JOURNAL ON COMMUNICATIONS

ISSN:1000-436X

Scopus®

REGISTERED

www.jocs.review

Machine learning algorithms in predicting piglet mortality and weight based on sow backfat thickness

William Armando Tapie[1**], Carlos-Leonardo Guerra-Marín[1], Juan Fernando Manrique-Hincapie[1], Sebastian Montoya Uribe[1]

[1]*Grupo de Investigación en agronomía y zootecnia-GIAZ, Facultad de Ciencias Agropecuarias, Universidad Católica de Oriente Sector 3, cra. 46 No. 40B 50, Rionegro 054040, Antioquia*

Keywords:

Machine learning; Backfat thickness; Cross-validation; Neural networks; Random Forest

ABSTRACT

The implementation of machine learning models to predict reproductive variables in sows represents a fundamental tool for optimizing management in swine production systems. The objective of this study was to evaluate the predictive capacity of random forest (RF) and dense neural network (DNN) models to predict live birth piglets (LBP), total piglets born (TPB), average piglet weight (APW), and piglet mortality rate (PMR), using backfat thickness (BFT) as the predictor variable. A total of 571 records from a commercial swine farm were analyzed, of which 348 were used after data cleaning. Models were developed in Google Colaboratory using cross-validation for hyperparameter tuning. Performance was evaluated using mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), and relative percentage error (RPE). For LBP, the MAE was 1.51 in both models, with MSEs of 3.62 (DNN) and 3.53 (RF), and RMSEs of 1.90 (DNN) and 1.88 (RF), respectively. The RPEs were 13.59% and 13.41%, respectively. For TPB, the MAE was 1.61 in both models, with MSE values of 3.99 (DNN) and 3.95 (RF), and RMSE values of 2.00 (DNN) and 1.99 (RF), respectively. The RPE was 13.74% and 13.64%, respectively. For APW, both models showed satisfactory fit with RMSE of 0.063 kg and RPE close to 4.5%. PMR prediction was poor, with RPE exceeding 92% in both models. RF and DNN models demonstrated predictive potential for LBP, TPB, and APW using only BFT.

INTRODUCTION

In swine production, the reproductive efficiency of sows represents a fundamental aspect for economic profitability. Among the multiple factors that condition reproductive performance, body condition represents a key indicator that reflects the energy reserves accumulated by the animal to sustain reproductive and productive functions (Ajay et al., 2023). Currently, body condition is quantified through backfat thickness (BFT) measurement, an objective and non-invasive method that allows estimation of this condition due to its relationship with physiological variables such as milk production, piglet weight at birth and weaning, and litter survival (Díaz et al., 2015; Hu and Yan, 2022).

In addition to its value as a productive indicator, BFT has been associated with the reproductive performance of sows. Previous studies have demonstrated that young animals with higher BFT values reach puberty at earlier ages and present larger litters than those with lower lipid reserves. This effect is mediated by greater sensitivity to reproductive hormones and more efficient follicular development, which favors ovulation and fertility (Roongsitthichai and Tummaruk, 2014).

Optimal BFT in sows ranges from 12-22 mm. However, deviations from this interval generate counterproductive consequences. Elevated BFT (≥ 25 mm) can create a lipotoxic metabolic environment in the placenta, leading to increased stillbirth rates and reduced birth weight (Ha et al., 2024; Zhou et al., 2018). Conversely, low BFT (< 12 mm) compromises milk production and elevates pre-weaning mortality, which can reach levels of 15% (Hu and Yan, 2022).

Within the framework of precision livestock farming, developing analytical tools capable of anticipating productive outcomes constitutes a priority, especially when their application is performed directly

*Corresponding author: William Armando Tapie

with data obtained under real swine production conditions. Facing this challenge, artificial intelligence (AI), and particularly machine learning (ML), has emerged as a promising tool for predicting productive and reproductive events in intensive systems.

Recent applications include the work by Gauthier et al. (2022) in six farms with 20 years of data, where litter weight at weaning was predicted with minimal error (ME = -0.14 kg) and minimal percentage error (MAPE = 9.01%) using ensemble algorithms. Using dense neural networks (DNN) with ultrasound data, litter size was predicted with a mean absolute error (MAE) of 2.58 (Kousenidis et al., 2022). Similarly, through artificial neural networks (ANN) and morphometric data, Landrace piglet weight was predicted with determination coefficients (R^2) between 0.85 and 0.91 (Preethi et al., 2023), while Random Forest (RF) algorithms successfully classified pig groups with high and low mortality risk with 0.90 precision (Magalhães et al., 2024).

Despite the known limitations of these approaches regarding the requirement for extensive and multivariate datasets, this study proposes to evaluate two widely validated ML algorithms, RF and DNN, to predict four farrowing-related variables: number of live and total births, average birth weight, and pre-weaning mortality. Both DNN and RF have demonstrated the capacity to learn complex and non-linear relationships, successfully predicting farrowing-related parameters in sows (Rahman et al., 2023; Taylor et al., 2023), representing an evolution from traditional techniques (Witten et al., 2016) and reflecting the growing application of ML in agricultural production (Gauthier et al., 2022; James et al., 2021).

The innovation of the present approach lies in the exclusive use of BFT as the sole input variable in a simplified scenario representative of field conditions, particularly in environments where information collection is limited. Rather than establishing a definitive model, this work seeks to elucidate the predictive potential of these methods, providing preliminary evidence for future research applied in precision swine production. In this regard, the objective was to evaluate the predictive capacity of DNN and RF to predict the number of live births (LBP), total births (TPB), average piglet weight (APW), and piglet mortality percentage (PMR), using sow BFT as the predictor variable.

MATERIALS AND METHODS

Exploratory data analysis

A database containing 571 records corresponding to three genetic lines from a commercial swine farm was used, collected over five months (January to May 2021). The information included sow backfat thickness (mm), number of live births, number of total births, average piglet weight at birth (kg), number of farrowings per female, and mortality percentage.

To minimize the effect of outliers that could bias results, the database was cleaned using inclusion criteria based on biologically plausible ranges for the variables: number of live births (minimum 10), number of total births (maximum 18), mortality percentage (below 15%), average piglet weight at birth (between 1.1 to 1.6 kg), and number of previous farrowings per female (maximum 7). After excluding records that did not meet these criteria, a final database of 348 records was obtained, which was used for training and validation of predictive models.

Machine learning models

Dense Neural Networks (DNN): To predict LBP, TPB, APW, and PMR from BFT as the sole predictor variable, a dense neural network was developed using the Keras library from TensorFlow. The model architecture included an input layer (standardized BFT), followed by four densely connected layers with 64 and 32 neurons (tanh activation), and 96 and 32 neurons (ReLU activation), respectively. Each layer incorporated batch normalization and 20% dropout to mitigate overfitting. The output layer consisted of four linear neurons, one for each target variable.

For training, a random split of 80% of the data for training and 20% for testing was applied, with a fixed seed to ensure reproducibility. All variables were standardized using StandardScaler from scikit-learn. The model was optimized with the Adam algorithm (Kingma and Ba, 2014), using mean squared error (MSE) and mean absolute error (MAE) as objective functions. EarlyStopping was implemented to halt training if validation loss did not improve for five consecutive epochs.

Random Forest (RF): An RF model was implemented using the scikit-learn library in Python. Hyperparameter optimization was performed through grid search (GridSearchCV), evaluating multiple combinations to maximize model performance. Adjusted hyperparameters included: number of trees in the forest (100, 200, and 300), maximum tree depth (5, 10, and 20 levels), minimum number of samples to split an internal node (2, 5, and 10), and minimum number of samples required to form a leaf (1, 2, and 4). These combinations allowed balancing the model's predictive capacity with its complexity, mitigating overfitting risks.

Once RF and DNN models were trained and predictions generated for training and test sets, target variables were rescaled to their original scale using the inverse_transform function from StandardScaler, to interpret results in initial units and perform comparative descriptive statistical analysis between actual and predicted values.

Model fitting criteria

Mean Squared Error (MSE): Equation 1 quantifies the average squared difference between observed values (Oi) and predicted values (Pi). This metric provides a measure of prediction variability relative to actual values, applying quadratic penalization to larger deviations (Ko et al., 2023).

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}(\mathrm{Oi} - \mathrm{Pi})^2}{n} \quad (1)$$

Mean Absolute Error (MAE): Equation 2 calculates the average absolute difference between

observed values (Oi) and predicted values (Pi). It represents a measure of the average magnitude of prediction errors without considering their direction, providing a more intuitive interpretation of error in the original measurement units (Magdalena et al., 2015).

$$MAE = \frac{\sum_{i=1}^{n} |(Oi - Pi)|}{n} \quad (2)$$

Relative Percentage Error (RPE): Equation 3 quantifies the average relative percentage error, calculating for each observation the absolute difference between actual and predicted values divided by the actual value, and subsequently averaging all percentage errors. This metric enables normalized model performance evaluation, facilitating comparison between variables with different measurement scales (Tapie et al., 2024).

$$RPE = \left(\frac{\sqrt{\frac{\sum_{i=1}^{n} (Oi - Pi)^2}{n}}}{\bar{X}} \right) * 100 \quad (3)$$

Root Mean Squared Error (RMSE): Equation 4 corresponds to the square root of MSE, providing a measure of average error in the same unit as the target variable. This metric is more intuitive than MSE by being expressed on the same scale as the original data, facilitating practical interpretation of results (Ko et al., 2023; Magalhães et al., 2023).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Oi - Pi)^2}{n}} \quad (4)$$

Development platform

Statistical analysis and model development were performed on the Google Colaboratory platform, which provides a cloud-based computing environment with access to processing resources. The following Python libraries were employed for model construction and evaluation: Scikit-learn (Pedregosa et al., 2012), an open-source machine learning library that offers algorithms, data preprocessing functions, and performance evaluation metrics; and TensorFlow with Keras for dense neural network development. Data analysis and visualization were complemented with additional standard libraries from Python's scientific ecosystem.

RESULTS AND DISCUSSION

Table 1 presents the descriptive statistics of the variables used for model development. Following data cleaning, mortality showed high variability (CV = 93.23%), in contrast to the other variables, which presented coefficients of variation below 14%, facilitating their predictive modeling.

This high variability behavior in mortality percentage was expected, considering that neonatal mortality in piglets is strongly influenced by multicausal factors not represented in the single predictor variable used in this study. Previous research has demonstrated that variables such as parity, previous prolificacy, sow immune status, and peripartum management conditions are determinants in piglet survival, and their omission may significantly limit the predictive capacity of any model (Andersen et al., 2007; Bell et al., 2015; Kinane et al., 2021; Yun et al., 2019). This inherent limitation should be considered when interpreting predictive results for the mortality variable.

Table 1 - Descriptive statistics of the variables employed in model development (n = 348)

An example of a column heading	Mean	Median	Standard deviation	Minimum	Maximum	CV %
BFT (mm)	13.74	14	1.59	9	17	11.55
Number of live births	14.00	14	1.90	10	18	13.57
Total births	14.60	15	2.00	10	18	13.73
Average piglet weight (kg)	1.39	1.40	0.06	1.2	1.5	4.55
Mortality (%)	4.46	1	4.16	1	14.29	93.23

BFT = backfat thickness; CV = the coefficient of variation

The DNN architecture results are presented in **Table 2**. The final neural network was composed of 14 layers, including an initial normalization layer, multiple dense layers with decreasing numbers of neurons, batch normalization layers, dropout layers for regularization, and an output layer with 4 neurons corresponding to the target variables. The architectural design prioritized preventing overfitting through the strategic implementation of dropout and batch normalization, aiming to achieve effective generalization for the simultaneous prediction of multiple reproductive response variables.

Table 2. Neural network architecture

layer	Input shape	Output shape	Notes
Batch normalization	(None, 1)	(None, 1)	Initial batch normalization.
Dense	(None, 1)	(None, 64)	Dense layer with 64 neurons.
Batch normalization	(None, 64)	(None, 64)	Batch normalization.
Dropout	(None, 64)	(None, 64)	Dropout regularization to prevent overfitting.
Dense	(None, 64)	(None, 32)	Dense layer with 32 neurons.
Batch normalization	(None, 32)	(None, 32)	Batch normalization.
Dropout	(None, 32)	(None, 32)	Dropout regularization.
Dense	(None, 32)	(None, 96)	Dense layer with 96 neurons.
Batch normalization	(None, 96)	(None, 96)	Batch normalization.
Dropout	(None, 96)	(None, 96)	Dropout regularization.
Dense	(None, 96)	(None, 32)	Dense layer with 32 neurons.
Batch normalization	(None, 32)	(None, 32)	Batch normalization.
Dropout	(None, 32)	(None, 32)	Dropout regularization.
Dense	(None, 32)	(None, 4)	Output dense layer with 4 neurons.

The RF model was optimized through grid search hyperparameter tuning, where the optimal configuration was identified with the following parameters: maximum depth of five levels per tree, a minimum of two samples for both internal node splitting and leaf formation, and an ensemble of 200 trees. The limited tree depth aimed to prevent overfitting by reducing individual estimator complexity, while the minimum sample requirement promoted model generalization capacity. The ensemble of 200 trees represented the optimal balance between variance reduction in predictions and associated computational cost. A summary of this architectural configuration is presented in **Figure 1**.

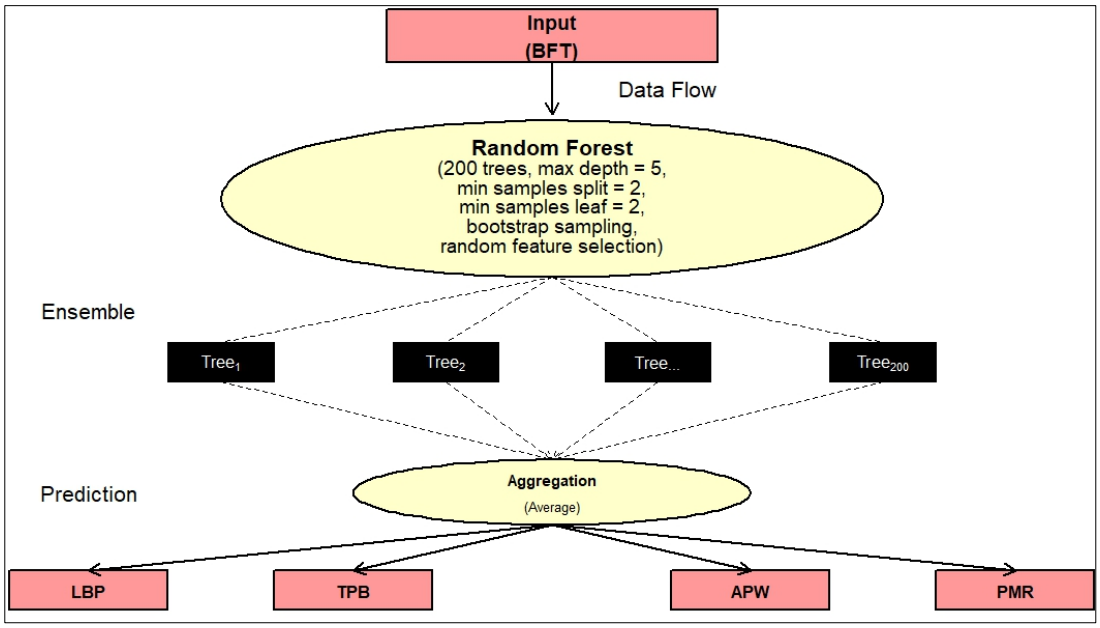

Figure 1- Random Forest model architecture

The performance of machine learning models, both DNN and RF, showed similar results in predicting LBP, TPB, and APW variables (Table 3). For average birth weight, both models (DNN and RF) achieved an RMSE of approximately 0.063 kg, MSE of 0.004, and RPE close to 4.5%. These values indicate high predictive accuracy, considering that only backfat thickness was used as an input variable. Similar studies reported by Preethi et al. (2023), using artificial neural networks with morphometric variables in Landrace piglets at three life stages (4th, 6th, and 8th week), reported determination coefficients (R^2) between 0.85 to 0.91 and MSE between 0.44 to 1.27. Meanwhile, Rahman et al. (2023), employing deep neural networks with multiple variables (BFT, parity, and sow age), reported RMSE values between 0.05 and 0.07 kg when estimating average litter weight. Despite not including multiple predictors, it was evidenced that BFT alone allows achieving precision levels comparable to multivariate studies, highlighting its utility as a biometric variable for predicting birth weight

under conditions of limited data availability. This finding underscores the potential of BFT as a robust predictive indicator in swine production systems with limited resources for information collection.

For LBP and total TPB, the models achieved mean absolute errors below 1.6 piglets, RPE between 13.4% and 13.7%, and RMSE between 1.8 to 2.0 piglets (**Table 3**). Although these metrics indicated lower precision compared to that obtained for average weight, the results remain adequate within the biological context and allow identification of general reproductive patterns useful for swine management. Compared to previous multivariate studies, such as Kousenidis et al. (2022), who used deep neural networks with ultrasonographic variables, including ovarian volume and follicle count in pregnant sows, to predict litter size, they achieved only an MAE of 2.58 piglets. These results demonstrate that, despite their model integrating multiple high-resolution physiological variables, the error margin exceeded that observed in the present study, highlighting the relative predictive value of BFT even as a single predictor. This suggests that, although BFT does not substitute complete multivariate characterization, it can constitute a useful and accessible tool in swine production contexts with limited resources for intensive physiological monitoring. The ability to obtain reasonably accurate predictions using a single easily measurable variable represents a significant advantage for production systems with technological or economic limitations.

PMR resulted in poor performance across all fitting criteria, with RPE values between 92% and 93%, indicating very low predictive capacity. According to Tapie et al. (2024), an RPE above 20% reflects low model precision, confirming the inadequacy of this univariate approach for predicting mortality. The MAE, which considers all individual differences with equal weight without distinguishing whether the model overestimates or underestimates actual values (Rosero-Noguera et al., 2022), indicated that for PMR, the model deviates on average 3.6 percentage points from actual values. This limited predictive capacity is attributed to the high variability of the variable (CV = 92%) and the restricted ability of BFT to explain a multifactorial phenomenon such as neonatal mortality. Similar results have been reported with Bayesian networks, where it was evidenced that BFT contributed less than 0.5% to the explained variance in mortality prediction, while factors such as parity, health status, and reproductive history of the sow were determinant (Teixeira et al., 2024). Nevertheless, a multivariate approach could significantly improve PMR predictions. Magalhães et al. (2024), using an RF model with multiple biological, environmental, and productive variables, achieved 90% precision in classifying mortality risk groups. These results suggest that PMR prediction requires a multivariate approach that integrates physiological, environmental, and management factors, beyond a single biometric measure such as BFT.

When jointly analyzing the fitting criteria of DNN compared to RF, equivalent performance was evidenced between both approaches, which is attributed to the simplicity of the prediction problem characterized by a single input variable and low complexity inherent to the data structure. Previous studies have indicated the tendency toward convergence of predictive model performance under low-dimensional conditions, where the advantage of more complex algorithms is limited by the simplicity of the feature space (Díaz-Uriarte and Alvarez de Andrés, 2006; Kong and Yu, 2018; Liu et al., 2013; May et al., 2011). These results indicate that algorithmic sophistication does not necessarily translate into substantial improvements in predictive performance in systems with low data dimensionality, confirming that modeling complexity should be appropriately adjusted to the inherent complexity of the problem. This observation has important practical implications, suggesting that in contexts of limited resources or when interpretability is required, simpler models like RF may be preferable without sacrificing predictive accuracy, while DNN could be reserved for problems that truly require modeling complex nonlinear relationships.

Table 3. Model fit criteria and predictive performance

Metrics	LBP		TPB		APW (Kg)		PMR (%)	
	DNN	RF	DNN	RF	DNN	RF	DNN	RF
MAE	1.513	1.509	1.608	1.609	0.038	0.035	3.707	3.648
MSE	3.621	3.525	4.024	3.966	0.004	0.004	17.194	16.889
RMSE	1.903	1.878	2.006	1.992	0.063	0.063	4.147	4.110
RPE	13.586	13.406	13.741	13.642	4.533	4.539	93.034	92.207

DNN = Deep Neural Networks; RF = Random Forest; MAE = Mean Absolute Error; MSE = Mean Squared Error; RMSE = Root Mean Squared Error; RPE = Relative Prediction Error.

Comparison of means between observed and predicted values revealed that both models showed similarity in their predictions, demonstrating that both DNN and RF adequately captured the central tendencies of the studied variables (**Table 4**). However, standard deviations of predicted values were consistently lower than those of observed values, particularly in the mortality variable. This observation suggests that the models, despite their capacity to predict average values with precision, tend to underestimate the real variability inherent in the data. This behavior is consistent with results from other studies on predictive models and their inherent difficulty in modeling stochastic events and complex interactions between biological variables (Baxter et al., 2012; Jakubovitz et al., 2019; Zhang et al., 2021). Notably, the RF model showed slightly higher variability in predictions compared to DNN, which could be an

intrinsic characteristic of ensemble models that integrate multiple estimators (Bai, 2017; Louppe, 2014). This greater dispersion in RF predictions suggests a slightly superior capacity to capture the natural variability of the data and, consequently, a potentially more representative fit of the underlying biological reality.

Table 4. Means and standard deviations of observed and predicted values

Model	LBP		TPB		APW (kg)		PMR (%)	
	Oi	Pi	Oi	Pi	Oi	Pi	Oi	Pi
DNN	14±1.90	14.11±0.1	14.60±2.0	14.61±0.24	1.39±0.06	1.39±0.006	4.46±4.16	4.56±0.31
RF	14±1.90	13.99±0.27	14.60±2.0	14.56±0.21	1.39±0.06	1.39±0.01	4.46±4.16	4.35±0.55

Oi = Observed values; Pi = Predicted values; DNN = Deep Neural Networks; RF = Random Forest.

Model implications

The developed predictive models, although demonstrating promising capacity to estimate LBP, TPB, and APW variables from a single predictor (BFT), faced limitations inherent to simplifying the multifactorial complexity of swine production systems. While their precision in these variables was valuable, their inability to adequately model high variability in phenomena influenced by multiple interactive factors highlighted a significant limitation, which could lead to underestimation of uncertainty associated with productive decision-making. The similarity in performance between RF and DNN can be interpreted as a consequence of the inherent simplicity of the univariate prediction problem, rather than as a limitation of the employed methodologies. This highlights the importance of carefully considering the complexity of the biological problem when selecting the most appropriate modeling methodology. Nevertheless, RF and DNN models can be useful for identifying general reproductive trends in resource-limited contexts. Their capacity to generalize and learn patterns from historical events, even with a single predictor variable, offers an accessible alternative for basic swine production management. These models represent a promising starting point for reproductive process optimization and operational cost reduction, especially in production systems where extensive data collection is limited by technological or economic constraints.

CONCLUSION

The RF model presented the best fit in predicting LBP, TPB, and APW variables, showing slightly superior performance to DNN. However, the main limitation of both models was evidenced in PMR prediction, where high relative errors were recorded (RPE of 93.03% for DNN and 92.20% for RF). This result highlights the need to incorporate additional variables of genetic, environmental, health, and management nature, whose influence has been documented as determinant in neonatal survival. Piglet mortality constitutes a highly multifactorial phenomenon, characterized by complex interactions and stochastic components, which significantly hinders its modeling through simplified approaches based on a single predictor variable such as BFT. The obtained results suggest that BFT, as a single biometric variable, provides valuable information for predicting basic reproductive variables, representing an accessible tool in resource-limited contexts. However, for complex phenomena such as neonatal mortality, multivariate approaches that integrate multiple physiological, environmental, and management factors are required to achieve clinically useful predictions.

DECLARATIONS

Corresponding Author

Correspondence and requests for materials should be addressed to; E-mail: watapiec@unal.edu.co; ORCID: https://orcid.org/0000-0001-7064-6753

Authors' Contribution

W.A. Tapie: Conceptualization, data curation, formal analysis, methodology, writing – original draft, writing – review and editing, investigation, project administration, resources, supervision, and validation. C.L Guerra-Marín: Conceptualization, data curation, formal analysis, methodology, writing – review and editing, funding acquisition. J.F Manrique-Hincapie: Conceptualization, data curation, formal analysis, methodology, writing – original draft, writing – review and editing, investigation, project administration, software, supervision, and validation. S Montoya-Uribe: Conceptualization, data curation, methodology, visualization, writing – review and editing, validation.

Funding

The experiment was financed by Universidad Católica de Oriente.

Conflict of interest

The authors have not declared any conflicts of interest.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

A la Facultad de Ciencias Agropecuarias de la Universidad Católica de Oriente.

REFERENCES

Ajay, A., Chauhan, A., Vaishnav, S., Rani, C., Kumar, B., De, U., Verma, M. R., Singh, M., and Gaur, G. (2023). Impact of body condition on sow and litter performance, postpartum physiological, hematological, and biochemical parameters in Landlly crossbred pigs. *Tropical Animal Health and Production*, *55*(6), 1–13. https://doi.org/10.1007/S11250-023-03772-4/METRICS

Andersen, I. L., Tajet, G. M., Haukvik, I. A., Kongsrud, S., and Bøe, K. E. (2007). Relationship between postnatal piglet mortality, environmental factors and management around farrowing in herds with loose-housed, lactating sows. *Acta Agriculturae Scandinavica A: Animal Sciences*, *57*(1), 38–45. https://doi.org/10.1080/09064700601159626

Bai, S. (2017). Growing random forest on deep convolutional neural networks for scene categorization. *Expert Systems with Applications*, *71*, 279–287. https://doi.org/10.1016/J.ESWA.2016.10.038

Baxter, E. M., Lawrence, A. B., and Edwards, S. A. (2012). Alternative farrowing accommodation: welfare and economic aspects of existing farrowing and lactation systems for pigs. *Animal*, *6*(1), 96–117. https://doi.org/10.1017/S1751731111001224

Bell, W., Urioste, J. I., Barlocco, N., Vadell, A., and Clariget, R. P. (2015). Genetic and environmental factors affecting reproductive traits in sows in an outdoor production system. *Livestock Science*, *182*, 101–107. https://doi.org/10.1016/J.LIVSCI.2015.10.025

Díaz, J. A. C., Nikkilä, M. T., and Stalder, K. (2015). Piglet mortality: causes and prevention. In *The Gestating and Lactating Sow*. Wageningen Academic. https://doi.org/10.3920/978-90-8686-803-2_19

Díaz-Uriarte, R., and Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, *7*(1), 1–13. https://doi.org/10.1186/1471-2105-7-3/FIGURES/1

Gauthier, R., Largouët, C., and Dourmad, J. Y. (2022). Prediction of litter performance in lactating sows using machine learning, for precision livestock farming. *Computers and Electronics in Agriculture*, *196*, 106876. https://doi.org/10.1016/J.COMPAG.2022.106876

Ha, S. H., Choi, Y. H., Mun, J. Y., Park, S. R., Kinara, E., Park, H. J., Hong, J. S., Kim, Y. M., and Kim, J. S. (2024). Correlation between reproductive performance and sow body weight change during gestation. *Journal of Animal Science and Technology*, *66*(3), 543. https://doi.org/10.5187/JAST.2023.E105

Hu, J., and Yan, P. (2022). Effects of Backfat Thickness on Oxidative Stress and Inflammation of Placenta in Large White Pigs. *Veterinary Sciences*, *9*(6), 302. https://doi.org/10.3390/vetsci9060302

Jakubovitz, D., Giryes, R., and Rodrigues, M. R. D. (2019). Generalization Error in Deep Learning. *Applied and Numerical Harmonic Analysis*, 153–193. https://doi.org/10.1007/978-3-319-73074-5_5

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning*. https://doi.org/10.1007/978-1-0716-1418-1

Kinane, O., Butler, F., and O'driscoll, K. (2021). Freedom to Grow: Improving Sow Welfare also Benefits Piglets. *Animals 2021, Vol. 11, Page 1181*, *11*(4), 1181. https://doi.org/10.3390/ANI11041181

Kingma, D. P., and Ba, J. L. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://arxiv.org/abs/1412.6980v9

Ko, E., Jeong, K., Oh, H., Park, Y., Choi, J., and Lee, E. (2023). A deep learning-based framework for predicting pork preference. *Current Research in Food Science*, *6*. https://doi.org/10.1016/J.CRFS.2023.100495

Kong, Y., and Yu, T. (2018). A Deep Neural Network Model using Random Forest to Extract Feature Representation for Gene Expression Data Classification. *Scientific Reports 2018 8:1*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-34833-6

Kousenidis, K., Kirtsanis, G., Karageorgiou, E., and Tsiokos, D. (2022). Evaluation of a Numerical, Real-Time Ultrasound Imaging Model for the Prediction of Litter Size in Pregnant Sows, with Machine Learning. *Animals 2022, Vol. 12, Page 1948*, *12*(15), 1948. https://doi.org/10.3390/ANI12151948

Liu, M., Wang, M., Wang, J., and Li, D. (2013). Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical*, *177*, 970–980. https://doi.org/10.1016/J.SNB.2012.11.071

Louppe, G. (2014). *Understanding Random Forests: From Theory to Practice*. https://arxiv.org/abs/1407.7502v3

Magalhães, E. S., Zhang, D., Moura, C. A. A., Trevisan, G., Holtkamp, D. J., López, W. A., Wang, C., Linhares, D. C. L., and Silva, G. S. (2024). Development of a pig wean-quality score using machine-learning algorithms to characterize and classify groups with high mortality risk under field conditions. *Preventive Veterinary Medicine*, *232*, 106327. https://doi.org/10.1016/J.PREVETMED.2024.106327

Magalhães, E. S., Zhang, D., Wang, C., Thomas, P., Moura, C. A. A., Trevisan, G., Holtkamp, D. J., Rademacher, C., Silva, G. S., and Linhares, D. C. L. (2023). Comparing forecasting models for predicting nursery mortality under field conditions using regression and machine learning algorithms. *Smart Agricultural Technology*, *5*, 100280. https://doi.org/10.1016/J.ATECH.2023.100280

Magdalena, S. N., Robert, E., Tadeusz, B., Mirosław, T., and Artur, P. (2015). Prediction of carcass meat percentage in young pigs using linear regression models and artificial neural networks. *Annals of Animal Science*, *2015*, 551–564. https://doi.org/10.1515/AOAS-2015-0069

May, R., Dandy, G., Maier, H., May, R., Dandy, G., and Maier, H. (2011). Review of Input Variable Selection Methods for Artificial Neural Networks. *Artificial Neural Networks - Methodological Advances and Biomedical Applications*. https://doi.org/10.5772/16004

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2012). *Scikit-learn: Machine Learning in Python*. http://arxiv.org/abs/1201.0490

Preethi, A. L., Tarafdar, A., Ahmad, S. F., Panda, S., Tamilarasan, K., Ruchay, A., and Gaur, G. K. (2023). Weight Prediction of Landlly Pigs from Morphometric Traits in Different Age Classes Using ANN and Non-Linear Regression Models. *Agriculture 2023, Vol. 13, Page 362*, *13*(2), 362. https://doi.org/10.3390/AGRICULTURE13020362

Rahman, M. T., Brown-Brandl, T. M., Rohrer, G. A., Sharma, S. R., Manthena, V., and Shi, Y. (2023). Statistical and machine learning approaches to describe factors affecting preweaning mortality of piglets. *Translational Animal Science*, *7*(1). https://doi.org/10.1093/TAS/TXAD117

Roongsitthichai, A., and Tummaruk, P. (2014). Importance of Backfat Thickness to Reproductive Performance in Female Pigs. *The Thai Journal of Veterinary Medicine*, *44*(2), 171–178. https://doi.org/10.56808/2985-1130.2557

Rosero-Noguera, R., Bedoya-Mazo, S., Posada-Ochoa, S. L., Rosero-Noguera, R., Bedoya-Mazo, S., and Posada-Ochoa, S. L. (2022). Predicción del consumo de materia seca de forraje en vacas lecheras mediante el uso de acelerómetros. *Información Tecnológica*, *33*(4), 63–72. https://doi.org/10.4067/S0718-07642022000400063

Tapie, W. A., Posada-Ochoa, S. L., Rosero-Noguera, J. R., and Muñoz-Tamayo, R. (2024). Desarrollo de un modelo dinámico mecanicista para predecir el crecimiento de cuyes (Cavia porcellus) machos del genotipo Perú. *Revista de La Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, *48*(189), 859–870. https://doi.org/10.18257/RACCEFYN.2997

Taylor, C., Guy, J., and Bacardit, J. (2023). Estimating individual-level pig growth trajectories from group-level weight time series using machine learning. *Computers and Electronics in Agriculture*, *208*, 107790. https://doi.org/10.1016/J.COMPAG.2023.107790

Teixeira Costa, C., Boulbria, G., Dutertre, C., Chevance, C., Nicolazo, T., Normand, V., Jeusselin, J., and Lebret, A. (2024). Estimating the individual stillborn rate from easy-to-collect sow data on farm: an application of the bayesian network model. *Porcine Health Management*, *10*(1), 1–9. https://doi.org/10.1186/S40813-024-00395-5/TABLES/5

Witten, I., Frank, E., Hall, M., and Christopher, P. (2016). Chapter 4: Clustering. *Data Mining : Practical Machine Learning Tools and Techniques*, 141–201.

Yun, J., Han, T., Björkman, S., Nystén, M., Hasan, S., Valros, A., Oliviero, C., Kim, Y., and Peltoniemi, O. (2019). Factors affecting piglet mortality during the first 24 h after the onset of parturition in large litters: Effects of farrowing housing on behaviour of postpartum sows. *Animal*, *13*(5), 1045–1053. https://doi.org/10.1017/S1751731118002549

Zhang, C., Bengio, S., Hardt, M., Recht, B., and Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, *64*(3), 107–115. https://doi.org/10.1145/3446776

Zhou, Y., Xu, T., Cai, A., Wu, Y., Wei, H., Jiang, S., and Peng, J. (2018). Excessive backfat of sows at 109 d of gestation induces lipotoxic placental environment and is associated with declining reproductive performance. *Journal of Animal Science*, *96*(1), 250–257. https://doi.org/10.1093/JAS/SKX041