Biosystems Engineering 263 (2026) 104381

Contents lists available at ScienceDirect

Biosystems Engineering

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Improved and interpretable accelerometer-based farrowing prediction

Elisabeth Mayrhuber[a,b,*], Kristina Maschat[d], David Brunner[a,b], Stephan M. Winkler[a], Maciej Oczak[c,d]

[a] Bioinformatics Research Group, PLFDoc, University of Applied Sciences Upper, Hagenberg, Austria
[b] Computer Vision Lab, TU Wien, Vienna, Austria
[c] Precision Livestock Farming Hub, University of Veterinary Medicine Vienna, Vienna, Austria
[d] Centre for Animal Nutrition and Welfare, University of Veterinary Medicine, Vienna, Vienna, Austria

ARTICLE INFO

Keywords:
Explainable artificial intelligence
Precision livestock farming
Symbolic regression

ABSTRACT

Predicting the onset of farrowing in sows is critical for improving animal welfare and optimising farm management. Methods driven by explainable artificial intelligence for detecting nest-building behaviour and predicting time to farrowing using accelerometer data from ear tags are presented. These methods are evaluated on a dataset containing farm management data and accelerometer data of 179 sows. During data collection the animals were kept in three different pen types with the possibility of temporary crating. By combining acceleration metrics with prepartum examinations and farm management data, a two-stage model was developed that first detects the onset of nest-building and subsequently predicted the remaining time until farrowing. Various methods, including cumulative sum (CUSUM), Bayesian estimation of abrupt change, seasonality, and trend (BEAST), and a custom model (NestDetect), were compared for nest-building detection, while symbolic regression and deep learning were used to predict farrowing time. For 82.6 % of the sows, it was possible to detect the start of nest-building behaviour in a 48-h window before the onset of farrowing. When nest-building was detected correctly, symbolic regression was able to predict the remaining time to farrowing with a mean absolute error of 9.4 h and delivered interpretable results, while NNs achieved a mean absolute error of 9.6 h without being inherently interpretable. This work emphasises the importance of model interpretability and explainability in precision livestock farming, highlighting that transparent models can facilitate timely, data-driven interventions, while having the same prediction power as non-interpretable models.

1. Science4Impact statement (S4IS)

This study presents a novel method for detecting nest-building behaviour in sows and predicting farrowing time using acceleration data from ear tags, integrated with prepartum examination and stable management data. The findings provide actionable insights that can enhance decision-making processes in the livestock industry, particularly for optimising farrowing management and improving animal welfare. In the future, the developed models can be applied in real-time on farms, facilitating timely interventions that reduce sow stress and piglet crushing events. This research supports farm staff by offering scientifically validated, technology-based solutions for improving the efficiency and welfare outcomes in swine production systems, aligning with broader societal goals of sustainable and ethical livestock management.

Nomenclature table

ANN	Artificial neural network
BEAST	Bayesian estimator of abrupt change, seasonality, and trend
CUSUM	Cumulative sum
DL	Deep learning
GP	Genetic programming
LIME	Local interpretable model-agnostic explanations
M	Magnitude of acceleration
MAE	Mean absolute error
Milli-g	Acceleration unit equal to 0.00981 m s^{-2}
MLP	Multi-layer perceptron
MSE	Mean squared error

(*continued on next page*)

This article is part of a special issue entitled: Technologies in Livestock published in Biosystems Engineering.

* Corresponding author. Bioinformatics Research Group, PLFDoc, University of Applied Sciences Upper, Hagenberg, Austria.
E-mail address: Elisabeth.Mayrhuber@fh-hagenberg.at (E. Mayrhuber).

https://doi.org/10.1016/j.biosystemseng.2025.104381
Received 20 December 2024; Received in revised form 23 December 2025; Accepted 29 December 2025
Available online 1 January 2026

(*continued*)

R^2	Coefficient of determination
PLF	Precision livestock farming
SHAP	Shapley additive explanations
SR	Symbolic regression
XAI	Explainable artificial intelligence

2. Introduction

In modern livestock production, ensuring the health and welfare of animals is paramount, not only for ethical reasons but also for optimising productivity, sustainability and economic returns (Larsen et al., 2021; Neethirajan, 2023). Among the various critical aspects of livestock management, monitoring the reproductive behaviour of sows is of significant importance (Morgan et al., 2021; Pastell et al., 2016; Van Erp-; van Der Kooij et al., 2023; Vargovic et al., 2022). The behaviour of sows during the prepartum period, particularly nest-building, serves as a crucial indicator of imminent farrowing (Oczak et al., 2015; Oczak, Maschat, & Baumgartner, 2019). Detecting the onset of nest-building behaviour and accurately predicting the time until farrowing can lead to better management practices, reducing stress for the animals and allowing for timely intervention by farm personnel.

When it comes to using artificial intelligence (AI) systems for predicting behaviour, it is crucial for these systems to be trustworthy in order to be useful for farmers and veterinarians (Gardezi et al., 2024). Gardezi et al. (2024) identified key issues, such as model transparency, accountability, fairness, and data privacy, as being central to fostering trust between farmers and AI providers. Elliott and Werkheiser (2023) highlighted the challenges posed by the complexity of precision livestock farming (PLF) systems and the potential biases in machine learning algorithms, favouring strategies that included stakeholder involvement and the use of explainable artificial intelligence (XAI) to enhance transparency.

In the field of pig production, a lot of work has been done in the area of detecting the onset of farrowing in different environments (Thompson et al., 2019) in order to improve animal welfare, e.g. by enabling nest-building behaviour while preventing early crating of sows and avoiding piglet crushing events (Cornou & Lundbye-Christensen, 2012; Pastell et al., 2016, 2016; Traulsen et al., 2018; Van Erp-van Der Kooij et al., 2023). In order to do so, a variety of different technologies have been used, such as sound cameras (Van Erp-van Der Kooij et al., 2023), accelerometers (Cornou & Lundbye-Christensen, 2012; Oczak, Maschat, & Baumgartner, 2019; Oliviero et al., 2008; Pastell et al., 2016; Traulsen et al., 2018) and computer vision (Oczak et al., 2023).

Different approaches have been researched over time to predict the onset of farrowing. Using a photocell, Oliviero et al. (2008) were able to show that during the last 24 h before farrowing, sows stand significantly longer than before the onset of nest-building. The activity is also significantly higher in this last 24 h time window, compared to previous days, as measured by a force sensor on the floor of the pen.

The diurnal activity pattern was modelled in previous research on farrowing (Cornou & Lundbye-Christensen, 2012) using threshold based models such as dynamic generalised linear models and cumulative sum (CUSUM) of hourly differences of activity. This method achieved good performance on a small dataset and was used as groundwork for other activity based farrowing prediction models.

Pastell et al. (2016) used 3-dimensional accelerometers placed in neck collars of sows to monitor activity levels before farrowing. CUSUM charts were used to detect significant change in the activity levels of the sows related to the onset of farrowing. Similarly, Traulsen et al. (2018) also used CUSUM charts to detect the onset of farrowing based on acceleration data but in their work the acceleration data was collected using SMARTBOW® ear tags (Smartbow/Zoetis LLC, Weibern, Austria).

The use of acceleration data to predict parturition has not only been investigated in pigs but also used to predict calving (Krieger et al., 2019). Krieger et al. (2019) predicted calving on the basis of acceleration features, such as rumination time, activity, lying time of individual animals, in relation to other animals. Krieger et al. (2019) applied SMARTBOW® sensors to collect these acceleration-based features to predict calving with transfer function models.

Recent studies have also outlined the potential of deep learning (DL) applications in pig production (Bao et al., 2024), but one crucial point in every support system for humans is the need for explainability and trustworthiness (Elliott & Werkheiser, 2023; Gardezi et al., 2024).

Despite progress in farrowing prediction, existing approaches face two major limitations: (i) changepoint detection methods such as CUSUM often lack robustness across diverse housing conditions, and (ii) time-to-farrowing prediction models are typically based on black-box algorithms, which hinder trust and practical adoption in farm settings. Furthermore, no interpretable model currently exists that combines accelerometer data with health and management information to predict the remaining time until farrowing.

The research presented in this paper addresses these gaps by developing a two-stage prediction framework that first detects the onset of nest-building behaviour and then estimates the remaining time to farrowing. The hypothesis is that XAI methods, specifically symbolic regression (SR), can deliver predictive performance comparable to deep learning models while providing inherent interpretability. By integrating accelerometer data with prepartum examination and management features, the proposed approach aims to improve prediction accuracy and transparency, enabling timely interventions that enhance sow welfare and farm efficiency.

3. Materials and methods

The presented work is split into two main parts: (i) detection of nest-building behaviour and (ii) time to farrowing prediction. An online test setting simulating the real time processing of data on a farm, using sliding windows, is designed to evaluate the developed models (Fig. 1). Both steps, nest-building detection and farrowing prediction, are based on a sliding window approach. In the initial phase, nest-building detection is performed at 3-h intervals to efficiently monitor general activity patterns. Once nest-building is detected, the interval is reduced to 1.5 h, as this marks the onset of the critical pre-farrowing phase. From this point on, more frequent updates are desirable to support timely decision-making in a practical farm setting. These interval values (3 h and 1.5 h) are not fixed and can be easily adjusted by modifying the configuration of the test setup. All implementation steps, including data preprocessing, feature engineering, model development, and evaluation, are performed in a Python 3.12 environment using standard scientific libraries (e.g., NumPy, pandas, scikit-learn) and specialised packages for symbolic regression and hyperparameter optimisation (Ec-KiTy and Optuna, respectively).

3.1. Animals and housing

Data collection was carried out on a teaching and research farm in Lower Austria from June 2014 to May 2016 within project Pro-SAU (see Heidinger et al. (2022) for further details). The aim of this project is to investigate the effects of different confinement durations and pen types on piglet mortality, sow and piglet welfare, economic efficiency, and occupational safety.

Experimental animals are Large White (n = 148) and Landrace × Large White crossbred pigs (n = 31). The age of the sows ranges from first to seventh parity, with 28 gilts and 151 older sows included in the study. Batch farrowing with all-in/all-out management is applied. Sows are moved to the farrowing compartment approximately 5 days before parturition and confined temporarily according to an experimental protocol (Maschat et al., 2020). On average, sows are crated for 3.7 (±2.5) days with the duration of peripartal confinement ranging from

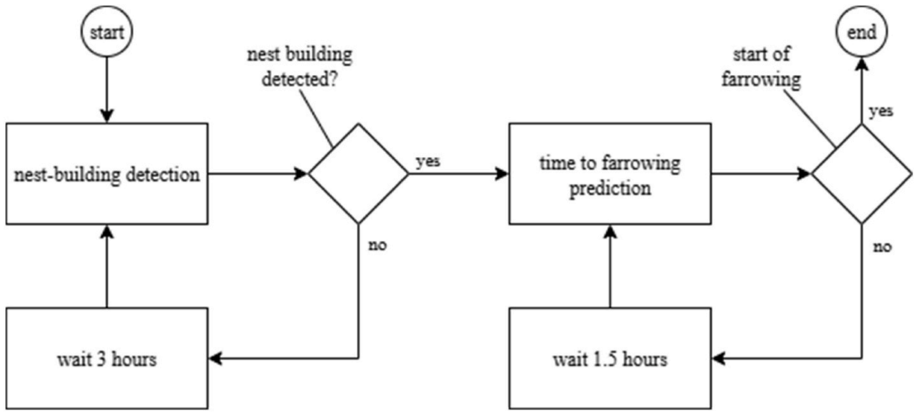

Fig. 1. The first step is the detection of nest-building where a new detection is done every 3 h until nest-building behaviour is detected. After nest-building behaviour is detected, then a time to farrowing prediction is done every 1.5 h until the actual beginning of farrowing.

0 to 8 days. Induction of farrowing is allowed only after 116 days of pregnancy. Cross-fostering is carried out 12–36 h postpartum. The weaning age of the piglets is four weeks. Sows are fed commercial feed according to a farm-specific feeding regime. Sows, as well as piglets, have *ad libitum* water access. Hay is offered as enrichment and nesting material in a rack and is available throughout the sows' stay in the farrowing pens. Racks are half filled in the morning and whenever they are empty. The farrowing compartment is equipped with an automatic ventilation system and average room temperature is kept at 22 °C. Health-related measures of welfare are recorded when sows are introduced to the pens and in weeks 1, 3 and 4 after parturition.

To evaluate whether seasonal factors influence sow behaviour, we analyse the duration between the onset of nest-building and farrowing across four seasons (spring, summer, autumn, winter). A one-way ANOVA is performed using Python's scipy.stats library to test for differences among seasonal groups. Statistical significance was set at $p < 0.05$.

3.2. Accelerometer data

Each sow was equipped with a SMARTBOW® ear tag containing an accelerometer sensor, through which the sow's movements along three axes (x, y, and z) are captured. Acceleration data is recorded with a frequency of 10 Hz. To measure the overall activity of a sow the magnitude of acceleration (M) is calculated as the square root of the sum of squared values from each axis ($M = \sqrt{x^2 + y^2 + z^2}$).

The data are split into a training and test sets with 143 sows (80 %) and 36 sows (20 %), respectively (Table 1). This ensured an even distribution of data from different pen types (SWAP, trapezoid, and wing) across the test set. Evaluating the model's performance on the unseen test set allows to assess its generalisability to new data.

A descriptive analysis is performed on the available activity data. The activity values, originally recorded in g, were converted m s^{-2}, where 1 g corresponds to 9.81 m s^{-2}. These values represent acceleration-based measurements used as indicators of the animals' movement intensity over time.

To characterise general activity patterns, we calculate summary statistics for acceleration magnitude across the entire dataset, including mean, standard deviation, and maximum values. Additionally, activity is analysed separately for daytime and nighttime intervals to capture diurnal variation. These descriptive analyses are performed using Python (NumPy and pandas) and serve to inform feature engineering and model development.

Table 1
Number of sows for which the data is used for the training and test split based on pen types.

Pen type	Training	Test
Wing	47	12
SWAP	44	12
Trapezoid	52	12
Total	**143**	**36**

3.3. Data preprocessing and feature engineering

For the detection of nest-building behaviour only the accelerometer data are used. Since noise in the data impedes detection on the raw dataset, rolling averages (12 h and 24 h) are used for all three tested methods for nest-building detection. The 12 h rolling averages show the day and night cycle of a sow, while the 24 h rolling average can identify major changes in the activity level, e.g. those at the onset of nest-building behaviour. For the time to farrowing prediction, the accelerometer data is combined with the provided examination data.

Regarding the acceleration data for the farrowing prediction, six time-related features are extracted every hour in the last 48 h before the onset of farrowing. These six features can be found in Table 2 as the activity related features for farrowing prediction. This results in a total of 48 instances for every sow with a total of 80 features (examples can be found in Table 2). The resulting training dataset contains 6864 instances. For each instance, the remaining time to farrowing in minutes is

Table 2
Features used for the detection of nest-building behaviour and farrowing prediction split by the data category.

Prediction target	Data category	Features
nest-building detection	activity	average during last 24 h average over 24 h without nest-building average during last 12 h average over 12 h without nest-building
farrowing prediction	activity	average during daytime average during nighttime average over 24 h without nest-building average over 48 h without nest-building average during last 12 h average during last 48 h
	farm management	pen type time since start of crating 4 additional features
	examination	litter number body condition score lameness 65 additional features

Table 3
Number of correct and incorrect nest-building detections for 24 h and 48 h before farrowing. The results are shown for each method and possible pen type. Throughout all pen types NestDetect achieves the best results (shown in bold).

Method	Pen Type	24 h window			48 h window		
		Early	On time	Late	Early	On time	Late
CUSUM	Wing	25	32	2	22	35	2
	SWAP	20	33	3	14	39	3
	Trapezoid	23	41	3	15	46	3
	All	68	103	8	51	120	8
BEAST	Wing	22	19	18	12	26	21
	SWAP	17	15	24	3	33	20
	Trapezoid	24	19	21	10	29	25
	All	63	53	63	25	88	66
NestDetect	Wing	11	**45**	3	8	**48**	3
	SWAP	9	**42**	5	4	**47**	5
	Trapezoid	12	**46**	6	5	**53**	6
	All	32	**133**	14	17	**148**	14

calculated and used as a target variable during model development. In previous studies only activity related features have been used for the prediction (Mayrhuber et al., 2024).

Since the time of interest is before the onset of farrowing, all examinations done after the piglets are born are disregarded. All categorical features that don't have natural ordering are one hot encoded, a method that converts each category into a separate binary column (0 or 1) to avoid introducing artificial ordinal relationships. These features are the pen type and the sows' breed. After removing all columns where the majority of values are missing and removing all the features that are only available after the piglets are born, the dataset is mapped with the acceleration data.

3.4. Detection of nest-building behaviour

Three different methods are implemented and compared to detect the beginning of nest-building behaviour i.e. CUSUM, Bayesian estimator of abrupt change, seasonality, and trend (BEAST) and NestDetect, which is developed specifically for the research presented in this paper. The detection of nest-building behaviour is started two days after the sow is introduced to the farrowing pen, as activity levels are typically different immediately after introduction to a new pen due to environmental changes (Oczak, Maschat, & Baumgartner, 2019).

To assess the correctness of nest-building detection, predefined detection windows of 24 h and 48 h prior to the onset of farrowing are used. If an algorithm detects nest-building activity within one of these time windows, it is considered a correct detection. This approach is chosen due to the high variability and subjectivity in manually determining the exact onset of nest-building behaviour. Consequently, no manually annotated ground-truth labels for nest-building are used. Instead, the detection is evaluated based on whether the algorithm identified activity within the relevant pre-farrowing timeframes (Oczak, Maschat, & Baumgartner, 2019).

3.4.1. CUSUM

The CUSUM method (Page, 1954) is a well-established technique for identifying shifts in the mean of a sequential dataset. It is particularly suited for online or real-time applications, where data are processed as it arrives, allowing for prompt detection of changes. In the current study, an existing CUSUM implementation (Khamesi, 2022) is used. The cumulative sum of deviations from a reference value, typically the mean of the process under consideration, is monitored. When the cumulative sum exceeds a predetermined threshold, a changepoint is detected, indicating a significant shift in the process mean. CUSUM uses two one-sided statistics to detect shifts in a certain trend. These statistics are upward CUSUM (S_t^+, Eq. (1)) and downward CUSUM (S_t^-, Eq. (2)). The initial value for S_t^+ and S_t^+ is 0.

$$S_t^+ = \max\left(0, S_{t-1}^+ + \frac{x_t - \mu}{\sigma} - k\right) \tag{1}$$

$$S_t^- = \max\left(0, S_{t-1}^- - \frac{x_t - \mu}{\sigma} - k\right) \tag{2}$$

In Eqs. (1) and (2), S_t is the cumulative sum at time t and x_t is the observed value at time t. For runtime improvements, μ and σ can be updated sequentially without storing all previous occurrences for x_t. Equations (3) and (4) show the decision rules for CUSUM.

$$S_t^+ > h \tag{3}$$

$$S_t^- > h \tag{4}$$

The CUSUM control parameter k, which controls the sensitivity, is set to 2. Smaller values of k lead to a higher sensitivity towards smaller changes. Since nest-building behaviour is shown as a big change in activity, k is set to 2 to reduce false alarms. The threshold (th) for the decision rule is set to 150, since the change point is incorrectly detected with lower values. These parameters are identified by testing different variations on the training dataset. Therefore, k is tested with 0.1, 0.25, 0.5, 1, 2 and 5 and th is tested with 50, 100, 150 and 200.

In previous studies, CUSUM is used to detect the beginning of farrowing (Cornou & Lundbye-Christensen, 2012; Pastell et al., 2016; Traulsen et al., 2018). In the current study, changepoint detection is used to detect the beginning of nest-building behaviour and from there on to predict the remaining time until farrowing with XAI and DL methods.

3.4.2. BEAST

BEAST (Zhao et al., 2019) is a robust time-series decomposition algorithm designed to detect and analyse changes in satellite data, specifically focusing on abrupt changes, seasonality, and trends. It utilises Bayesian model averaging (Fragoso et al., 2018) to combine multiple competing models, which allows it to account for model uncertainty and avoid the pitfalls of relying on a single "best" model. BEAST applies a general linear model framework, with trend components modelled as piecewise linear functions and seasonal components as harmonic functions. The algorithm uses Markov Chain Monte Carlo sampling (Van Ravenzwaaij et al., 2018) to estimate the posterior distribution of model parameters, enabling it to identify changepoints, detect nonlinear trends, and quantify uncertainties. It is primarily designed for batch processing rather than online or real-time settings. BEAST employs computationally intensive algorithms that typically require the entire dataset to be available before analysis can be performed. This makes it less suitable for online settings where data arrives sequentially, and decisions need to be made in real-time. However, it was decided to test it in the current study, as promising results were shown in the literature (Zhao et al., 2019).

The BEAST model is set up with a periodicity of one day. The seasonality is periodic and the maximum number of allowed changepoints is set to one since only one changepoint should be detected to detect the onset of nest-building behaviour. The implementation was done using the python package published by the authors of the original BEAST publication (Zhao, 2025 & 2025).

3.4.3. NestDetect

This study proposes the NestDetect model, which works by processing the accelerometer data with a sliding window and detects the onset of nest-building behaviour based on two conditions. A nest-building behaviour alarm is triggered if either of the two specified conditions (Eqs. (5) and (6)) are evaluated as TRUE. In these equations, a_t represents the acceleration at time t, and a_{t-12} corresponds to the acceleration 12 h prior to t. The base dataset for these calculations is the rolling averages over 12 h (Section 2.3). The constant values of 0.8 and 1.75 were determined by using an evolution strategy, with these values yielding the best detection performance on the training dataset. Detection performance is measured as number of detected nest-building behaviours in the training set. The nest-building detection is correct if the alarm is triggered within 48 h before farrowing.

$$\mathrm{mean}([a_{t-24}; a_t]) \geq \left(\mathrm{mean}\left(\left[a_{t_0}; a_{t-24}\right]\right) + \mathrm{std}\left(\left[a_{t_0}; a_{t-24}\right]\right)\right) * 0.8 \tag{5}$$

$$\max\left([a_{t-12}; a_t]\right) > \max\left(\left[a_{t_0}; a_{t-12}\right]\right) * 1.75 \tag{6}$$

3.5. Time to farrowing prediction

For the time to farrowing prediction, two methods are used, namely SR and a ANN, specifically a multi-layer perceptron (MLP) model. Both methods model the relationship between acceleration, examination data and farm management information (Table 2). Furthermore, these methods are evaluated in terms of predictive power and interpretability of the resulting models.

For both, SR and MLP a hyperparameter optimisation is conducted using Optuna (Akiba et al., 2019, pp. 2623–2631). For a fitness function, the mean squared error (MSE, Eq. (7)) is used.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2 \tag{7}$$

For the evaluation of the predictive power of both models, the mean absolute error (MAE, Eq. (8)) and the coefficient of determination (R^2, Eq. (9)) are used. The MAE is used because it is easier to interpret than the MSE, but the MSE is more sensitive to high errors, which is why it is used during the model development. For Eqs. (7)–(9), y_i is the real time to farrowing of the instance i and p_i is the predicted value.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_i - p_i) \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - p_i)^2}{\sum_{i=1}^{n}(p_i - y_i)^2} \tag{9}$$

To further ensure robust model training, a grouped 5-fold cross-validation is performed (Fig. 2). This is used to avoid one sow's data being split into different groups (Refaeilzadeh et al., 2009).

To evaluate the performance of the time to farrowing prediction models, the online test setting is used (Fig. 1). To eliminate the uncertainty of the nest-building detection for the time to farrowing prediction additional tests are conducted in order to assess the performance of the whole pipeline as well as the time to farrowing prediction alone. To assess the performance of the time to farrowing prediction alone the same dataset is used, but for individuals, where the NestDetect method can not detect nest-building behaviour, human annotations are used.

3.5.1. Symbolic regression by genetic programming

Symbolic regression is a form of regression analysis that searches for mathematical expressions that best fit a given dataset. Unlike traditional regression methods, which assume a specific model form (e.g., linear or polynomial), SR does not start with a predefined model structure. Instead, it explores a vast space of possible mathematical expressions to find the one that best describes the relationship between input variables and the output (Kronberger et al., 2024).

To perform this search, SR often employs genetic programming (GP), an evolutionary algorithm inspired by the process of natural selection (Hu, 2023). GP iteratively evolves a population of candidate solutions, mathematical expressions in this case, through processes analogous to biological reproduction, mutation, and selection (see Fig. 3).

SR has several advantages over traditional regression methods.

- **Model Flexibility:** It does not require a predefined model structure, allowing it to discover novel and potentially more accurate representations of the data (Kronberger et al., 2024).
- **Interpretability:** The resulting models are often in the form of simple mathematical expressions, making them easier to interpret and analyse compared to complex machine learning models such as artificial neural networks (ANN) (Zhou & Hu, 2024).

In the context of predicting the time to farrowing in sows, SR by genetic programming is hypothesised to be particularly useful because it can uncover complex, nonlinear relationships between various input variables (e.g., acceleration data, prepartum examination data) and the target variable (time to farrowing). By evolving expressions that combine these inputs in meaningful ways, the model can generate accurate and interpretable predictions that can be directly applied to improve decision-making in livestock management.

A SR toolbox, named Ec-KiTy (Sipper et al., 2023) is used to train SR models. Based on the hyperparameter optimisation using Optuna, the following hyperparameters are applied to train the final models. The population size is set to 953 and the training is done in 184 generations with a crossover probability of 0.59 and a mutation probability of 0.41. The elitism rate is set to 0.027 and the full function set provided by Ec-KiTy is used except OR.

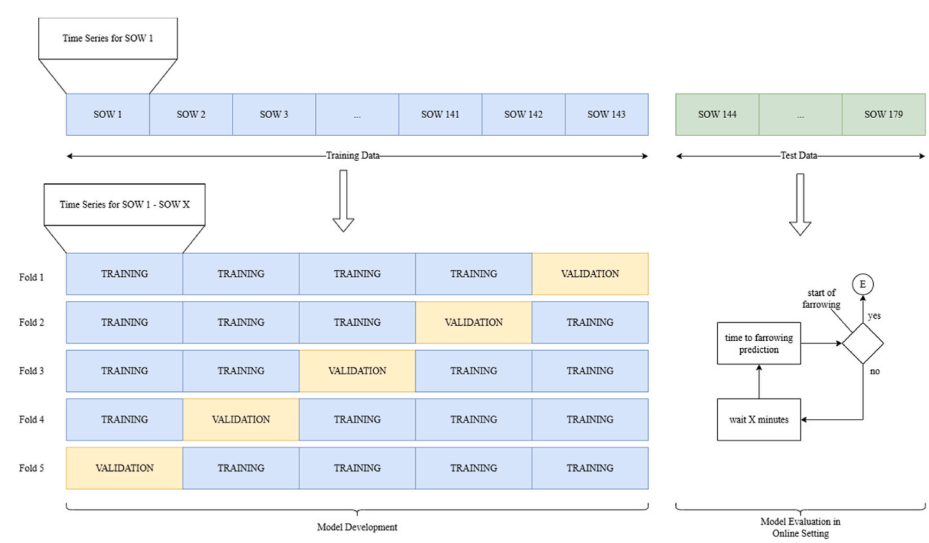

Fig. 2. Schematic illustrating the split of the data set using grouped 5-fold cross-validation for robust time to farrowing model development.

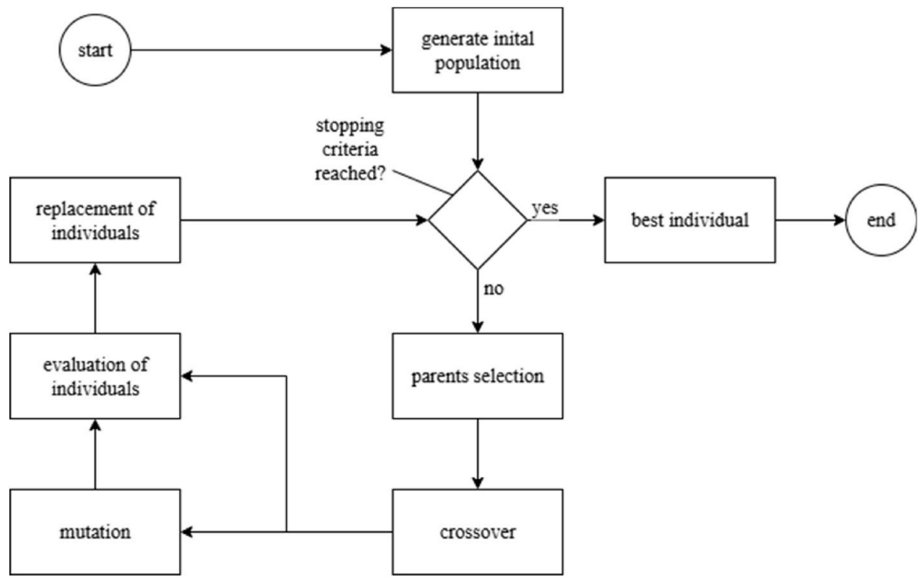

Fig. 3. Workflow of a typical genetic programming (GP) cycle.

3.5.2. Deep learning

Deep learning is a subset of machine learning that focuses on using ANN with multiple layers (hence "deep") to model complex patterns in data. It is particularly well-suited for tasks involving large datasets and high-dimensional data, where traditional machine learning methods may struggle to achieve similar performance (Goodfellow et al., 2023). DL has revolutionised many fields, including computer vision, natural language processing, and precision agriculture and livestock management (Bao et al., 2024).

ANNs are computational models inspired by the human brain and consist of layers of interconnected nodes (or "neurons"), where each connection has an associated weight. The network learns to perform tasks by adjusting these weights based on the data it is trained on. Due to the complex nature of neural networks, they are often called black-box, or closed box models (Fig. 5). This means that these models are not inherently interpretable and need post-hoc explainability methods to understand the models' internals and the predictions done by those models.

Similarly to the SR method, Optuna is used for the hyperparameter optimisation. Since the time to farrowing prediction is a regression task, the model can be kept small. Therefore, only one hidden layer is used in the models (Fig. 4).

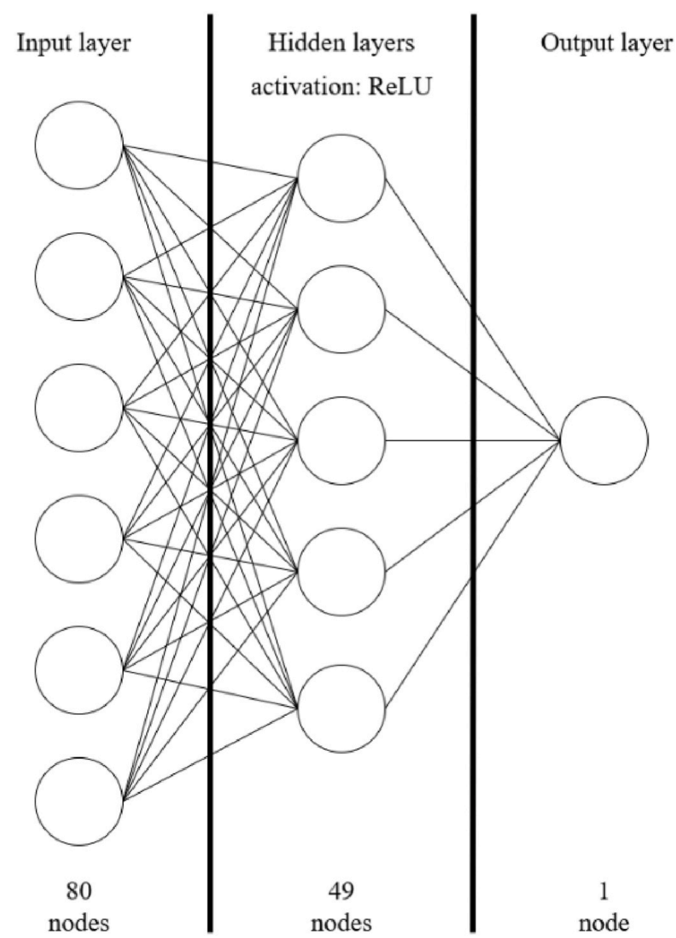

Fig. 4. Model architecture after hyperparameter optimisation using Optuna. The resulting model has 3 layers where the hidden layer has a node size of 49.

3.6. Interpretability versus explainability

In some discussions, the terms "explainability" and "interpretability" are often used interchangeably, as noted by Mei et al. (2023). However, in other works, including this one, these concepts are treated as distinct (Hu, 2023).

Interpretability refers to the ability to understand the inner workings of an open-box model by examining its structure, which allows a comprehension of why a model is making certain decisions. However, as the complexity of a model increases, this intrinsic interpretability can become limited. Some studies suggest that as model performance improves, there is often a trade-off with interpretability (Mei et al., 2023). To build trustworthy machine learning systems, it is essential to enhance model performance while simultaneously ensuring that model decisions are understandable (Petkovic, 2023).

Explainability, on the other hand, pertains to the ability to explain the decisions of closed-box models using techniques such as Shapley Additive exPlanations (SHAP; Lundberg & Lee, 2017) or Local Interpretable Model-Agnostic Explanations (LIME; Ribeiro et al., 2016). GP can also be employed to generate explanations for closed-box models. In this approach, GP approximates a model's decisions based on its outputs and then utilises the internal structure of the GP model to provide a global explanation for the closed-box model (Mei et al., 2023). The

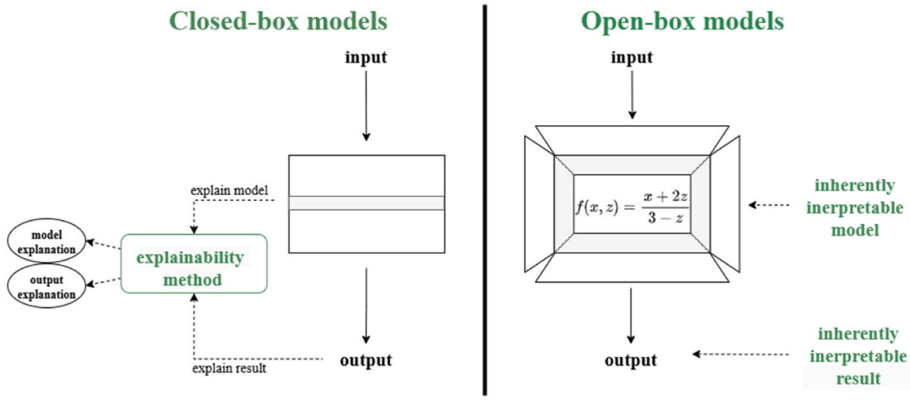

Fig. 5. When using closed-box models (left) explainability methods are needed to explain the model internals and results (explainability). On the other hand, open-box models (right) can be understood by looking at the internals of the model (interpretability).

distinction between explainability and interpretability is illustrated in Fig. 5.

Explainability can be classified into two types.

- **Global post-hoc explainability:** Involves explaining the structure of a closed-box model after training. This can include training transparent surrogate models on the closed-box model outputs (Evans et al., 2019; Mei et al., 2023). In order to more easily understand such surrogate models, partial dependence plots, that show feature importances using partial effects, can be used (Aldeia & De França, 2021).
- **Local post-hoc explainability**: Focuses on explaining individual predictions rather than the entire model. Beyond common methods, such as SHAP and LIME, evolutionary computation approaches and large language models (LLMs) are used to generate human-readable explanations. For example, GP4NLDR, which is an XAI dashboard, uses a combination of state-of-the-art GP and an LLM-powered chatbot to provide user-centred explanations for GP-based non-linear dimensionality reduction processes, enhancing model explainability (Maddigan et al., 2024).

In this research SHAP was utilised to visualize the internals of closed-box as well as open-box models with a high complexity. Although the internals of open-box models naturally do not need post-hoc explainability methods, since these models are inherently transparent, it was used in the current study to compare the differences between SR and ANN in predicting the remaining time to farrowing.

4. Results

The following section describes the results for both the detection of nest-building behaviour and the time to farrowing prediction. These results were obtained using the online test setting shown in Fig. 1.

Across the entire dataset, a mean activity of 0.415 m s^{-2} was observed, with a standard deviation of 0.077 m s^{-2}. The maximum recorded activity reached 5.383 m s^{-2}, while the corresponding standard deviation during peak activity episodes was found to be 1.188 m s^{-2}, indicating considerable variability in high-movement periods.

For a more detailed analysis, the data was divided into daytime and nighttime intervals. During daytime periods, a higher mean activity of 0.562 m s^{-2} was recorded (SD = 0.130 m s^{-2}), whereas during nighttime, the mean activity dropped to 0.268 m s^{-2} (SD = 0.047 m s^{-2}). This diurnal variation is consistent with typical circadian behaviour patterns in sows and highlights the importance of considering temporal structure in the modelling process.

To assess whether seasonal factors influence sow behaviour, we compare the duration between the onset of nest-building and farrowing across four seasons using a one-way ANOVA. No statistically significant differences was found ($p > 0.05$), indicating that moderate seasonal variation did not affect nest-building behaviour or farrowing timing. These findings support the robustness of the prediction models under typical on-farm conditions.

4.1. Detection of nest-building behaviour

The detection of nest-building behaviour was a true positive, if nest-building was detected within 48 h before the beginning of farrowing. The NestDetect model had the best detection performance with 82.68 % correctly detected instances followed by the CUSUM method with 67.04 %. The BEAST model achieved the worst performance with 49.16 % of correctly detected nest-building behaviours (Fig. 6).

An example for the nest-building detection based on changepoint detection with all three methods is shown in Fig. 7. The first subplot shows the nest-building detection with CUSUM. The red, vertical line shows the detected changepoint which was interpreted as the onset of nest-building. This detection with the CUSUM methods shows that the alarm was triggered rather late, shortly before the onset of farrowing. The changepoint detection performed by BEAST seemed similar to the one performed by the NestDetect model. Since BEAST is

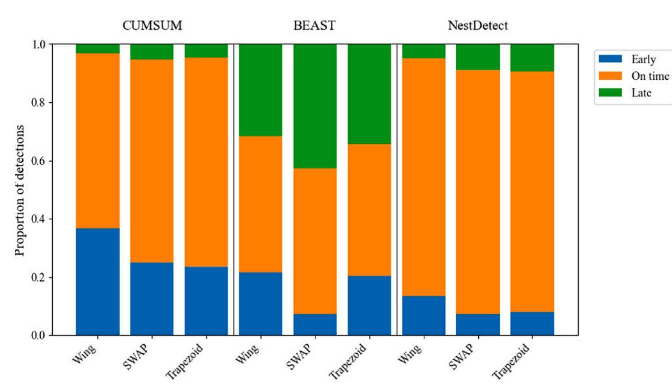

Fig. 6. Detection of nest-building behaviour in a 48-h window for the three methods (CUSUM, BEAST and NestDetect) using different farrowing pen types. Bars show the ratio of falsely and correctly detected nest-building behaviours as a percentage (the absolute numbers can be found in Table 3). Blue shows the ratio of too early detected nest-building behaviours for all used methods, while orange shows the ratio of correctly detected nest-building behaviour. The green bars show the ratio that is either detected too late or not at all, which is highest for BEAST. NestDetect shows the highest ratios of correctly detected nest-building behaviours. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

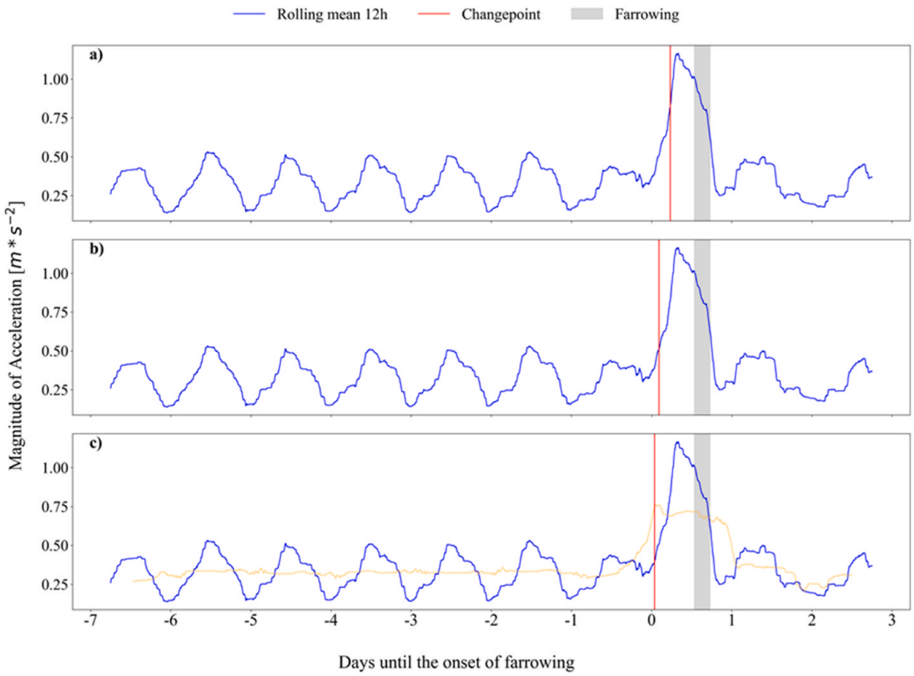

Fig. 7. Results for the nest-building detection with CUSUM (a), BEAST (b) and NestDetect (c) for one animal. The red lines in every subgraph show the detected changepoint for each method and the grey area shows the time of farrowing. For the shown animal all three methods detect the onset of nest-building correctly (within 24 h before the onset of farrowing), but CUMSUM detects the nest-building behaviour significantly later than BEAST and NestDetect. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

nondeterministic, the result for this detection can change with every run. In contrast, CUSUM and NestDetect are fully deterministic. Therefore, the results for the same individual were always the same and the interpretation was more straightforward and reliable. NestDetect detected the onset of nest-building behaviour the earliest. The highest runtime of all three methods of changepoint detection was found for BEAST, highlighting the computational complexity of this method. This resulted from the more complex algorithm behind the method. NestDetect performed best in detecting nest-building behaviour for a 48-h tolerance window, as well as for a 24-h tolerance window (Table 3). The ratio of correctly detected samples is shown in Fig. 6. For the sows where the NestDetect model did not detect nest-building behaviour correctly, an unusual change in the day-night cycle was detected which caused the alarm to be triggered too early.

4.2. Time to farrowing prediction

To predict the time to farrowing, the two methods, ANN and SR, are compared in terms of model performance and explainability. The performance indicators were calculated by running the test setting five times, both on the training and the test set. The ANN achieved an MAE of 9.6, and a R^2 of 0.77 over all pen types on the test dataset, where SR achieved a MAE of 9.42, and a R^2 of 0.64 (Table 4). The models have a similar R^2 and MAE on the test set and the ANN achieved the best results on the Wing pens.

Table 4
Comparison of model performance for time to farrowing prediction for all pet types. The bold numbers show the pen type where each method works best.

Method	Pen Type	MAE [h]		R^2	
		Training Data	Test Data	Training Data	Test Data
Neural Network	Wing	18.94	**3.66**	0.77	**0.90**
	SWAP	**13.25**	12.84	**0.85**	0.77
	Trapezoid	15.02	9.99	0.76	0.62
	All	16.03	9.64	0.79	0.76
Symbolic Regression	Wing	21.40	8.3	0.69	**0.88**
	SWAP	**15.58**	10.84	**0.76**	0.58
	Trapezoid	16.54	**7.85**	0.71	0.50
	All	18.12	9.47	0.72	0.64

The models perform better on the test set than on the training set when using the online test setting (Fig. 1).

The results for one sow for the time to farrowing prediction can differ a lot between the SR results and the ANN (Fig. 8). For the two examples displayed in Fig. 8, SR (8a) worked better on the sow from the training set in terms of the MAE. This was different for the sow in the test data set. Here both methods, SR and the ANN seemed to fit the 1:1 line similarly, but SR had a higher MAE in this example. Fig. 8b shows a more constant prediction trend across the observation period compared to Fig. 8a; however, it exhibits a higher error shortly before farrowing. In contrast, Fig. 8a does not display this pattern and maintains lower error values near the onset of farrowing. Since accurate predictions in the final hours before farrowing are critical for practical application, the model represented in Fig. 8a is considered more suitable for real-world use. This highlights the importance of evaluating not only overall fit but also performance during the most time-sensitive phase.

To eliminate the uncertainty of the nest-building behaviour detection for the assessment of the time to farrowing prediction performance, human annotations are added to the original dataset, where the NestDetect model cannot detect nest-building behaviour correctly. On this annotated dataset, a significantly improved performance for the SR model was observed, with a MAE of 7.49 h on the training set and 7.89 h on the test set. However, the predictions were found to be strongly biased toward overestimation, with 1416 overestimations and only 382 underestimations across 1798 total estimations. To address this imbalance, a series of tests was performed to introduce constant correction factors to the SR model predictions. The constant correction factors were added to the originally trained models (without retraining). The best results were achieved with a correction factor of 5 h, reducing the MAE to 5.89 h on the training set and 6.13 h on the test set. Moreover, this adjustment led to a more balanced prediction distribution with 834 underestimations and 955 overestimations, suggesting that such a

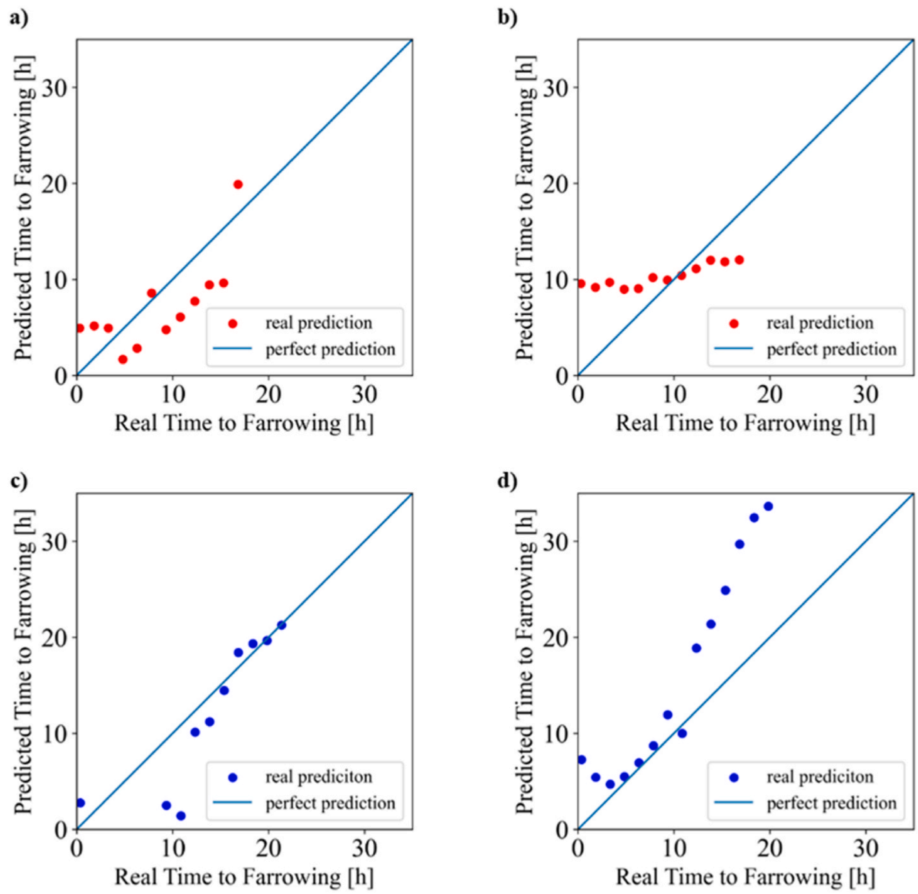

Fig. 8. Model performance on the training and test set for symbolic regression and a neural network with the online test setting. Subplot a) shows the results for the symbolic regression model on an individual from the training set (MAE 1.4 h). The same individual presented in b) with the results for the neural network (MAE 1.8 h). Subplots c) and d) show the results for an individual of the test set, where c) is the symbolic regression result (MAE 4.1 h) and d) is the neural network result (MAE 6.3 h).

correction could enhance both accuracy and practical applicability. A detailed comparison of different correction factors is provided in Table 5.

4.3. Evaluation of interpretability and explainability

The difference between open-box and closed-box models is that open-box models are inherently interpretable while closed-box models need an explainability method to explain the internals of a model (Fig. 5) (see Table 6). The ANN is a closed-box model, while SR is a an open-box model. This means that the resulting SR model was inherently interpretable and can be displayed as a formula.

Table 5
Impact of a constant correction factor on the prediction result. The correction is subtracted from every prediction in the test scenario. Furthermore, the number of big errors (>10 h) are shown below for a total number of 1798 observations. These tests were performed on a nest-building ground truth and therefore do not include the error for the first part of the prototype.

Correction Factor [h]	MAE [h]		Number of Over-estimations	Number of Under-estimations	Number of Errors >10 h
	Training Data	Test Data			
0	7.49	7.89	1416	382	516
2	6.55	6.97	1231	567	405
4	6.01	6.32	1059	739	337
5	5.89	6.13	955	843	319
6	5.89	6.14	856	942	309
8	6.22	6.38	650	1148	343

Table 6
Model performance and interpretability, where for the deep learning model the size is the total number of coefficients in the network and depth is the number of layers. For the symbolic regression the size is the number of nodes in the tree representation of the formula and the depth is the maximum tree depth. For both models the number of features describe the number of different features that are used by the model.

Model	Performance		Interpretability		
	R^2	MAE	Size*	Depth*	Features
Neural Networks	0.76	9.64	3969	3	80
Symbolic Regression	0.64	9.47	559	19	18

The ANN consisted of 3969 coefficients spread across three layers and used all 80 features (Table 6). The SR model had a size of 559, which is the number of nodes in the resulting tree representation of the resulting symbolic expression. This resulting tree had a depth of 19 and used 18 of the available 80 features for the prediction. Although the ANN was kept very small (only one hidden layer) it was still seven times more complex than the SR model.

Fig. 9 shows the influence of the features on the model output of the SR model. It shows that “mean activity during last 24 h” and “confinement period”, had the highest impact on the prediction result. These two were closely followed by the activity trend of the last 12 and 24 h. If the trend values were low, they had a high positive contribution on the prediction and if the values were high, they had a high negative impact in the model output. This meant that if sow activity changed a lot in the last 24 h then a lower time to farrowing was predicted.

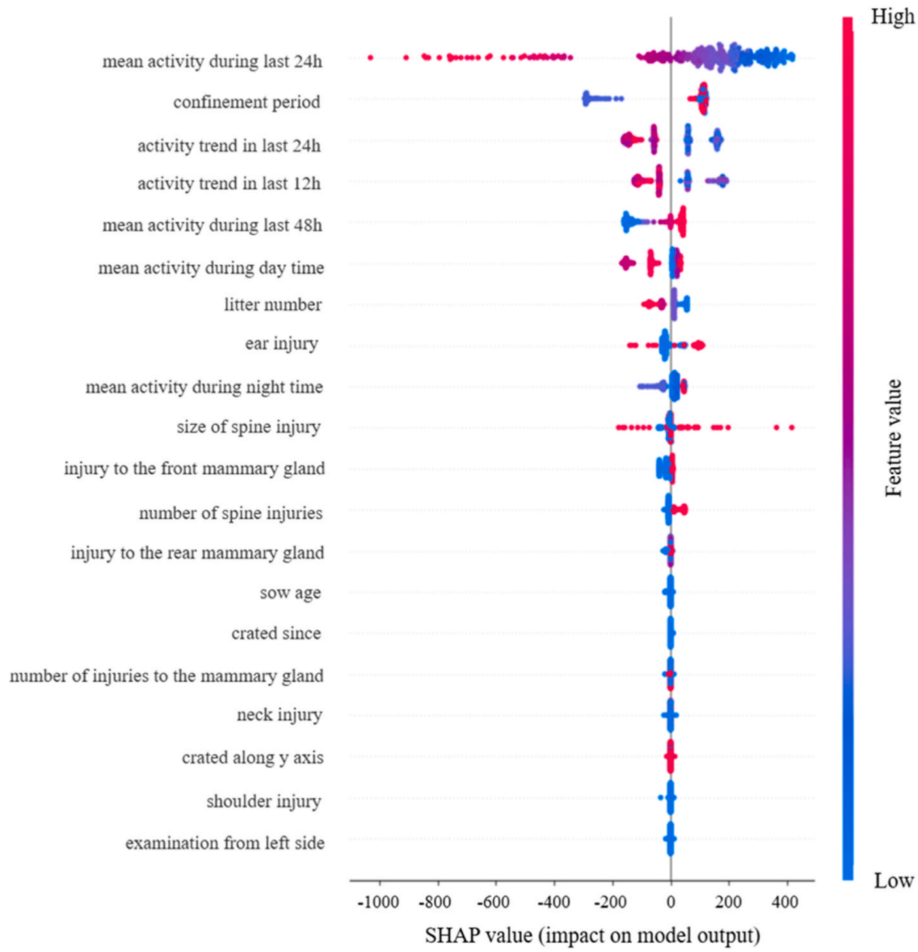

Fig. 9. Global explanation generated using SHAP for the symbolic regression model, showing the features with the highest impact on the prediction. Blue values indicate a low feature value and red values show the impact of a high feature value in the result. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

In contrast to the SR model, the feature with the highest impact on the result for the ANN was "crated since", which expresses the amount of time the sow was already crated in minutes (Fig. 10). Similar to the SR model, the mean activity during the last 24h was also a very important feature for the ANN where a high feature value had a high negative impact on the result. When looking at the general impacts of the features in the ANN, the impact that certain values had on the prediction result were more distinct for the SR model.

5. Discussion

The prediction models achieved a mean absolute error (MAE) of approximately 9 h on the test set. While this may initially seem high, it represents a substantial improvement over current farm practices, where farrowing time is typically estimated based on gestation length and visual observation, both of which can vary by more than a day. Importantly, the system provides continuous updates, and errors decrease as farrowing approaches. This is critical because accurate predictions close to the actual farrowing time are most valuable for farmers, enabling timely interventions to reduce sow stress and prevent piglet mortality. Compared to manual monitoring, the proposed approach offers a data-driven, automated solution that reduces labour and improves decision-making.

The evaluation of prediction performance in this study was based on the best-fit regression line rather than a fixed 1:1 line. The 1:1 line provides a direct assessment of systematic bias and shift between predicted and observed values, whereas the best-fit line reflects the actual relationship between predictions and observations under the given model assumptions. This approach allows for an intuitive interpretation of the coefficient of determination (R^2), which quantifies how well the model explains variance in the data.

However, using the best-fit line can mask potential bias and systematic error. Readers should interpret the reported R^2 values with this limitation in mind. The choice of the best-fit line was motivated by the primary goal of comparing relative performance between models (symbolic regression vs. neural network) rather than assessing absolute agreement. Future work should incorporate additional evaluation metrics, including deviation from the 1:1 line, to provide a more comprehensive assessment of model bias.

"Mean activity during the last 24h" was the variable with the highest impact on the prediction of "time to farrowing" in the SR model and the one with the second highest impact in the ANN model. This can be explained by the fact that nest-building starts 24 h before parturition and reaches its peak 6–12 h prepartum before decreasing as parturition gets closer (Castrén et al., 1993). "Confinement period" was also identified as a feature with a high impact on the SR model. While 74 out of 79 of the sows included in the dataset were not crated before parturition, the majority were confined one day before expected due date (on group level). This is most probably the reason why confinement before parturition led to a higher SHAP value. In the DL model, "crated since" was determined as one of the features with the highest effect on the prediction outcome. Confinement during the last few days before actual parturition and especially during nest-building phase can negatively affect the expression of nest-building behaviour (Hansen et al., 2017),

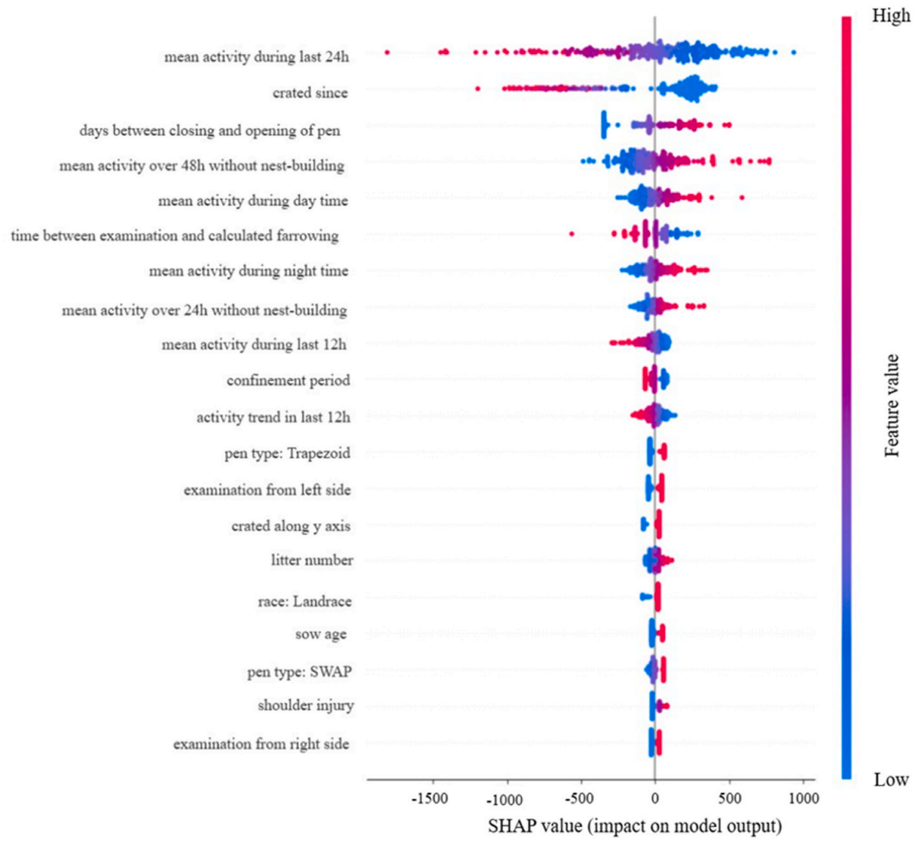

Fig. 10. Global explanation generated using SHAP for the neural network, showing the features with the highest impact on the prediction. Blue values indicate a low feature value and red values show the impact of a high feature value in the result. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

which might explain why high values of the variable “crated since” were associated with low predicted values.

In comparison with other studies (Manteuffel et al., 2015; Oczak, Maschat, & Baumgartner, 2019), this works utilised interpretable regression models to predict the remaining time until farrowing in minutes. Manteuffel et al. (2015) also used activity data based on light barriers of 34 sows to predict the time until farrowing and detect nest-building behaviour. Differently to the approach presented here, Manteuffel et al. (2015) used a classification approach to provide alarms for practitioners. They showed that body-mounted accelerometer sensors provided a higher accuracy. Oczak, Maschat, and Baumgartner (2019) implemented a two-staged system that triggered alarms before the onset of farrowing, but no time to farrowing prediction was developed. The accuracy of their alarms was 69.23 % for the first-stage alarm and 65.38 % for the second-stage alarm, with a tolerance window of 48 h before farrowing. In contrast, the presented NestDetect model achieved 82.68 % accuracy for the nest-building behaviour detection.

The extraction of sow activity levels from video data has already been proven possible (Oczak et al., 2023), and the utilisation of computer vision holds potential to further improve farrowing prediction models. This could be done by extracting features from video data such as pain or specific nest-building behaviours. Incorporating these behavioural changes could improve the prediction performance and also the interpretability of the models, which is crucial for them to be useful to domain experts, such as farmers and veterinarians.

Although the complexity of the SR model that resulted from this work was quite high with a tree depth of 19 and 559 nodes (Table 4) it was still more inherently interpretable than the ANN. Therefore, another aspect of further development are new presentation methods to make the SR result more intuitively interpretable. This can be done either by textual explanations of the models and predictions or by graphics that outline the key factors that lead to certain predictions (Lee et al., 2023). Such enhanced explanations will improve the usability of these systems for farmer and veterinarians, especially for large production systems.

Within this study all models were developed and tested within the same production system. Cross-validation and test set evaluation suggest good generalisation, external validation on different farms and housing conditions is necessary to confirm robustness and is. If substantial variability exists across production systems, retraining or adaptation may be required, which could limit scalability. Future work should therefore focus on multi-farm datasets and domain adaptation techniques to improve general applicability.

6. Conclusion

The developed approach of a two-stage farrowing prediction enables a farmer or veterinarian to be notified when the onset of nest-building behaviour is detected and from then on receive regular predictions on how much time is left until the onset of farrowing. Nest-building behaviour was detected with an accuracy of 82.68 % and the interpretable time to farrowing prediction achieved a PPC of 0.8 and a MAE of 9.48 h.

By using transparent models, it was possible to interpret why the resulting models give certain predictions, which is crucial in PLF systems using artificial intelligence. The comparison of ANN and SR results showed that similar results can be achieved, but that the SR results were inherently interpretable and there was no trade-off between performance and interpretability, as is often described in literature (Petkovic, 2023). The detection of the onset of farrowing has been previously researched before but no time to farrowing prediction has previously been developed. A new method for detecting the start of nest-building was proposed which is simpler but better performing, than previously researched methods and shown to be effective in one particular production system.

CRediT authorship contribution statement

Elisabeth Mayrhuber: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Conceptualization. **Kristina Maschat:** Writing – review & editing, Writing – original draft, Resources, Data curation. **David Brunner:** Writing – review & editing, Conceptualization. **Stephan M. Winkler:** Writing – review & editing, Supervision, Conceptualization. **Maciej Oczak:** Writing – review & editing, Supervision, Data curation, Conceptualization.

Data statement

The data used in this study were collected using a commercial product (SMARTBOW® ear tags) under license agreements with the manufacturer. Due to proprietary restrictions and confidentiality obligations, the raw data cannot be shared publicly. Researchers interested in similar data should contact the manufacturer for access under their terms.

Declaration of generative AI and AI assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT by OpenAI to increase the readability of the text. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded in whole or in part by the Austrian Science Fund (FWF) [https://doi.org/10.55776/DFH34]. For open access purposes, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

Project "Pro-SAU", where the used data originates from, was financed by former Bundesministerium für Gesundheit und Frauen (BMGF, Federal Ministry for Health and Women) and former Bundesministerium für Land-und Forstwirtschaft, Umwelt und Wasserwirtschaft (BMLFUW, Federal Ministry of Agriculture, Forestry, Environment and Water Management, research projects 100964, 100986 and 101062; BMLFUW-LE. 1.3.2/0086-II/1/2013). The open-access funding for this manuscript was given by FWF (austrian science fund).

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge Discovery & data mining.* https://doi.org/10.1145/3292500.3330701

Aldeia, G. S. I., & De França, F. O. (2021). Measuring feature importance of symbolic regression models using partial effects. *Proceedings of the genetic and evolutionary computation conference* (pp. 750–758). https://doi.org/10.1145/3449639.3459302

Bao, Y., Llagostera, P., & Plà-Aragonès, L. M. (2024). Is Deep Learning useful for decision making in pig production? *Internet of Things, 26*, Article 101229. https://doi.org/10.1016/j.iot.2024.101229

Castrén, H., Algers, B., De Passillé, A.-M., Rushen, J., & Uvnäs-Moberg, K. (1993). Preparturient variation in progesterone, prolactin, oxytocin and somatostatin in relation to nest building in sows. *Applied Animal Behaviour Science, 38*(2), 91–102. https://doi.org/10.1016/0168-1591(93)90059-X

Cornou, C., & Lundbye-Christensen, S. (2012). Modeling of sows diurnal activity pattern and detection of parturition using acceleration measurements. *Computers and Electronics in Agriculture, 80*, 97–104. https://doi.org/10.1016/j.compag.2011.11.001

Elliott, K. C., & Werkheiser, I. (2023). A framework for transparency in precision livestock farming. *Animals, 13*(21), 3358. https://doi.org/10.3390/ani13213358

Evans, B. P., Xue, B., & Zhang, M. (2019). What's inside the black-box?: A genetic programming method for interpreting complex machine learning models. *Proceedings of the genetic and evolutionary computation conference* (pp. 1012–1020). https://doi.org/10.1145/3321707.3321726

Fragoso, T. M., Bertoli, W., & Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review, 86*(1), 1–28. https://doi.org/10.1111/insr.12243

Gardezi, M., Joshi, B., Rizzo, D. M., Ryan, M., Prutzer, E., Brugler, S., & Dadkhah, A. (2024). Artificial intelligence in farming: Challenges and opportunities for building trust. *Agronomy Journal, 116*(3), 1217–1228. https://doi.org/10.1002/agj2.21353

Goodfellow, I., Bengio, Y., & Courville, A. (2023). *Deep learning.* Alanna Maldonado.

Hansen, C. F., Hales, J., Weber, P. M., Edwards, S. A., & Moustsen, V. A. (2017). Confinement of sows 24h before expected farrowing affects the performance of nest building behaviours but not progress of parturition. *Applied Animal Behaviour Science, 188*, 1–8. https://doi.org/10.1016/j.applanim.2017.01.003

Heidinger, B., Maschat, K., Kuchling, S., Hochfellner, L., Winckler, C., Baumgartner, J., & Leeb, C. (2022). Short confinement of sows after farrowing, but not pen type affects live-born piglet mortality. *Animal, 16*(2), Article 100446. https://doi.org/10.1016/j.animal.2021.100446

Hu, T. (2023). Genetic programming for interpretable and explainable machine learning. In L. Trujillo, S. M. Winkler, S. Silva, & W. Banzhaf (Eds.), *Genetic programming theory and practice XIX* (pp. 81–90). Singapore: Springer Nature. https://doi.org/10.1007/978-981-19-8460-0_4.

Khamesi, V. (2022). Ocpdet: A Python package for online changepoint detection in univariate and multivariate data. *Zenodo*. https://doi.org/10.5281/ZENODO.7632721 [Computer software] Version 0.0.5. .

Krieger, S., Oczak, M., Lidauer, L., Berger, A., Kickinger, F., Öhlschuster, M., Auer, W., Drillich, M., & Iwersen, M. (2019). An ear-attached accelerometer as an on-farm device to predict the onset of calving in dairy cows. *Biosystems Engineering, 184*, 190–199. https://doi.org/10.1016/j.biosystemseng.2019.06.011

Kronberger, G., Burlacu, B., Kommenda, M., Winkler, S. M., & Affenzeller, M. (2024). *Symbolic regression* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315166407

Larsen, M. L. V., Wang, M., & Norton, T. (2021). Information technologies for welfare monitoring in pigs and their relation to welfare quality®. *Sustainability, 13*(2), 692. https://doi.org/10.3390/su13020692

Lee, B. C. G., Downey, D., Lo, K., & Weld, D. S. (2023). LIMEADE: From AI explanations to advice taking. *ACM Transactions on Interactive Intelligent Systems, 13*(4), 1–29. https://doi.org/10.1145/3589345

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

Maddigan, P., Lensen, A., & Xue, B. (2024). *Explaining genetic programming trees using large Language models* (arXiv:2403.03397). *arXiv*. http://arxiv.org/abs/2403.03397.

Manteuffel, C., Hartung, E., Schmidt, M., Hoffmann, G., & Schön, P. C. (2015). Towards qualitative and quantitative prediction and detection of parturition onset in sows using light barriers. *Computers and Electronics in Agriculture, 116*, 201–210. https://doi.org/10.1016/j.compag.2015.06.017

Maschat, K., Dolezal, M., Leeb, C., Heidinger, B., Winckler, C., Oczak, M., & Baumgartner, J. (2020). Duration of confinement and pen-type affect health-related measures of welfare in lactating sows. *Animal Welfare, 29*(3), 339–352. https://doi.org/10.7120/09627286.29.3.339

Mayrhuber, E. M., Oczak, M., Iwersen, M., Brunner, D., Dorfer, V., & Winkler, S. (2024). Exploring the potential of symbolic regression in precision livestock farming. *European conference on precision livestock farming*.

Mei, Y., Chen, Q., Lensen, A., Xue, B., & Zhang, M. (2023). Explainable artificial intelligence by Genetic Programming: A Survey. *IEEE Transactions on Evolutionary Computation, 27*(3), 621–641. https://doi.org/10.1109/TEVC.2022.3225509

Morgan, L., Meyer, J., Novak, S., Younis, A., Ahmad, W. A., & Raz, T. (2021). Shortening sow restraint period during lactation improves production and decreases hair cortisol concentrations in sows and their piglets. *Animal, 15*(2), Article 100082. https://doi.org/10.1016/j.animal.2020.100082

Neethirajan, S. (2023). AI in sustainable pig farming: IoT insights into stress and gait. *Agriculture, 13*(9), 1706. https://doi.org/10.3390/agriculture13091706

Oczak, M., Maschat, K., & Baumgartner, J. (2019). Dynamics of sows' activity housed in farrowing pens with possibility of temporary crating might indicate the time when sows should be confined in a crate before the onset of farrowing. *Animals, 10*(1), 6. https://doi.org/10.3390/ani10010006

Oczak, M., Maschat, K., & Baumgartner, J. (2023). Implementation of computer-vision-based farrowing prediction in pens with temporary Sow confinement. *Veterinary Sciences, 10*(2), 109. https://doi.org/10.3390/vetsci10020109

Oczak, M., Maschat, K., Baumgartner, J., O'Brien, B., Hennessy, D., & Shalloo, L. (2019). *Monitoring of approaching farrowing in pens with possibility of temporary crating on the basis of ear tag acceleration data* (pp. 322–329).

Oczak, M., Maschat, K., Berckmans, D., Vranken, E., & Baumgartner, J. (2015). Classification of nest-building behaviour in non-crated farrowing sows on the basis of accelerometer data. *Biosystems Engineering, 140*, 48–58. https://doi.org/10.1016/j.biosystemseng.2015.09.007

Oliviero, C., Pastell, M., Heinonen, M., Heikkonen, J., Valros, A., Ahokas, J., Vainio, O., & Peltoniemi, O. A. T. (2008). Using movement sensors to detect the onset of farrowing. *Biosystems Engineering, 100*(2), 281–285. https://doi.org/10.1016/j.biosystemseng.2008.03.008

Page, E. S. (1954). Continuous INSPECTION SCHEMES. *Biometrika, 41*(1–2), 100–115. https://doi.org/10.1093/biomet/41.1-2.100

Pastell, M., Hietaoja, J., Yun, J., Tiusanen, J., & Valros, A. (2016). Predicting farrowing of sows housed in crates and pens using accelerometers and CUSUM charts. *Computers and Electronics in Agriculture, 127*, 197–203. https://doi.org/10.1016/j.compag.2016.06.009

Petkovic, D. (2023). It is not "Accuracy vs. Explainability"—We need both for Trustworthy AI Systems. *IEEE Transactions on Technology and Society, 4*(1), 46–53. https://doi.org/10.1109/TTS.2023.3239921

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. In L. Liu, & M. T. Özsu (Eds.), *Encyclopedia of database systems* (pp. 532–538). Springer US. https://doi.org/10.1007/978-0-387-39940-9_565.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you?. *Explaining the predictions of any classifier*. https://doi.org/10.48550/ARXIV.1602.04938

Sipper, M., Halperin, T., Tzruia, I., & Elyasaf, A. (2023). EC-KitY: Evolutionary computation tool kit in Python with seamless machine learning integration. *SoftwareX, 22*, Article 101381. https://doi.org/10.1016/j.softx.2023.101381

Thompson, R. J., Matthews, S., Plötz, T., & Kyriazakis, I. (2019). Freedom to lie: How farrowing environment affects sow lying behaviour assessment using inertial sensors. *Computers and Electronics in Agriculture, 157*, 549–557. https://doi.org/10.1016/j.compag.2019.01.035

Traulsen, I., Scheel, C., Auer, W., Burfeind, O., & Krieter, J. (2018). Using acceleration data to automatically detect the onset of farrowing in sows. *Sensors, 18*(1), 170. https://doi.org/10.3390/s18010170

Van Erp-van Der Kooij, E., De Graaf, L. F., De Kruijff, D. A., Pellegrom, D., De Rooij, R., Welters, N. I. T., & Van Poppel, J. (2023). Using sound location to monitor farrowing in sows. *Animals, 13*(22), 3538. https://doi.org/10.3390/ani13223538

Van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin & Review, 25*(1), 143–154. https://doi.org/10.3758/s13423-016-1015-8

Vargovic, L., Athorn, R. Z., Hermesch, S., & Bunter, K. L. (2022). Improving sow welfare and outcomes in the farrowing house by identifying early indicators from pre-farrowing assessment. *Journal of Animal Science, 100*(11). https://doi.org/10.1093/jas/skac294. skac294.

Zhao, K. (2025). *Rbeast* [C]. https://github.com/zhaokg/Rbeast (Original work published 2019).

Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., Li, Y., Toman, E., Mallick, B., Zhang, X., & Brown, M. (2019). Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sensing of Environment, 232*, Article 111181. https://doi.org/10.1016/j.rse.2019.04.034

Zhou, R., & Hu, T. (2024). Evolutionary approaches to explainable machine learning. In W. Banzhaf, P. Machado, & M. Zhang (Eds.), *Handbook of evolutionary machine learning* (pp. 487–506). Singapore: Springer Nature. https://doi.org/10.1007/978-981-99-3814-8_16.