancy was oppositely associated with the genetic factor of Zebu origin. The taurine genetic component was used as the reference category and, therefore, does not appear explicitly in the graph.

Pregnant heifers were primarily concentrated on the left side of the biplot, in the same direction as the vectors representing body weight, body condition score, crossbred animals, and age at the beginning of the breeding season, suggesting that pregnancy was associated with greater age, improved body development, and hybrid vigor. In contrast, non-pregnant heifers were concentrated in the upper and lower right quadrants of the biplot, in the same direction as the vector representing Zebu genetic origin, suggesting that this factor may be associated with less favorable conditions for pregnancy in heifers aged 420 to 480 days (14-16 months).



Crossbred - genetic origin of crossbred animals; IBS - body condition score at the start of the breeding season; FBS - body condition score at the end of the breeding season; FBW - body weight at the end of the breeding season; IBW - body weight at the start of the breeding season; AGE - age at the start of the breeding season; Zebu - genetic origin of zebu animals.

Figure 1 - PCA biplot of beef heifers by pregnancy status.

The receiver operating characteristic (ROC) curve was used to evaluate the predictive performance of the random forest model for classifying pregnancy status in beef heifers. The ROC curve was constructed by plotting the true positive rate against the false positive rate across different classification thresholds. The model achieved an area under the curve (AUC) of 0.80, indicating good discriminatory ability (Figure 2). This AUC value indicates that, when randomly selecting a pair of heifers, the model has an 80% probability of correctly assigning a higher pregnancy likelihood to a pregnant heifer than to a non-pregnant heifer. The shape of the ROC curve and its distance from the diagonal reference line further demonstrate the ability of the model to discriminate between pregnant and non-pregnant heifers. Thus, the curve represents the relationship between sensitivity and specificity across a range of classification thresholds.

The most important variable in the random forest (RF) was IBW, followed by FBW, IBS and FBS (Table 1). Immediately following the variables related to development and, consequently, nutrition, the predominance of the zebu and taurine genetic origins stands out, as well as the age of the heifers at the start of mating.

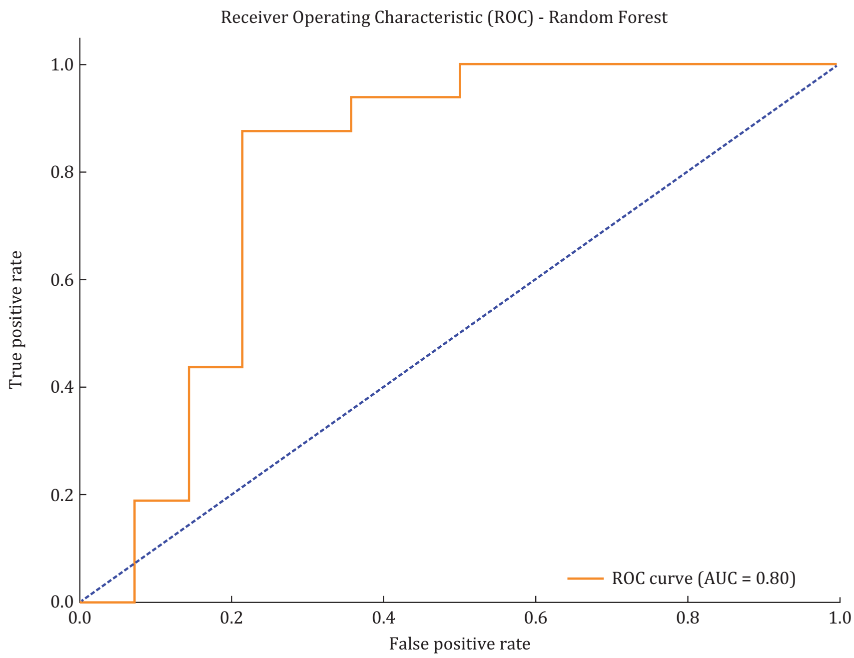


Figure 2 - Receiver operating characteristic (ROC) curve of the random forest model developed to classify pregnancy in beef heifers.

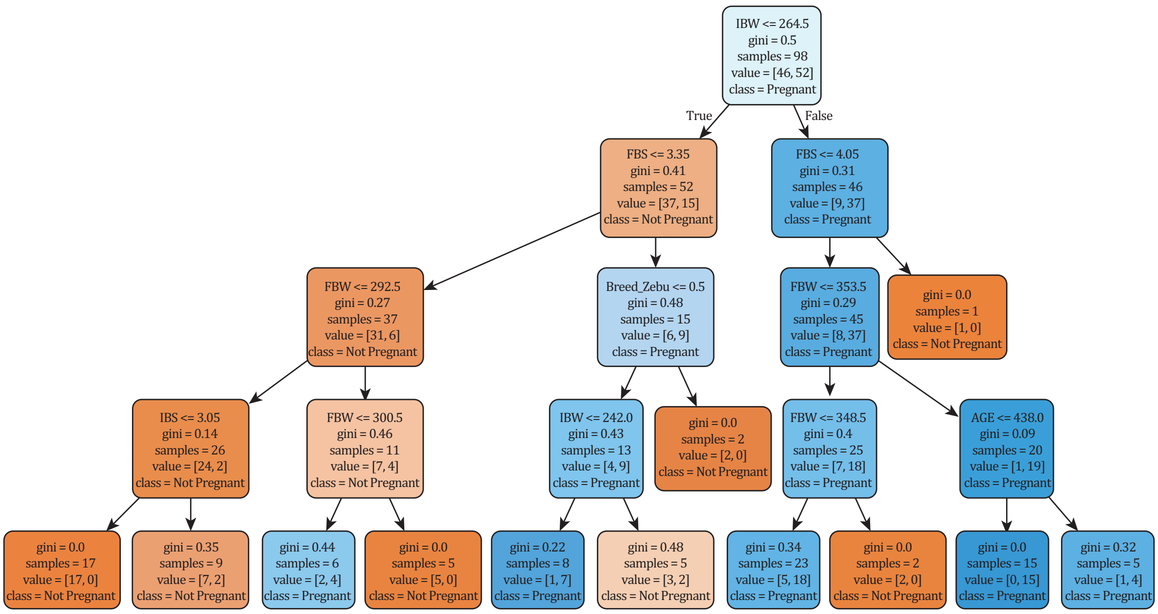
Table 1 - Importance of the variables evaluated using random forest

Variable	Importance (%)
Weight at the start of the mating season (IBW)	30.1
Weight at the end of the mating season (FBW)	19.9
Body score at the start of the mating season (IBS)	14.4
Body score at the end of the mating season (FBS)	12.4
Cross with predominantly zebu genetic origin	11.0
Age at mating (AGE)	9.6
Cross with predominantly taurine genetic origin	2.7

The most representative decision tree for predicting pregnancy in beef heifers, based on the variables body weight, age and predominance of genetic origin (taurine or Zebu) considered body weight at the beginning of the breeding season as the main reference for the model (Figure 3). The tree with a depth of four levels was extracted from a composite of 1,000 trees for the purposes of interpretation, all created by the random forest model.

The weight at the beginning of the reproductive period was the main segmentation variable at the root node, demonstrating its importance as the primary splitting criterion in the model. Heifers were initially separated into groups with body weight ≤264 kg or >264 kg and subsequently directed along different classification paths. In the first branch, heifers with IBW >264 kg and FBS >4.05 points were classified as "non-pregnant". The conclusive result observed in this branch, characterized by a pure node (Gini = 0.0), suggests a negative effect of excessive body condition score on fertility within the evaluated dataset. In contrast, pregnancy classification was associated with heifers presenting IBW >264 kg, FBS <4.05 points, and FBW >353.5 kg, possibly combined with AGE <438 days.

For the second path, heifers with an IBW ≤ 264 kg depend on FBS being ≤ 3.35 points, i.e. if FBS is ≤ 3.35, the heifer is considered not pregnant. On the other hand, heifers with an IBW ≤ 264 kg but FBS greater than 3.35 points must have less than 50% Nellore genes in their genetic makeup to be pregnant; with more than 50%, they are considered not pregnant as long as they have an IBW ≤ 264 kg.



IBW - weight at the start of the breeding season; FBS - body condition score at the end of the mating season; FBW - weight at the end of the mating season; AGE - age at the start of the mating season; Breed_Zebu - percentage zebu (0 = taurine, 1 = zebu); FBS - body condition score at the end of the mating season; IBS - body condition score at the start of the mating season; Gini - Gini index (measure of impurity in split data); Samples - percentage of samples or data analysed from the total database at each node of the tree; Value - percentage of samples of each class (pregnant or not pregnant) at each node; Class - final classification of the heifer.

Figure 3 - Representative decision tree extracted from a random forest model trained with a maximum depth restricted to four levels to predict the likelihood of pregnancy in beef heifers.

4. Discussion

The explanation of 60.7% of the total variance (PC1 = 44.6% and PC2 = 16.1%) afforded by the first two components of the principal component analysis (PCA) is considered satisfactory given the multifactorial nature of pregnancy in beef heifers. Studies using PCA in animal science have reported different proportions of explained variance for the first two components, reflecting the complexity of the zootechnical traits evaluated. To understand the components that explained egg production, the value of the two principal groups explained 68% of the results (Paiva et al., 2010), confirming the complexity of correlating production characteristics. Investigating the components that influence the variability of part-Nellore carcasses, PCA was able to explain 51.91% of the variation with the first two components, underlining the impact of the environment on animal production (Silva et al., 2015). Reinforcing the complexity and variability of production-related traits, Vaz et al. (2024), evaluating characteristics that increase the risk of bruising in cattle carcasses, found only 36.67% of the occurrences explained by the first two principal components.

The associations observed in PC1 (initial and final body weight and heifer age during the mating season) and PC2 (initial and final body condition scores during the mating season and heterozygosity resulting from different crossbreeding generations) indicate that more-developed crossbred heifers with better body condition scores and greater physiological maturity were more likely to become pregnant. These results suggest that adequate nutritional management during the rearing period is essential to ensure that heifers begin the mating season with greater body weight and improved body condition, factors directly associated with higher pregnancy rates.

Although PCA does not have predictive capability, it provided a robust exploratory basis for the subsequent development of supervised models, such as random forest (RF). The AUC value of 0.80 achieved by the random forest classifier demonstrates a strong ability to discriminate between pregnant and non-pregnant heifers. Random forest provides robust analytical performance through the combination of multiple decision trees, enabling the capture of non-linear interactions while reducing overfitting, often outperforming linear approaches such as logistic regression (Breiman, 2001; Hastie et al., 2009). The region of the ROC curve closest to the upper left corner (Figure 2) indicates that the model maintained high sensitivity even at moderate false-positive rates, which is desirable in reproductive management systems where maximizing pregnancy rates is a priority. These results support the application of random forest as a decision-support tool in breeding programs, particularly for identifying heifers with greater probability of conception based on phenotypic and genetic characteristics.

In terms of the predictor variables, random forest highlighted the IBW, FBS, FBW and AGE of the heifers as the factors with the most influence on pregnancy. In a second instance, the percentage of zebu blood plays a part in the reproductive success of beef heifers. These results confirm that when heifers start the mating season with more weight and a better body condition score, they tend to have higher a rate of fertility (Vaz et al., 2012; Landarin et al., 2016; Silva et al., 2018; Dickinson et al., 2019).

Studies that stratified herds of heifers mated between 420-480 days (14-16 months) of age into weight ranges (Vaz end Lobato, 2010; Vaz et al., 2012) found higher pregnancy rates in the heavier weight classes. According to the NRC (2016), a taurine heifer needs to reach 60% of its adult weight to be suitable for breeding. For breeds of European origin, the weight of the heifers is close to those predicted as the cutoff point by random forest. For zebu breeds, which mature later, the recommended cutoff for breeding is around 65% of the adult weight. This was confirmed in the present study since the pregnancy rate in females with a higher degree of zebu blood was lower. In addition, it should be noted that heifers of both genetic origins were in a subtropical climate, where breeds of European origin naturally tend to produce more than do breeds of zebu origin (Restle et al., 2001). Gregianini et al. (2021) demonstrated the importance of a minimum weight for Nellore heifers mated at 330-420 days (11-14 months), who achieved a reproductive index of 42% with an average of 280.17 kg, a difference of 12.43 kg compared to non-pregnant heifers.

The inclusion of IBW and FBW as important predictors suggests that not only the initial weight, but also the capacity for weight gain during the mating season contribute to better rates of pregnancy. However, IBW alone does not ensure reproductive success if heifers in the event of weight loss during the mating season, as demonstrated by the importance of weight and body condition score in the PCA (Vaz et al., 2012).

However, body weight alone is not the only determinant of pregnancy at 420 days (14 months) of age, as the physiological maturity of the heifers also plays an important role. Older heifers generally exhibit higher pregnancy rates and greater likelihood of conception (Rocha and Lobato, 2002; Vaz and Lobato, 2010; Landarin et al., 2016; Dickinson et al., 2019). Pregnancy at 420-480 days (14–16 months) of age is therefore the result of an interaction between age and weight, with weight being closely associated with the body condition score.

In many cases, age is considered one of the main factors influencing reproductive tract maturity and has been described as a predictor of pubertal status in female cattle (Holm et al., 2015). It should also be noted that heifers younger than 368 days at first mating exhibited reduced reproductive success (Dickinson et al., 2019). This finding is consistent with the present study, in which the model classified females above this age threshold as pregnant.

Zebu genetic background was associated with lower pregnancy rates in heifers bred at 420-480 days (14-16 months) of age, indicating that breed composition plays an important role in the reproductive performance of beef heifers. The lower reproductive performance observed in purebred Nellore heifers, as well as in animals with a greater proportion of Zebu genes, may be associated with poorer adaptation to the environmental conditions in which the study was conducted. The characteristics of the production system and regional climatic conditions, together with the genetic background of Nellore heifers, may not provide ideal conditions for adequate pre-mating development. In particular, the heifers were exposed to periods of adverse environmental conditions, including low temperatures, to which this genetic group may be less adapted.

The early identification of females with lower probability of pregnancy allows the implementation of targeted management interventions, such as nutritional supplementation or the selective culling of unproductive animals, thereby optimizing return on investment. This approach also enables the allocation of more expensive genetic resources to heifers with greater reproductive potential. In addition, achieving pregnancy at an early age contributes to faster amortization of cow production costs, as a cow is estimated to require the production of six calves to offset these costs (Boyer et al., 2020), thereby maximizing productivity and economic return.

The ability of the model to distinguish between pregnant and non-pregnant animals, with an AUC of 0.80, highlights the potential of machine learning for optimizing reproductive management programs in cattle production systems. Similarly, an AUC of 0.75 was reported when applying machine learning algorithms to predict insemination success in Holstein cows (Shahinfar et al., 2014).

For future studies, the use of larger and more robust datasets, as well as the inclusion of temperament assessments in heifers, is recommended to improve predictive accuracy. Furthermore, the performance of random forest models may be enhanced by incorporating additional sources of information, such as physiological indicators and sensor-derived data (Rabaglino et al., 2023; Marques et al., 2024).

Although the results obtained were satisfactory, some limitations should be acknowledged due to their implications for the generalization of the findings. This study was conducted under subtropical environmental conditions, using animals from a single herd and a limited sample size. Therefore, additional studies are required to validate the proposed analyses, as reproductive performance is strongly influenced by environmental conditions. Further improvements in the analyses should include the evaluation of additional breeds and different management systems, as well as the incorporation of new reproductive-related variables, such as the hormonal status of the heifers.

5. Conclusions

In the specific context of the present dataset and management conditions, phenotypic and nutritional factors proved to be more decisive for pregnancy than the differences associated with the genetic group. The decision tree model efficiently predicted pregnancy in beef heifers, demonstrating its potential as an effective decision-support tool, particularly for reproductive and mating management systems.

Data availability

Please contact author for data requests.

Author contributions

Conceptualization: Vaz, R. Z. and Ramos, G. U. **Data curation:** Ramos, G. U.; Borges, M. A. and Rodrigues, B. T. **Formal analysis:** Ramos, G. U.; Pacheco, R. F.; Espigolan, R.; Borges, M. A. and Restle, J. **Funding acquisition:** Vaz, R. Z. and Restle, J. **Investigation:** Borges, M. A. and Rodrigues, B. T. **Methodology:** Vaz, R. Z.; Ramos, G. U.; Pacheco, R. F.; Faturi, C. and Restle, J. **Supervision:** Vaz, R. Z. **Writing – original draft:** Vaz, R. Z.; Ramos, G. U. and Restle, J. **Writing – review & editing:** Vaz, R. Z.; Ramos, G. U. and Restle, J.

Conflict of interest

The authors declare no conflict of interest.

Financial support

The authors thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - grant 310987/2020-2 and 308963/2021-0.

Declaration of generative AI in scientific writing

Generative artificial intelligence tools (ChatGPT Plus, Data Analyst GPT) were used exclusively for auxiliary tasks such as script formatting, command verification, and figure preparation. No part of the analytical design, statistical modelling, or interpretation of the results involved AI-based decision-making.

References

Alvares, C. A.; Stape, J. L.; Sentelhas, P. C.; Moraes; G. J. L. and Sparovek, G. 2013. Köppen's climate classification map for Brazil. Meteorologische Zeitschrift 22:721-728.

Boyer, C. N.; Griffith, A. P. and Delong, K. L. 2020. Reproductive failure and long-term profitability of spring-and fall-calving beef cows. Journal of Agricultural and Resource Economics 45:78-91.

Brand, W.; Wells, A. T.; Smith, S. L.; Denholm, S. J.; Wall, E. and Coffey, M. P. 2021. Predicting pregnancy status from mid-infrared spectroscopy in dairy cow milk using deep learning. Journal of Dairy Science 104:4980-4990. https://doi.org/10.3168/jds.2020-18367

Breiman, L. 2001. Random forests. Machine Learning 45:5-32. https://doi.org/10.1023/A:1010933404324

Cullen, B. R.; Eckard, R. J.; Timms, M. and Phelps, D. G. 2016. The effect of earlier mating and improving fertility on greenhouse gas emissions intensity of beef production in northern Australian herds. The Rangeland Journal 38:283-290. https://doi.org/10.1071/RJ15063

Dickinson, S. E.; Elmore, M. F.; Kriese-Anderson, L.; Elmore, J. B.; Walker, B. N.; Dyce, P. W.; Rodning, S. P. and Biase, F. H. 2019. Evaluation of age, weaning weight, body condition score, and reproductive tract score in pre-selected beef heifers relative to reproductive potential. Journal of Animal Science Biotechnology 10:18. https://doi.org/10.1186/s40104-019-0329-6

Eloy, L. R.; Bremm, C.; Lobato, J. F. P.; Pötter, L. and Laca, E. A. 2022. Direct and indirect nutritional factors that determine reproductive performance of heifers and primiparous cows. Plos One 17:e0275426. https://doi.org/10.1371/journal.pone.0275426

Farias, P. P.; Ferreira, O. G. L.; Kröning, A. B.; Azevedo, G. M.; Costa, P. T.; Fluck, A. C.; Vaz, R. Z. and Restle, J. 2023. Consumers' perception of beef from cattle raised in native grasslands in the Pampa Biome. Observatório de La Economía Latinoamericana 21:1848-1867. https://doi.org/10.55905/oelv21n4-007

Gregianini, H. A. G.; Carneiro Junior, J. M.; Pinto Neto, A.; Costa Filho, L. C. C.; Gregianini, J. T. F.; Pinheiro, A. K. and Trenkel, C. K. G. 2021. Sexual precocity of Nellore heifers in a herd under selection in the State of Acre, Brazil. Research, Society and Development 10:e16310413945. https://doi.org/10.33448/rsd-v10i4.13945

Hastie, T.; Tibshirani, R. and Friedman, J. 2009. The elements of statistical learning. Springer Series in Statistics. Springer, New York, NY. Available at: <http://link.springer.com/10.1007/978-0-387-84858-7>. Accessed on: July 01, 2025.

Holm, D. E.; Nielen, M.; Jorritsma, R.; Irons, P. C. and Thompson, P. N. 2015. Evaluation of pre-breeding reproductive tract scoring as a predictor of long term reproductive performance in beef heifers. Preventive Veteterinary Medicinal 118:56-63. https://doi.org/10.1016/j.prevetmed.2014.10.016

Jaurena, M.; Durante, M.; Devincenzi, T.; Savian, J. V.; Bendersky, D.; Moojen, F. G.; Pereira, M.; Soca, P.; Quadros, F. L. F.; Pizzio, R.; Nabinger, C.; Carvalho, P. C. F. and Lattanzi, F. A. 2021. Native grasslands at the core: A new paradigm of intensification for the Campos of Southern South America to increase economic and environmental sustainability. Frontiers in Sustainable Food Systems 5:547834. https://doi.org/10.3389/fsufs.2021.547834

Kgari, R. D.; Muller, C. J. C.; Dzama, K. and Makgahlela, M. L. 2020. Evaluation of female fertility in dairy cattle enterprises – A review. South African Journal of Animal Science 50:819-829. https://doi.org/10.4314/sajas.v50i6.8

Lalman, D. and Stein, D. 2024. Body condition scoring of cows. Oklahoma Cooperative Extension Service. 4p. Available at: <https://extension.okstate.edu/fact-sheets/print-publications/afs/body-condition-scoring-of-cows-afs-3283-a.pdf>. Accessed on: July 01, 2025.

Landarin, C. M.; Lobato, J. F. P.; Tarouco, A. K.; Tarouco, J. U.; Eloy, L. R.; Pötter, L.; Rosa, A. A. G. 2016. Growth and reproductive performance of 14- to 15-month-old Hereford heifers. Revista Brasileira de Zootecnia 45:667-676. https://doi.org/10.1590/S1806-92902016001100005

Marques, T. C.; Marques, L. R.; Fernandes, P. B.; Lima, F. S.; Paim, T. P. and Leão, K. M. 2024. Machine learning to predict pregnancy in dairy cows: An approach integrating automated activity monitoring and on-farm data. Animals 14:1567. https://doi.org/10.3390/ani14111567

NRC - National Research Council. 2016. Nutrient requirements of beef cattle. National Academy Press, Washington, DC.

Paiva, A. L. C.; Teixeira, R. B.; Yamaki, M.; Menezes, G. R. O.; Leite, C. D. S. and Torres, R. A. 2010. Análise de componentes principais em características de produção de aves de postura. Revista Brasileira de Zootecnia 39:285-288. https://doi.org/10.1590/S1516-35982010000200009

Pravia, M. I.; Ravagnolo, O.; Urioste, J. I. and Garrick, D. J. 2014. Identification of breeding objectives using a bioeconomic model for a beef cattle production system in Uruguay. Livestock Science 160:21-28. https://doi.org/10.1016/j.livsci.2013.12.006

Rabaglino, M. B.; Sánchez, J. M.; McDonald, M.; O'Callaghan, E. and Lonergan, P. 2023. Maternal blood transcriptome as a sensor of fetal organ maturation at the end of organogenesis in cattle. Biology of Reproduction 109:749-758. https://doi.org/10.1093/biolre/ioad103

Restle, J.; Polli, V. A. and Senna, D. B. 1999. Efeito de grupo genético e heterose sobre a idade e peso à puberdade e sobre o desempenho reprodutivo de novilhas de corte. Pesquisa Agropecuária Brasileira 34:701-707. https://doi.org/10.1590/S0100-204X1999000400023

Restle, J.; Vaz, R. Z.; Alves Filho, D. C.; Bernardes, R. A. L. C.; Pascoal, L. L.; Senna, D. B. and Polli, V. A. 2001. Desempenho de vacas Charolês e Nelore desterneiradas aos três ou sete meses. Revista Brasileira de Zootecnia 30:499-507. https://doi.org/10.1590/S1516-35982001000200029

Rocha, M. G. and Lobato, J. F. P. 2002. Avaliação do desempenho reprodutivo de novilhas de corte primíparas aos dois de idade. Revista Brasileira de Zootecnia 31:1388-1395. https://doi.org/10.1590/S1516-35982002000600009

Schleder, G. R. and Fazzio, A. 2021. Machine learning na física, química, e ciência de materiais: descoberta e design de materiais. Revista Brasileira de Ensino de Física 43:e20200407. https://doi.org/10.1590/1806-9126-RBEF-2020-0407

Shahinfar, S.; Page, D.; Guenther, J.; Cabrera, V.; Fricke, P. and Weigel, K. 2014. Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. Journal of Dairy Science 97:731-742. https://doi.org/10.3168/jds.2013-6693

Silva, F. V.; Soares, F. D. S.; Oliveira, L. L. S.; Toral, F. L. B.; Reis, S. T.; Alves, D. D.; Rocha Júnior, V. R. and Gomes, R. R. R. 2015. Componentes principais das características de carcaças de bovinos anelorados e fontes de variação em lesões. Pesquisa Veterinária Brasileira 35:148-152. https://doi.org/10.1590/S0100-736X2015000200009

Silva, F. M. B.; Lopes, D. T.; Ferraz, H. T.; Viu, M. A. O.; Ramos, D. G. S.; Saturnino, K. C.; Fontana, C. A. P.; Silva, J. M. A. and Leso, F. V. 2018. Estratégias para antecipação da puberdade em novilhas *Bos taurus indicus* pré-púberes. Pubvet 12:a225.

Vaz, R. Z. and Lobato, J. F. P. 2010. Efeito da idade de desmame no desempenho reprodutivo de novilhas de corte expostas à reprodução aos 13/15 meses de idade. Revista Brasileira de Zootecnia 39:142-150. https://doi.org/10.1590/S1516-35982010000100019

Vaz, R. Z.; Restle, J.; Pacheco, P. S.; Vaz, F. N.; Pascoal, L. L. and Vaz, M. B. 2012. Ganho de peso pré e pós-desmame no desempenho reprodutivo de novilhas de corte aos quatorze meses de idade. Ciência Animal Brasileira 13:272-281. https://doi.org/10.5216/cab.v13i3.17527

Vaz, R. Z.; Conceição, V. G. D.; Espigolan, R. and Dutra, M. M. M. 2024. Abordagem multivariada de variáveis potencializadoras de contusões em carcaças bovinas. Revista Thêma et Scientia 14:261-274.

Zhang, M.; Wang, X.; Feng, H.; Huang, Q.; Xiao, X. and Zhang, X. 2021. Wearable Internet of Things enabled precision livestock farming in smart farms: A review of technical solutions for precise perception, biocompatibility, and sustainability monitoring. Journal of Cleaner Production 312:127712. https://doi.org/10.1016/j.jclepro.2021.127712