R. Bras. Zootec., 53:e20240001, 2024
https://doi.org/10.37496/rbz5320240001

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

Comparison of supervised machine learning and variable selection methods for body weight prediction of growth pigs using image processing data

Eula Regina Carrara[1], Polliany da Costa Santos Oliveira[1], Layla Cristien de Cássia Miranda Dias[1], Weverton Gomes da Costa[2], Aline Rabello Conceição[1], Pedro Henrique Silva Braga[1], Mario Luiz Chizzotti[1], Renata Veroneze[1], Erica Beatriz Schultz[1]*

[1] Universidade Federal de Viçosa, Departamento de Zootecnia, Viçosa, MG, Brasil.
[2] Universidade Federal de Viçosa, Departamento de Estatística, Viçosa, MG, Brasil.

***Corresponding author:**
erica.schultz@ufv.br

Received: January 29, 2024
Accepted: July 23, 2024

How to cite: Carrara, E. R.; Oliveira, P. C. S.; Dias, L. C. C. M.; Costa, W. G.; Conceição, A. R.; Braga, P. H. S.; Chizzotti, M. L.; Veroneze, R. and Schultz, E. B. 2024. Comparison of supervised machine learning and variable selection methods for body weight prediction of growth pigs using image processing data. Revista Brasileira de Zootecnia 53:e20240001.
https://doi.org/10.37496/rbz5320240001

Editors:
Marcos Inácio Marcondes
Valdir Ribeiro Junior

ABSTRACT - This research aimed to compare statistical methods (random forest, RIDGE, LASSO, and elastic net regression) for the prediction of body weight in purebred and crossbred pigs reared in Brazil. This prediction was based on dorsal-view images obtained from video image processing. The study involved 69 animals belonging to breeds such as Large White, Piau, Duroc × Large White, and Piau × Large White. The data collection spanned 144 days, with measurements taken at approximately 20-day intervals, totaling eight measurements for each animal throughout their growth stages. Image acquisition was carried out in individual pens using an Intel RealSense Depth D435 digital camera. The features back area, back perimeter, back width, and body depth were extracted from the images. Pearson's correlation analysis was conducted to assess the relationship between live weight and these features. The dataset was randomly divided into a training dataset (65%) and a test dataset (35%), and model training was performed by five-fold cross-validation balanced according to the growth stage, which was divided into three groups. This procedure was repeated 100 times, and the resulting metrics were taken as the average of the 100 repetitions. Although with a slight difference, the random forest method outperformed the others with the highest average R^2 value (0.87), as well as the lowest average RMSE (14.32) and average MAE (10.13) values. Consequently, the random forest algorithm proved to be the most effective in predicting body weight. The back area, back width, and back perimeter were the most important variables in the model.

Keywords: 2D image, back area, crossbred pig, penalized regression, precision livestock farming, random forest

1. Introduction

Traditionally, pig weighing is performed using manual methods that require the animal to be physically restrained. This approach is labor-intensive and time-consuming, especially on large farms with many animals (Li et al., 2014). Usually, manual weight measurements are taken at the end of each phase and in many commercial farms, only pen weights are registered. It affects the ability of the farm to detect short-term fluctuations in animal body weight, resulting in the loss of individual variability (Fernandes et al., 2019).

In this way, the use of techniques with potential for more efficient and accurate recording of body weight can help overcome these limitations. Two-dimensional video images have some advantages for body weight recording, such as low cost, ease of use and no need to handle animals, avoidance of stress and less labor, and the possibility of more frequent data collection. In addition, current automated phenotyping technologies permit the documentation of an infinity of non-standard phenotypes, including images and videos, which yield thousands of complex phenotypes.

The application of machine learning techniques and penalized regression emerges as an approach for the analysis of complex phenotypes. Supervised statistical learning with regularization, such as tree-based methods, boosting, and penalized regression with variable selection, have been widely used in studies involving production animals (Gorczyca et al., 2018; Nguyen et al., 2020; He et al., 2021). Among them, the random forest (RF) algorithm is frequently used for data mining and prediction analysis (Chen and Ishwaran, 2012). The RF method combines a bagging sampling approach and random feature selection to assemble a set of decision trees to provide controlled variation in the modeling process (Breiman, 2001).

In regression problems, the aim is to minimize the sum of squared errors (SSE). This objective, when augmented by the inclusion of a lambda (λ) penalty, gives rise to the development of penalized regression methods: ridge regression (RIDGE; Hoerl and Kennard, 1970), least absolute shrinkage and selection operator regression (LASSO; Tibshirani, 1996), and elastic net regression (ENET; Zou and Hastie, 2005).

Currently, a restricted quantity of ongoing research is dedicated to constructing predictive models for estimating pig body weight using images (Brandl and Jørgensen, 1996; Fernandes et al., 2019; Yu et al., 2021). Additionally, most of the studies are carried out using only commercial breeds. No investigations have utilized images and machine learning methods to predict the body weight of fat-type pig breeds such as Piau, a Brazilian breed, or their associated crossbreeds.

Thus, this study aimed to compare the statistical methods RF, RIDGE, LASSO, and ENET to predict the body weight of purebred and crossbred pigs based on dorsal-view images obtained from video image processing.

2. Material and Methods

Research on animals was conducted according to the institutional committee on animal use (014/2022).

2.1. Data collection

The experiment was conducted in Viçosa, Minas Gerais, Brazil (−20.77683, −42.85908). Initially, 69 animals of the Large White (LL; n = 16), Piau (PP; n = 14), Duroc × Large White (DL; n = 18), and Piau × Large White (PL; n = 21) breeds were evaluated. The animals were allocated in individual concrete floor pens (2.41 × 1.36 m), with free access to water and feed.

The live weight was obtained using a digital floor weighing scale. Data was collected during 144 days at intervals of approximately 20 days, comprising eight measurements per animal from the weaning to the finishing phases. The growth stages were categorized into three groups to balance the dataset partition of the model training, as outlined: group 1 - weaning (28.84±1.76 days old) and nursery period (48.85±1.78 days old); group 2 - at end of the nursery (63.84±1.76 days old) and three measurements during growth (78.95±1.86, 98.95±1.86, 119.97±1.87 days old); and group 3 - during finishing (140.95±1.86 days old) and in the end of finishing for slaughter (172.77±1.84 days old). The births occurred for one week, contributing to the observed variation in the age of the animals.

During the experimental phase, some animals died, and some records were missing, resulting in a varying number of measurements for each day. There were 483 measurements (Table 1), being 138, 241, and 104 in groups 1, 2, and 3, respectively. The complete data (i.e., considering all growth stages) was evaluated to ensure a larger dataset.

Table 1 - Number of animals measured at each group, by breed

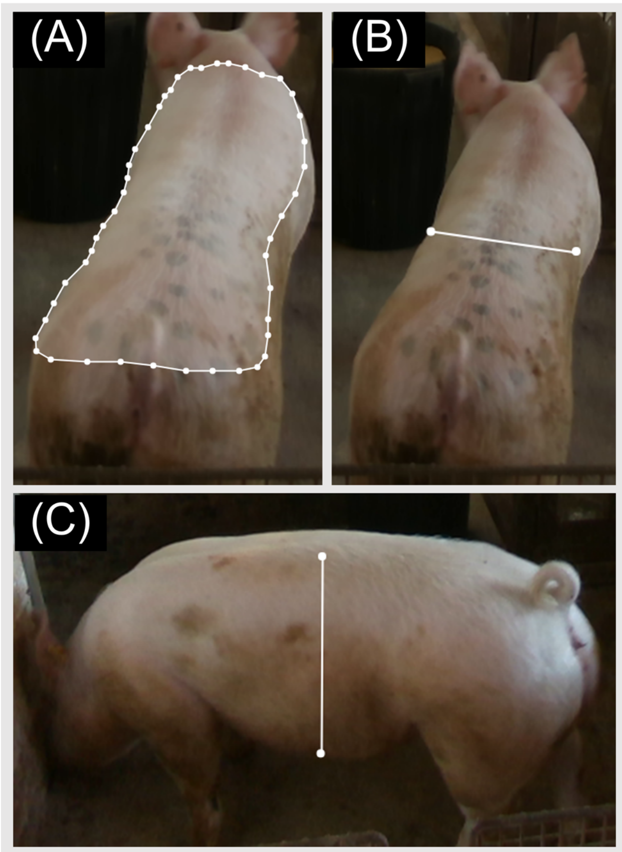
Breed	Group 1		Group 2				Group 3	
	Weaning	Nursery	Leaving the nursery	Growth 1	Growth 2	Growth 3	Finishing	Leaving finishing
LL	16	15	15	15	15	14	14	15
PP	14	15	16	13	13	13	13	10
DL	18	18	18	16	16	15	16	13
PL	21	21	20	14	14	14	13	10
Total for stage	69	69	69	58	58	56	56	48

LL - Large White; PP - Piau; DL - Duroc × Large White; PL - Piau × Large White.
Group 1 = weaning and nursery period; group 2 = at end of the nursery and three measurements during growth; group 3 = during finishing and at the end of finishing for slaughter.

2.2. Video, frame processing, and features extracting

Immediately after the animals were taken to the weighing scale to measure their live weight, individual imaging was collected using an Intel RealSense Depth D435 digital camera with 1920 × 1080 pixels resolution. The camera was positioned on a tripod at a distance and a height of 1.5 m from the animals. Videos were taken lasting between 30 and 40 s for each animal, focusing on the dorsal and lateral regions.

The videos were processed to select manually the best frames of individual dorsal and lateral positions of each animal. This step was performed using the Intel RealSense Viewer video software. The features of the back area, back perimeter, back width, and body depth (Figure 1) were extracted



A - back perimeter and back area; B - back width; C - body depth.

Figure 1 - Examples of the features extracted from the images of pigs.

from the images, all of them in pixels. The features were extracted semi-automatically using the ImageJ free software. The back area, back perimeter, and back width were extracted from the back region located with reference to the first thoracic and last caudal vertebrae, and the body death was extracted from the lateral region between the 12th and 13th thoracic vertebrae.

2.3. Statistical analysis

Firstly, a Pearson's correlation was performed at a 5% significance level among live weight and the features. In sequence, the data were partitioned randomly into two parts: the training dataset (65%) and the test dataset (35%), balanced by grouped growth stage (group 1, group 2, and group 3), and four statistical methods were used to construct predictive models: RF and the penalized regression methods RIDGE, LASSO, and ENET. The analyses were performed using the R packages *caret* (Kuhn, 2008), *randomForest* (Liaw and Wiener, 2002), and *glmnet* (Friedman et al., 2010).

The data was partitioned into a training set and a test set in a balanced manner by group, as this approach yielded superior results in a previous analysis (not shown). This analysis partitioned the data by group, by breed, and by group + breed. The partitioning by group alone yielded the most optimal metrics. Similarly, partitioning ratios spanning from 50 to 90% in increments of 5 were evaluated, with the 65% ratio exhibiting the most optimal metrics (not shown).

The RF assembly was performed following these steps: a collection of bootstrap samples (*ntree*) from the initial dataset was generated; construction of an individual tree for each bootstrap dataset with random selection of variables (*mtry*) at each node with a specified number of cases (*nodesize*); building predictions for new data with the information gathered from the ensemble of trees; and utilizing the data that was not included in the original bootstrap sample (test data) to compute the out-of-bag (OOB) error rate. In our study, the following hyperparameters were tested in a training dataset by grid search 5-fold cross-validation: *ntree* equal to 150, 250, 350, and 500; *mtry* equal to 2, 3, and 4; and *nodesize* equal to 2, 4, 6, and 8. The final RF model used *ntree* equal to 500 and *mtry* and *nodesize* equal to 2 as best hyperparameters.

In penalized regressions with variable selection, a penalty was added in the SSE each time the parameter had a high value, causing the parameter shrinkage:

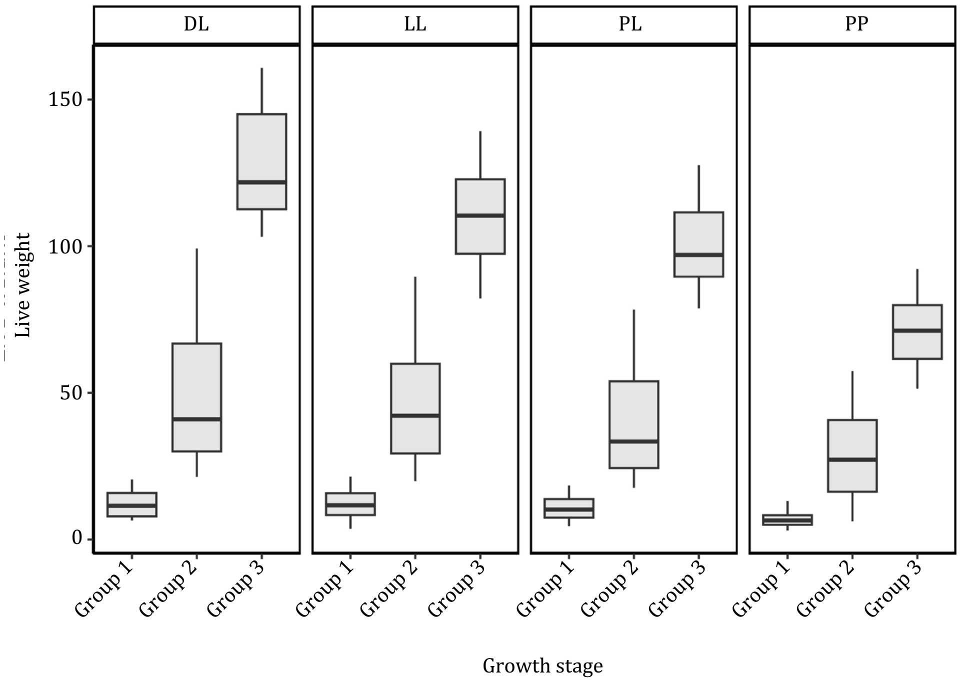
$$SSE = \sum_i^n (y_i - \hat{y}_i)^2 + \lambda((1 - \alpha) \sum_j^p \beta_j^2 + \alpha \sum_j^p |\beta_j|),$$

in which the intensity of the penalty is controlled by the λ value, and the α value defines the method, being: $\alpha = 0$ is the RIDGE, in which the quadratic coefficients of the parameters (β_j^2) are penalized, and no variables are selected; $\alpha = 1$ is the LASSO, which penalizes the absolute values of the parameter coefficients ($|\beta_j|$) and, as a consequence, less informative predictors have their coefficient close to or equal to zero; and $0 < \alpha < 1$ is the ENET, which combines the RIDGE and LASSO penalties, i.e., α represents the weight of RIDGE or LASSO regularization. In our study, the optimal λ value for each regression method (RIDGE, LASSO, and ENET) was selected through five-fold cross-validation in the training dataset. Additionally, the α value for ENET ranged from 0.1 (ENET1) to 0.9 (ENET9), i.e., $\alpha = 0.2$ (ENET2), $\alpha = 0.3$ (ENET3), etc., totaling nine ENET models.

After applying the trained model to the test dataset, predicted pig weights were generated for each algorithm. The predicted and observed weights were then compared using simple linear regression, which provided the metrics coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). All the steps, from the train-test split to the obtaining of the metrics, were carried out 100 times. This was done to ensure the reliability of the results and avoid any potential bias or high prediction performance due to chance if we had only conducted one train-test split. The final metrics were obtained by averaging the 100 repetitions. Based on the highest average R^2 score and the lowest average RMSE and average MAE values, the most effective algorithm for weight prediction was chosen.

3. Results

The growth pattern of the animals varied among the breeds (Figure 2, Table 2). Concerning the regression metrics between the predicted and observed data, there was minimal variation throughout the 100 repetitions of the train-test partition (Figure 3). The prediction algorithms showed little difference between the average metrics for precision and accuracy, with R^2 values between 0.85 and 0.87, RMSE values between 14.32 and 15.23, and MAE values between 10.13 and 11.19. Furthermore,



DL - Duroc × Large White; LL - Large White; PL - Piau × Large White; PP - Piau.
Group 1 = weaning and nursery period; group 2 = at end of the nursery and three measurements during growth; group 3 = during finishing and at the end of finishing for slaughter.

Figure 2 - Live weight measurements by breed and by group.

Table 2 - Mean, standard deviation (SD), median, and minimum (MIN) and maximum (MAX) values of live weight by group and by breed

Group	Breed	Mean ± SD	Median	MIN	MAX
Group 1	DL	12.00±4.38	11.48	6.50	20.45
	LL	11.99±4.60	11.65	3.65	21.45
	PL	10.79±3.88	10.20	4.55	18.45
	PP	6.95±2.62	6.50	3.05	13.10
Group 2	DL	51.26±24.32	41.00	21.35	99.20
	LL	47.24±20.55	42.20	19.90	20.55
	PL	41.71±18.90	33.40	17.65	78.40
	PP	27.83±13.77	27.20	6.15	57.40
Group 3	DL	129.59±19.17	121.80	103.20	160.80
	LL	111.86±16.11	110.40	82.20	139.20
	PL	101.43±14.33	97.00	78.80	127.60
	PP	71.06±12.16	71.20	51.40	92.20

LL - Large White; PP - Piau; DL - Duroc × Large White; PL - Piau × Large White.
Group 1 = weaning and nursery period; group 2 = at end of the nursery and three measurements during growth; group 3 = during finishing and at the end of finishing for slaughter.

there was no difference between the ENET models with different alpha values (Table 3). Although with a slight difference, the random forest method outperformed the others with the highest average R^2 value (0.87), as well as the lowest average RMSE (14.32) and average MAE (10.13) values (Figure 4).

The correlation between live weight and the features related to the back (back area, back perimeter, and back width) was high and positive, ranging from 0.78 to 0.92 (Figure 5). However, the correlation between lateral height and live weight was low (0.42). The higher correlation was between live weight and back area and was equal to 0.97. The most important feature in building the RF predictive model

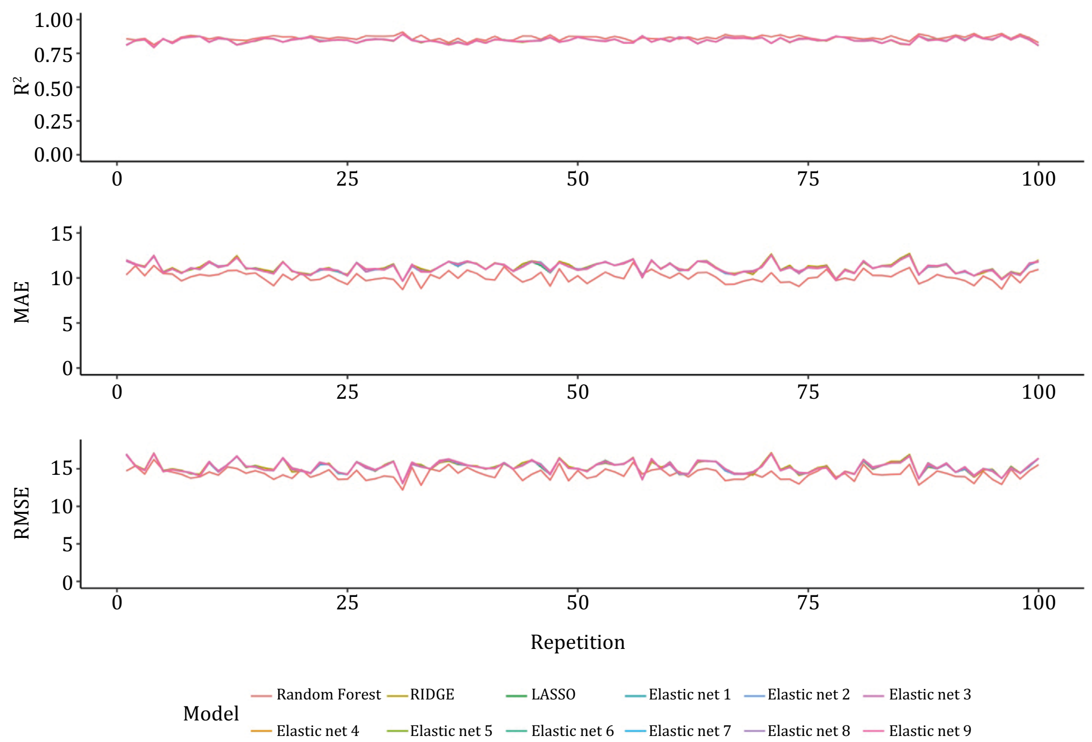


Figure 3 - Regression metrics between predicted and observed data throughout the 100 repetitions of train-test partition.

Table 3 - Descriptive[1] of metrics coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) considering 100 repetitions

Model	R^2	RMSE	MAE
RF	0.87±0.02 [0.81:0.91]	14.32±0.75 [12.19:16.24]	10.13±0.60 [8.73:11.72]
RIDGE	0.85±0.02 [0.79:0.89]	15.22±0.80 [13.03:17.15]	11.19±0.59 [9.67:12.71]
LASSO	0.85±0.02 [0.80:0.89]	15.23±0.80 [13.10:17.06]	11.13±0.59 [9.65:12.52]
ENET1	0.85±0.02 [0.80:0.89]	15.18±0.80 [13.05:17.04]	11.14±0.59 [9.65:12.55]
ENET2	0.85±0.02 [0.80:0.89]	15.18±0.80 [13.05:17.04]	11.14±0.59 [9.65:12.55]
ENET3	0.85±0.02 [0.80:0.89]	15.18±0.80 [13.05:17.04]	11.14±0.59 [9.65:12.55]
ENET4	0.85±0.02 [0.80:0.89]	15.18±0.80 [13.05:17.04]	11.14±0.59 [9.65:12.55]
ENET5	0.85±0.02 [0.80:0.89]	15.18±0.80 [13.05:17.04]	11.14±0.59 [9.65:12.55]
ENET6	0.85±0.02 [0.80:0.89]	15.18±0.80 [13.05:17.04]	11.14±0.59 [9.65:12.55]
ENET7	0.85±0.02 [0.80:0.89]	15.18±0.80 [13.05:17.04]	11.14±0.59 [9.65:12.55]
ENET8	0.85±0.02 [0.80:0.89]	15.18±0.80 [13.05:17.04]	11.14±0.59 [9.65:12.55]
ENET9	0.85±0.02 [0.80:0.89]	15.18±0.80 [13.05:17.04]	11.14±0.59 [9.65:12.55]

RF - random forest; RIDGE - ridge regression; LASSO - least absolute shrinkage and selection operator regression; ENET1 - elastic net with α = 0.1; ENET2 - elastic net with α = 0.2; ENET3 - elastic net with α = 0.3; ENET4 - elastic net with α = 0.4; ENET5 - elastic net with α = 0.5; ENET6 - elastic net with α = 0.6; ENET7 - elastic net with α = 0.7; ENET8 - elastic net with α = 0.8; ENET9 - elastic net with α = 0.9.
[1] Mean ± standard deviation [minimum value: maximum value].

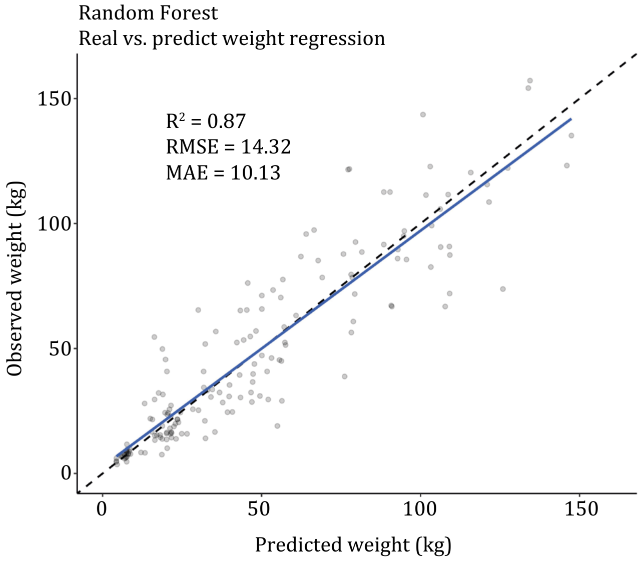
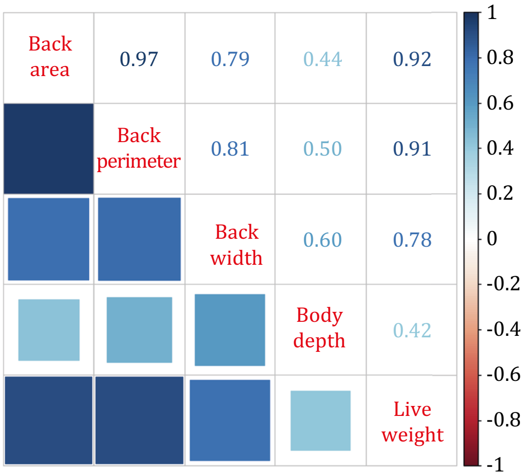


Figure 4 - Linear regression of the observed weight and the weight predicted by random forest regression approach (best performance), with the respective coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE).



All the correlations were significant (P-value < 0.001).

Figure 5 - Pearson's correlation among live weight and the features obtained from image processing for back area, back perimeter, back width, and body depth of evaluated pigs.

was the back area (Figure 6). The importance of variables in the RIDGE, LASSO, and all ENET approaches presented the same pattern as that in the RF approach (Figure 7). There was no difference between the ENET models (ENET1 to ENET9), and thus, only one graph was plotted for ENET1-9 (Figure 7).

4. Discussion

Four statistical methods (RF and RIDGE, LASSO, and ENET) were evaluated to predict the body weight of purebred and crossbred pigs based on dorsal-view images obtained from video image processing.

Regarding live weights, the purebred PP exhibited the lowest values across all growth stages, as expected, due to their smaller size. The average weaning weight in this population of the PP breed is 6.60 kg with a standard deviation of 1.84 kg (Oliveira et al., 2023). In our study, the average weight

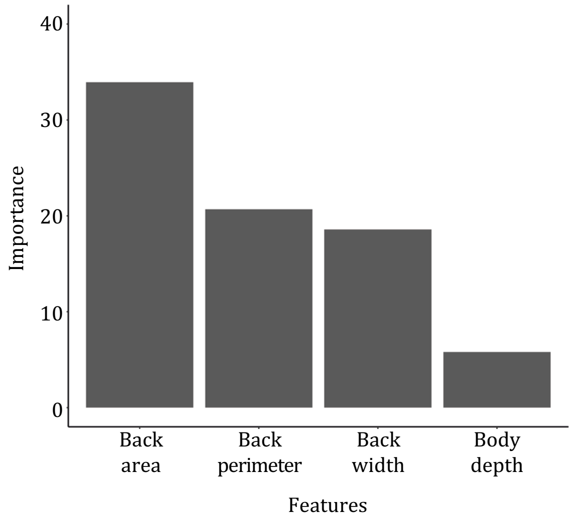
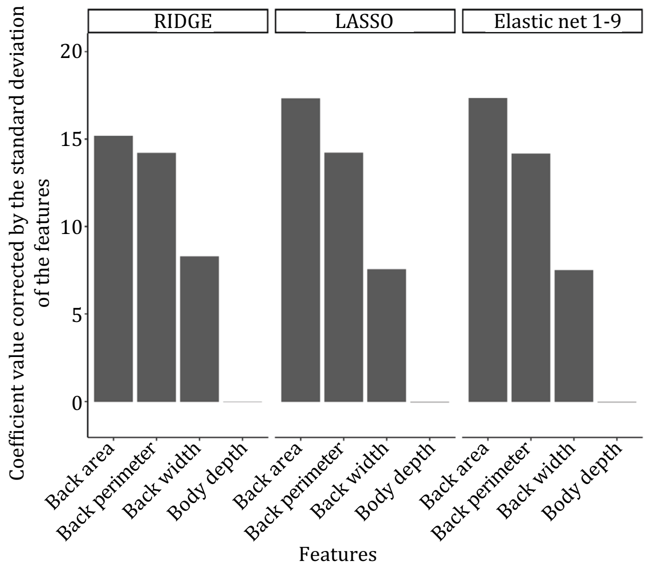


Figure 6 - Importance of the image features back area, back perimeter, back width, and body depth in the prediction model for body weight in pigs using the random forest regression approach.



The α value for elastic net ranged from 0.1 (elastic net 1) to 0.9 (elastic net 9), i.e., α = 0.2 (elastic net 2), α = 0.3 (elastic net 3), etc., totaling nine elastic net models. However, there was no difference among these models, and thus only one graph was plotted for elastic net 1-9.

Figure 7 - Contribution of the features in the RIDGE, LASSO, and elastic net regression models.

at the weaning stage for PP was slightly higher, with a mean of 6.95 kg and a standard deviation of 2.62 kg. This is because animals in both the weaning (leaving the farrowing stage) and nursery stages were included in the same "group", increasing the mean and the deviation. This was required because the data volume would be lower when assessed separately according to growth phases. The average weights at the growth (27.83±13.77 kg) and finishing (71.06±12.16 kg) stages are following previous studies for the Piau breed at the same stages (35.00±4.00 kg and 65.20±4.20 kg, respectively) (Silva et al., 2019).

The purebred LL exhibited intermediate live weight values compared with the crosses PL and DL, with lower values for PL and higher values for DL. The lower live weight values for PL can be attributed to the use of the smaller purebred PP as parental. Similarly, the DL crosses exhibited the highest live weight values throughout their growth, as both purebred parentals are larger-sized animals. In fact, Duroc animals can show an average daily weight gain of 1,062 g and reach 100 kg in 137 days, while Large White animals can show an average daily weight gain of 1,016 g and reach 100 kg in 147.5 days (Tretyakova et al., 2021).

The RF method was slightly better than RIDGE, LASSO, and all ENET (ENET1 to ENET9), with a higher average R^2 value and lower average RMSE and average MAE values. Thus, RF was the most effective algorithm for predicting body weight. There were no disparities within the penalized regression with variable selection methods, likely due to the limited number of assessed features, and consequently, a low chance of parameter penalization across the different methods.

Other studies point to the superiority of RF for weight prediction problems using features extracted from digital images, in plants and animals. Duc et al. (2023) used several features extracted by digital images (e.g., area size, perimeter length, length, width, and others) to predict soybean seed weight. They demonstrated the superiority of the RF method over the RIDGE, LASSO, and ENET methods. Sant'Ana et al. (2021) used eight machine learning models to predict body weight in sheep using a variety of features related to the shape, size, and angles of digital images, and the RF model was the approach that obtained the best performance.

Although the precision was high in the RF approach (87%), the MAE pointed to a variation of up to 10.13 kg, indicating that the model may not be accurate, mainly at younger ages. There is greater variability in the observed weight in group 3 (i.e., growing and finishing stages) (Figure 2). This greater dispersion leads to greater variability in the predicted weights and, consequently, increases the prediction error. Additionally, only 11 animals (2.28%) had a live weight above 140 kg, which makes it difficult to predict heavier animals. Additionally, it is important to note that data variability is crucial in training robust models, while data with little variability may negatively impact their predictability.

In this sense, a study conducted by Fernandes et al. (2019) used features of body measurements and shape descriptors extracted from digital images to predict body weight in pigs in the nursery and finishing stages. The authors reported an average R^2 of 0.92 and MAE of 0.35 for the models including all animals (nursery and finishing stages). An average R^2 of 0.80 and MAE of 0.30, when considering only animals in the finishing stage, suggest that the inclusion of animals at younger ages (with less variability) contributes to increasing the accuracy of the model, without substantially modifying the MAE.

Given the results from Fernandes et al. (2019), we re-analyzed our database using the RF model and only the growing and finishing data (n = 345). We found a lower average R^2 (0.74) and higher average RMSE (18.89) and average MAE (14.91) compared with the analyses performed with the complete database, i.e., in our study, precision and accuracy are higher when we include all animals (nursery, growing, and finishing) in the analysis. The high accuracy of the RF model can be explained by the correlation between live weight and the features evaluated, which corroborates with the importance of each feature in building the predictive model, which pointed to the back area as the most important feature. Similarly, body depth showed the lowest correlation with live weight and was the feature of the lowest importance in building the predictive model.

The greater importance of features related to the back region can be explained by the fact that the region is large and representative of the pig's body size, as they accompany animal growth. In the study by Brandl and Jørgensen (1996), the area and perimeter of the back of pigs were used to create a predictive model for body weight using spline functions, and the model showed an R^2 of 0.98, indicating high precision in predicting body weight using these features. Fernandes et al. (2019) used various measurements of the back of pigs, such as area and various length and height measurements, to build a predictive model for body weight using three-dimensional images and reported a high precision of 0.92, with an MAE of 3.5%. The dorsal region is widely explored in animal prediction studies, since cameras are usually fixed to the top of barns or imaging is performed by drones, where the dorsal area is best captured in these situations.

The lower importance of body depth in predicting live weight in our study can be explained by the greater difficulty in obtaining lateral images of the animals in our case. The space for taking the images was limited and the images could only be taken from above and could not capture the curvature of the belly, for example. In addition, the animals had longer body lengths and smaller body depths, increasing the importance of the back area for body weight prediction models and reducing the importance of body depth in that case.

In any case, the features related to the dorsal area were sufficient to predict the body weight of growing pigs with a precision of over 87% using the RF method. It is hoped that the advent of real-time data collection using images will contribute to advances in body weight monitoring in pigs, especially images related to the animals' backs.

5. Conclusions

The random forest machine learning algorithm was slightly better than RIDGE, LASSO, and elastic net penalized regression algorithms for predicting body weight of pigs. It was possible to predict the pigs' body weight by using image measurements and the random forest algorithm with an R^2 of 87%, with the area, width, and perimeter of the back being the most important variables.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization: Schultz, E. B. **Data curation:** Carrara, E. R.; Oliveira, P. C. S.; Dias, L. C. C. M.; Conceição, A. R. and Braga, P. H. S. **Formal analysis:** Carrara, E. R. and Costa, W. G. **Funding acquisition:** Chizzotti, M. L.; Veroneze, R. and Schultz, E. B. **Investigation:** Carrara, E. R. and Veroneze, R. **Methodology:** Carrara, E. R.; Oliveira, P. C. S.; Costa, W. G. and Schultz, E. B. **Project administration:** Chizzotti, M. L. and Schultz, E. B. **Supervision:** Veroneze, R. and Schultz, E. B. **Writing – original draft:** Carrara, E. R.; Oliveira, P. C. S.; Veroneze, R. and Schultz, E. B. **Writing – review & editing:** Carrara, E. R.; Oliveira, P. C. S.; Dias, L. C. C. M.; Costa, W. G.; Conceição, A. R.; Braga, P. H. S.; Chizzotti, M. L.; Veroneze, R. and Schultz, E. B.

Acknowledgments

We acknowledge the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, Process: RED-00172-22, APQ-01457-22, BPD-00477-22); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant number 312454/2022-8); and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, PROEX grant number 32002017011P9).

References

Brandl, N. and Jørgensen, E. 1996. Determination of live weight of pigs from dimensions measured using image analysis. Computers and Electronics in Agriculture 15:57-72. https://doi.org/10.1016/0168-1699(96)00003-8

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

Chen, X. and Ishwaran, H. 2012. Random forests for genomic data analysis. Genomics 99:323-329. https://doi.org/10.1016/j.ygeno.2012.04.003

Duc, N. T.; Ramlal, A.; Rajendran, A.; Raju, D.; Lal, S. K.; Kumar, S.; Sahoo, R. N. and Chinnusamy, V. 2023. Image-based phenotyping of seed architectural traits and prediction of seed weight using machine learning models in soybean. Frontiers in Plant Science 14:1206357. https://doi.org/10.3389/fpls.2023.1206357

Fernandes, A. F. A.; Dórea, J. R. R.; Fitzgerald, R.; Herring, W. and Rosa, G. J. M. 2019. A novel automated system to acquire biometric and morphological measurements and predict the body weight of pigs via 3D computer vision. Journal of Animal Science 97:496-508. https://doi.org/10.1093/jas/sky418

Friedman, J. H.; Hastie, T. and Tibshirani, R. 2010. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33:1-22. https://doi.org/10.18637/jss.v033.i01

Gorczyca, M. T.; Milan, H. F. M.; Maia, A. S. C. and Gebremedhin, K. G. 2018. Machine learning algorithms to predict core, skin, and hair-coat temperatures of piglets. Computers and Electronics in Agriculture 151:286-294. https://doi.org/10.1016/j.compag.2018.06.028

He, Y.; Tiezzi, F.; Howard, J. and Maltecca, C. 2021. Predicting body weight in growing pigs from feeding behavior data using machine learning algorithms. Computers and Electronics in Agriculture 184:106085. https://doi.org/10.1016/j.compag.2021.106085

Hoerl, A. E. and Kennard, R. W. 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12:55-67. https://doi.org/10.1080/00401706.1970.10488634

Kuhn, M. 2008. Building predictive models in R using the caret package. Journal of Statistical Software 28:1-26. https://doi.org/10.18637/jss.v028.i05

Li, Z.; Luo, C.; Teng, G. and Liu, T. 2014. Estimation of pig weight by machine vision: a review. p.42-49. In: Li, D. and Chen, Y. (eds). Computer and Computing Technologies in Agriculture VII. CCTA 2013. IFIP Advances in Information and Communication Technology, v. 420. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-54341-8_5

Liaw, A. and Wiener, M. 2002. Classification and regression by random forest. R News 2:18-22.

Nguyen, Q. T.; Fouchereau, R.; Frénod, E.; Gerard, C. and Sincholle, V. 2020. Comparison of forecast models of production of dairy cows combining animal and diet parameters. Computers and Electronics in Agriculture 170:105258. https://doi.org/10.1016/j.compag.2020.105258

Oliveira, L. F.; Lopes, P. S.; Dias, L. C. C. M.; Silva, L. M. D.; Silva, H. T.; Guimarães, S. E. F.; Marques, D. B. D.; da Silva, D. A. and Veroneze, R. 2023. Estimation of genetic parameters and inbreeding depression in Piau pig breed. Tropical Animal Health and Production 55:14. https://doi.org/10.1007/s11250-022-03428-9

Sant'Ana, D. A.; Pache, M. C. B.; Martins, J.; Soares, W. P.; de Melo, S. L. N.; Garcia, V.; Weber, V. A. M.; Heimbach, N. S.; Mateus, R. G. and Pistori, H. 2021. Weighing live sheep using computer vision techniques and regression machine learning. Machine Learning with Applications 5:100076. https://doi.org/10.1016/j.mlwa.2021.100076

Silva, H. T.; Silva, F. F.; Ferreira, A. S.; Veroneze, R. and Lopes, P. S. 2019. Evaluation of Bayesian models for analysis of crude protein requirement for pigs of Brazilian Piau breed. Scientia Agricola 76:208-213. https://doi.org/10.1590/1678-992x-2017-0256

Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58:267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tretyakova, O.; Degtyar, A.; Avdeyev, A.; Ovchinnikov, D. and Morozyuk, I. 2021. Features of the growth and development of young pigs of various breeding. E3S Web Conferences 273:02012. https://doi.org/10.1051/e3sconf/202127302012

Yu, H.; Lee, K. and Morota, G. 2021. Forecasting dynamic body weight of nonrestrained pigs from images using an RGB-D sensor camera. Translational Animal Science 5:txab006. https://doi.org/10.1093/tas/txab006

Zou, H. and Hastie, T. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology 67:301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x