A pipeline with pre-processing options to detect behaviour from accelerometer data using Machine Learning tested on dairy goats.

Mauny S[1*], Kwon J[2], Friggens NC[1], Duvaux-Ponter C[1], Taghipoor M[1]

[1] Université Paris-Saclay, INRAE, AgroParisTech, UMR Modélisation Systémique Appliquée aux Ruminants, 91120, Palaiseau, France

[2] Université Paris-Saclay, INRAE, AgroParisTech, UMR MIA-Paris-Saclay, 91120, Palaiseau, France

*Corresponding author

Correspondence: sarah.mauny@inrae.fr

Abstract

Animal behaviour is a significant component in the evaluation of animal welfare. Conducting continuous observations of animal behaviour is a time-consuming task and may not be feasible over extended periods for all animals. Thus, new technologies like sensors and cameras can be used to assess individual behaviour continuously. Combined with Artificial Intelligence (AI), accelerometers are promising to continuously and individually detect animal behaviour from the acceleration signals and characteristics of the behaviour. Such devices are commercialised for cattle but they have not been widely developed for small ruminants. Being able to automatically monitor behaviour at an individual scale represents a crucial step towards an objective assessment of animal welfare.

This paper aims to present the use of a pipeline called ACT4Behav (Accelerometer-based Classification Tool for identifying Behaviours) involving a supervised classification algorithm for automatically characterising specific animal behaviours using accelerometer data, and to explore the best pre-processing steps for each behaviour. This algorithm is designed to be general-purpose and applicable with different species, behaviours and accelerometers. This paper presents the use of this pipeline with eight indoor-housed goats equipped with ear-mounted accelerometers. Rumination, head-in-the-feeder, standing and lying behaviours were continuously sampled from camera recordings for 11 consecutive hours for each goat using The Observer software. The developed pipeline was used to identify optimal descriptive features and data preparation steps for each prediction model, one for each behaviour. A sensitivity analysis was conducted to assess the impact of the processing techniques and parameter value on the resulting AUC (Area Under the Curve) score, used as the performance score of the models. This analysis allowed the identification of the adequate filtering techniques, time-window segmentations, application of various transformations to raw data, and feature selections for each behaviour. Tuning the data pre-processing for each behaviour enhanced the ability to predict rumination (AUC score=0.800), head in the feeder (AUC score=0.819), lying (AUC score=0.829) and standing (AUC score=0.823) behaviours. When the application of the models on goats that did not participate in the training was tested by training the models on six goats and testing it on the two other goats, the AUC score for the four behaviours decreased (0.644, 0.733, 0.741 and 0.749 respectively for rumination, head in the feeder, lying and standing).

Keywords: animal behaviour, accelerometer, ruminant, dairy goat, Machine Learning, processing techniques

Abbreviations: ACCx, acceleration on the x-axis; ACCy, acceleration on the y-axis; ACCz, acceleration on the z-axis; AI, Artificial Intelligence; AUC, Area Under the Curve; FP rate, False Positive rate; ML, Machine Learning; PLF, Precision Livestock Farming; RFID, Radio Frequency Identification; ROC curve, Receiver Operating Characteristic curve; ts, time-series; TP rate, True Positive rate.

1. Introduction

Climate change and global population growth are leading agriculture to evolve and adapt towards more sustainable systems while ensuring health and welfare, and global food security. Europe is promoting agroecology as a key concept to transform our agricultural and food system. Despite the high interest of this transition, the animal will be faced with more environmental variations, and animal welfare could potentially be impacted. Accordingly, there is a growing societal demand to ensure the welfare of livestock animals. A further issue is that as farms grow in terms of number of animals and production, it is more difficult for farmers to care for each individual. Moreover, in these intensification conditions, welfare assessment represents a complex challenge given the focus of current welfare evaluation protocols on the herd or on the farming system rather than individual assessment (Winckler, 2019). Animal behaviour measures are key to assessing animal welfare (Bercksmans, 2014; Buller et al., 2020). For example, González et al. (2008) showed that feeding behaviour can be used as an early sign in the detection of sick cows. Feeding behaviour in calves has been used to detect gastrointestinal infections before the appearance of clinical signs (Svensson and Jensen, 2007). Similarly, a change in the pattern of drinking in pigs can be associated with a diarrhoea outbreak (Madsen and Kristensen, 2005). Thereby, behavioural information can provide a valuable indication of the physiological state of animals (Frost et al., 1997, Hansen, 2015) and can be used as an early sign of variation in animal welfare. Indeed, behavioural assessment plays a crucial role for farmers in evaluating the health and well-being of their livestock but it is not feasible for all the animals and over extended periods (Müller & Schrader, 2003).

Advancements in technology, particularly Precision Livestock Farming (PLF) technologies, have paved the way for transformative changes in how we approach animal behaviour assessment (Berckmans, 2014; Vogt, 2019). Real-time and near-time tracking technologies have unlocked the potential for collecting a diverse range of data over extended periods, encompassing not only production metrics and physiological states but also behavioural insights (Bailey et al., 2021). These advancements significantly empower livestock productivity and welfare enhancement, according to Bailey et al. (2021).

Multiple sensors have been developed to monitor animal behaviour and the 3-axis accelerometer has quickly become popular to assess behavioural states in various animal species. These sensors measure changes in an object's acceleration on the three axes (x-, y- and z-axis). As a result, combined with Artificial Intelligence (AI), commercial devices are developed to quantify livestock behaviour, especially for species with economic interest (Borchers et al., 2016; Pereira et al., 2018) such as cattle, with CowManager (Agis, Harmelen, the Netherlands) detecting eating behaviour and rumination, the HOBO Data Logger (HOBO Pendant G Acceleration Data Logger, Onset Computer Corporation, Pocasset, MA), the AfiAct Pedometer Plus (Afimilk, Kibbutz Afikim, Israel) and the Track A Cow (ENGS, Rosh Pina, Israel) to monitor lying time, the CowAlert IceQube (IceRobotics Ltd., Edinburgh, Scotland) to quantify lying time and detect lameness (Thorup et al., 2015), and the Smartbow (Smartbow GmbH, Jutogasse, Austria) to monitor rumination time and to detect oestrus (Schweinzer et al., 2019). Most of these systems are used in the context of health alteration, for example detection of disease, or of behaviours that could impact animal health and performance. Since the reliability of these systems is not always reported, it is not possible to adapt them to other species such as small ruminants. Those limitations suggest that further research is needed to adapt existing algorithms to a wide range of behaviours on small ruminants based on accelerometer data.

Riaboff et al. (2022) described the three-step methodology commonly used in the literature for predicting ruminant behaviour from accelerometer data. These are data collection, data pre-processing that prepares the data for input in the behavioural classification model, and the development of the behavioural classification model. Numerous studies have compared various processing techniques and model development strategies, such as the best features to use (Smith et al. 2016; Kamminga et al., 2018; Riaboff et al., 2019), the Machine Learning (ML) algorithms (Lopez et al., 2013; Smith et al. 2016; Kamminga et al., 2018; Sakai et al., 2019), the signal filtering techniques (Riaboff et al., 2019) and the time-windows size (Smith et al. 2016; Riaboff et al., 2019). To date, no established methodology systematically searches for the best features and processing techniques to predict each targeted behaviour accurately from raw accelerometer data. The main objective of this study was to use a pipeline, called ACT4Behav (Accelerometer-based Classification Tool for identifying Behaviours), that integrates a comprehensive exploration of each data processing step to determine the best-fitted model for each behaviour, applied here to raw accelerometer data to characterise certain goat behaviours. To do so, a “sensitivity analysis” was used to evaluate the impact of variations of key factors at the data pre-processing step. These factors were filtering techniques, time windows sizes, inclusion of additional time series data, data transformation methods, and variations in feature selection. The performance score, which assesses the quality of behaviour predictions, served as an indicator of the model's reliability within the same experimental context. In this work, the main performance score was the AUC (Area Under the Curve) score. Accuracy, balanced accuracy, F1-score, sensitivity and specificity were also calculated. ~~However, in this study, detecting behaviours in goats that were not part of the model's training was also tested to reflect the ability of the model to generalise across different animals.~~ Moreover, in this study, detection of behaviours in goats that were not part of model’s training dataset was also tested to reflect the ability of the model to generalise detection of behaviours on data from new goats. The model was trained with a dataset of six goats and its performance was tested on the two remaining goats.

2. Methods

2.1. Animals

Data collection was carried out in the experimental goat unit of INRAE in Thiverval-Grignon, France in March 2022. Eight indoor-housed dairy Alpine goats were equipped with MSR145 3D-accelerometer (TH Industrie) sensors that can measure accelerations in the positive and negative directions, with a maximum magnitude of 15 g. The goats were normally group-housed in a straw bedded area but for the purpose of having reference measures of feeding behaviour, they were group-housed in the same slatted-floor pen of

8 animals for 24 days with continuous access to feed and water. Each goat had its own feed trough that was released via the electronic ear-tag when the goat allocated to the trough placed its head next to the antenna. This work was carried out under licence (Apafis number #24314-2019120915403741 v5).

2.2. Acceleration data collection

Goats were equipped with accelerometers attached to the RFID ear tag of the animal (**Figure 1**). The accelerometers were 27 × 16 × 53 mm in size and weighed approximately 20 g. The accelerometers were powered with a 230 mAh lithium-polymer battery and were programmed to record the acceleration on the x-y-z-axis at a frequency of 5 Hz. Data from eight animals was collected for 24 consecutive hours.

2.3. Behavioural observation

Video recordings were made using cameras, one placed above every pen to provide visual information on the goat behaviour (**Figure 2**). The time of the video recording system was synchronised to the same time as the accelerometers by manually and vigorously agitating the accelerometers in front of the cameras. The shake pattern was set as a reference time to match the video recordings to the accelerometer data. During the observation process, each animal was uniquely marked with a number on its back using animal spray paint. This numbering system allowed the observer to track and differentiate the goats throughout the observation. The videos were labelled by a single trained observer using The Observer® XT software version 16.0 (Noldus Information Technology, 2022), using a pre-established ethogram. The ethogram of all the annotated activities is available in **Appendix 1**.

Figure 1 - A goat with an accelerometer fixed with tape to the right RFID ear tag in lateral view.

Figure 2 - Screenshot of the video system above the goat pens. Each goat has a number painted on its back with yellow animal spray paint.

The ethogram was designed to describe the goats' activities within this experimental system. The behaviours were categorised into five distinct groups, ensuring an exhaustive description of all the activities. These categories included (i) feeding behaviour, (ii) social behaviour covering interactions between goats, such as grooming and fighting, (iii) position and movement behaviour describing activities related to the goats' posture and movement, such as standing, walking, and lying down and (iv) the "other behaviour" category accounted for actions that did not fit into the previous groups. Lastly, the "disturbance" category captured any unexpected or disruptive events that could affect the goats' behaviour. It is worth to mention that at any given time during the observation period, each goat was expected to exhibit a behaviour from each category. The behaviours « ruminating », « lying », « standing » and « head in the feeder » were selected to develop the classification model because of their large representation in the observation period to maximise available data for model development, and for their relevance regarding welfare and health status of the animals. In this context, it has been shown that general activity levels and time spent feeding are reduced under bad health conditions such as lameness (Thorup et al., 2016), while time spent lying can increase. The recordings were analysed only during the presence of daylight, which was approximatively 11 hours per day for each of the eight goats. **Figure 3** shows the cumulative time spent in each behaviour by all the eight goats during 11 hours each.

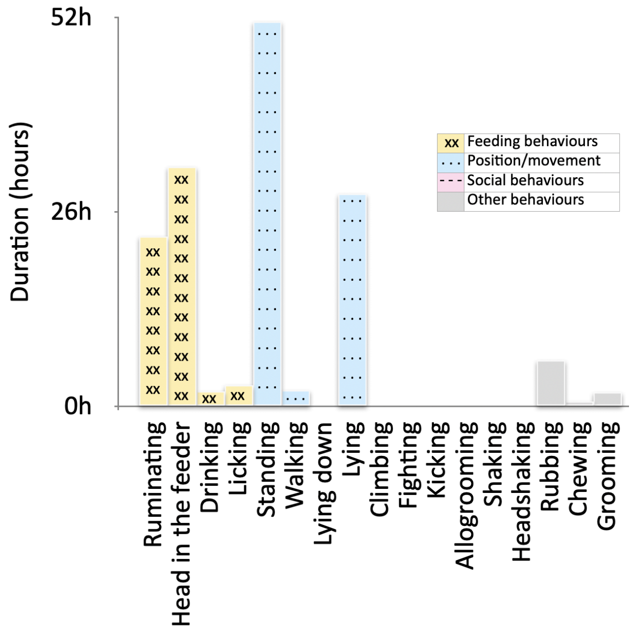


Figure 3 - Total duration of the behaviours on video recordings in eight goats during 11 hours each. The bars of the histogram are colour-coded according to behaviour categories. On the x-axis are the various observed behaviours. On the y-axis, the total cumulative time spent by the eight animals on each behaviour.

Raw data, i.e. accelerometer data and corresponding behaviours, is available online (https://doi.org/10.57745/LGZBM1) (Mauny et al, 2024a).

2.4. Data pre-processing

In this section, the different steps of the pipeline that prepare data for use in the ML algorithm are detailed. The data was pre-processed separately for each behaviour, resulting in four binary classification models, one for each behaviour.

2.4.1. Additional time-series

Additional time-series calculation involved generating supplementary time-series derived from the original accelerometer data (acceleration on the x-axis, on the y-axis and on the z-axis). These additional time series served to enhance the depth of information available by calculating time-series independently from the sensor orientation or time-series related to the two main components of the acceleration. Additional time-series were calculated from the raw or filtered acceleration data.

Norm. The ear placement of the sensors on the identification tag of the animals was made without considering the orientation of the sensors. The ear is quite mobile on the goat, making the sensor highly responsive to any movement of the animal. Consequently, it was proposed to compute the Euclidean norm, referred to as the magnitude of the acceleration, which is an additional time series that is not affected by the orientation of the sensor, e.g. Fida et al., 2015.

$a_{mag} = \sqrt{acc_x^2 + acc_y^2 + acc_z^2}$ with acc_i the acceleration on the i-axis.

The magnitude of the signal was calculated to account for the variety of sensor orientations.

High-passed Filtered data. To focus on the behaviour of an animal, the dynamic acceleration of the signal can be extracted by subtracting the static acceleration from the raw acceleration (Lush et al., 2018). In the recorded signal, we expect static acceleration to vary slowly and therefore to correspond to low frequencies, which can be eliminated using a high-pass filter, as proposed by Smith et al. (2016). The performance of the model was evaluated and compared with raw accelerometer data and with filtered acceleration data.

Pitch and roll angles. According to Riaboff et al. (2022), pitch and roll angles are useful to predict behaviours that involve different postures and head tilt like putting the head in the feeder, standing or lying.

Rotated acceleration data. To ensure accurate and comparable analysis, a strategy was developed to standardise the orientation of the accelerometer. If we consider that the typical posture of a goat is head straight and upright, then the goal is to align the data so that the vertical axis (0, 0, 1) corresponds to a uniform reference direction. The multivariate median or mean is used as an estimator of the central tendency of the data, considering all three acceleration axes, assuming that most of the time, the goat is in this typical posture, i.e., head straight and upright. Subsequently, a rotation matrix was computed based on the difference between the calculated multivariate median and the desired reference direction (0, 0, 1).

These additional time-series were calculated from the formulas indicated in **Table 1**. Pitch and roll angles, along with a transformation of the acceleration data, were calculated from each acceleration value and added as new variables to the dataset. The influence on the performance scores of the predictive model for each behaviour when adding these additional time-series was compared to the performance of the model without adding additional time-series (**Figure 4**). Combining the time-series was also tested.

Table 1 - Additional time-series calculated from the raw or filtered acceleration

Time-series	***Formula***	***Description***
Pitch angle	$pitch = arctan(\frac{-acc_x}{\sqrt{acc_y{}^2 + acc_z{}^2}})$	***Characterises the rotation around the z-axis***
Roll angle	$roll = arctan(\frac{acc_y}{acc_z})$	***Characterises the rotation around the x-axis***
Rotated acceleration data (mean)	$R = rotation_{mean} = [mean(acc_x), mean(acc_y), mean(acc_z)]$	***Transformation of the data to standardise the sensor orientation***

Rotated acceleration data (median)	$R = rotation_{median} = [median(acc_x), median(acc_y), median(acc_z)]$ $[acc_x, acc_y, acc_z] * R = [0,0,1]$ $acc_{rotated} = acc_i * R^{-1}$

Note:** acc_i **is the acceleration on the** i**-axis, R is the rotation matrix which is computed based on the mean or the median

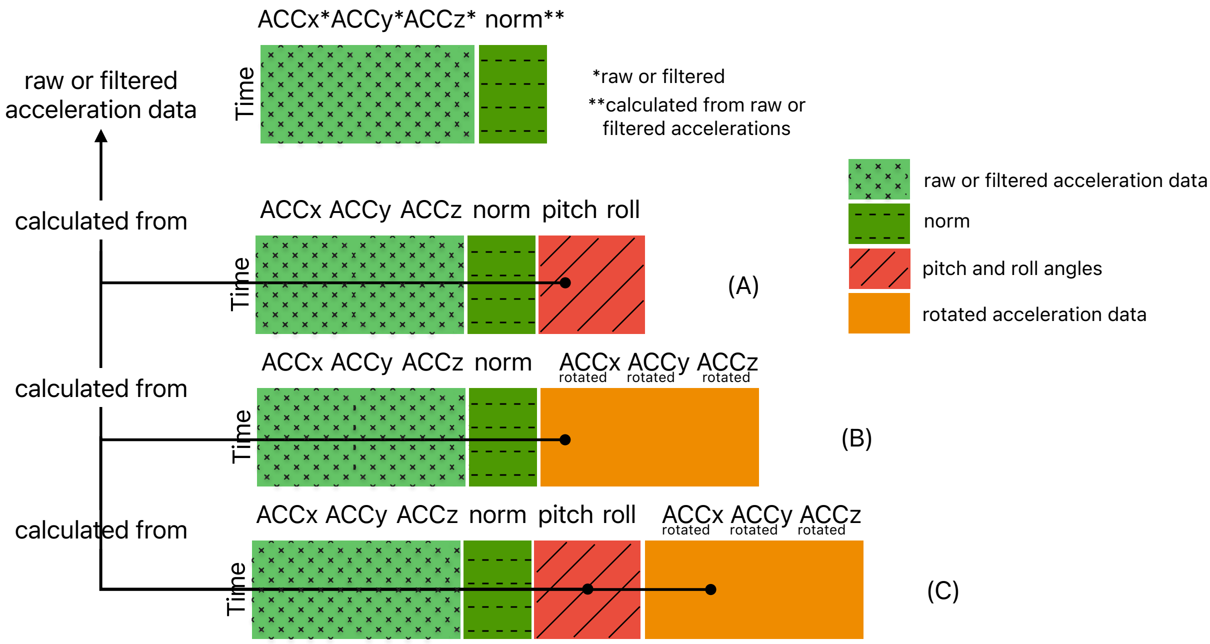


Figure 4 - Overview of the dataset with raw or filtered acceleration data (light green) and the norm (dark green) without additional time-series and overview of the dataset when adding pitch and roll angles (red) as additional time-series (A) or when adding rotated acceleration data (orange) as additional time-series (B) from raw or filtered acceleration data (light green) or combining them (C).

2.4.2. Features extraction and segmentation into time-windows

Predictions were not made on the raw acceleration data: an important pre-processing step is the transformation of the raw accelerometer data into 'features', which can be various descriptive quantities and which are statistical variables for the ML models. The calculated features constitute the input of each ML model. To do so and for each animal, the accelerometer time-series data were segmented into time-windows. Each window was the statistical unit of the dataset thereafter and the features were calculated on each time-window. The literature on recognizing human activities via accelerometer data highlights a diversity of possible time-window segmentations (Bersch et al., 2014) such as fixed-size overlapping, or not, sliding window (Pietka, 1988) or variable-size sliding window (Laguna et al., 2011). However, in the context of classifying ruminant behaviours, fixed-size windows are predominantly used, mostly without overlap (Riaboff et al., 2022). In this study, in order to find which size of time-window gives the best performance score for each behaviour, several sizes of time-windows were tested for each behaviour, ranging from 10 seconds to 120 seconds. The prediction of a given behaviour was approached as one binary classification model, where label 1 corresponded to the behaviour happening during more than 50% of the time-window.

In the field of ruminant behaviour prediction, previous studies have relied on a predefined set of features in both the time and frequency domains. In this study, an extensive feature extraction on each time-serie of the dataset was chosen to explore features that were never considered in previous studies. This was followed by a feature selection based on the importance of each feature during the model development rather than limiting the number of features before the prediction.

A wide range of 777 different features was automatically calculated (Python (version 3.10) package tsfresh (Christ et al., 2018)) on each time-window and on each extracted time-series (**Figure 5**). The calculated features are descriptive statistics features (mean, median, standard deviation, etc.), temporal characteristics (number of crossings, count above/below thresholds, etc.), value distribution features (percentiles, quantiles, histograms, etc.), spectral domain features (features derived from the frequency domain using Fast Fourier Transform), pattern measures (pattern counting, entropy-based calculations, etc.), signal characteristics (energy-related characteristics), anomaly detection and regression features (autoregressive models coefficients, polynomial fitting coefficients, etc.).

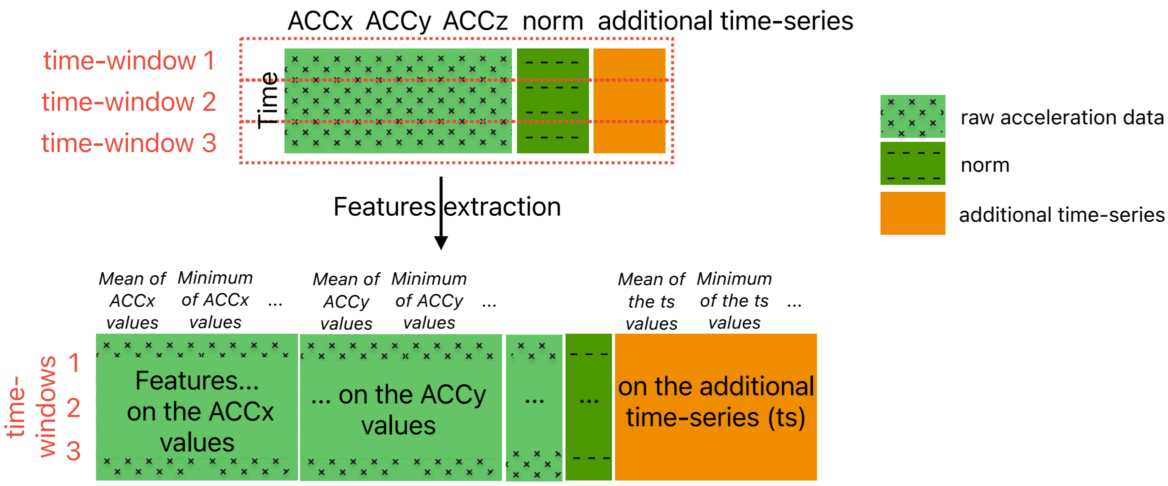


Figure 5 - Overview of the features extraction process of every time-series (ts), i.e. raw acceleration data (light green), norm (dark green) and additional ts (orange), and on each time-window (represented in red dashed lines) of the dataset

2.5. Behaviour classification model

In supervised learning, the ML model is developed by training the algorithm on a labelled dataset called the training set, where each input data point is associated with a corresponding label, i.e. a behaviour. The pre-processed dataset is therefore split into a training set, a validation set, and a test set.

For training, the Catboost algorithm (Prokhorenkova et al., 2019) was used, which is an implementation of the Gradient Boosting algorithm for classification and regression. It is based on the principle of aggregating a large number of decision trees (a flowchart structure where each node represents a decision, each branch the outcome of the decision and each leaf a class in classification tasks) that best predict the training set with constraints to limit overfitting. Those constraints are given by hyperparameters, e.g. the number of leaves, the maximum depth of each decision tree and the minimum number of samples required to be at a leaf node. A thorough hyperparameter tuning was performed using the validation set ("validation set" in **Figure 6**; "val" set in **Figure 7**) and based on the highest AUC Area Under the Curve) score obtained.

After training, a model was generated and the validation set ("val") was used to calculate preliminary performance scores of the model to tune the hyperparameters. Once the model was trained and tuned, the test set ("test set" in **Figure 6;** "X_test", "y_test" in **Figure 7**) was used to evaluate the final performance of the model with unseen data: the model predicts the behaviour from the test set data without having the corresponding labels. Then, the predicted behaviours are compared to the corresponding observed labels to calculate the final performance score.

In this work, two different types of models were developed based on two data split approaches (**Figure 6**):

(1) The dataset was divided based on the time-windows of all goats for sensitivity analysis, meaning that data from all goats was used during the training of the model and predictions were done for the same goats on different periods of time.
After the data pre-processing, time-windows for each goat were computed and 20% of those time-windows were excluded to create the test set, the 80% left was split into a training set (75%) and a validation set (25%).

(2) Performance scores were also computed using a goat-based split. In this case, the training was done on data from a group of goats and predictions were made on another group in the same context. Two goats were selected for testing, and the data of the six other goats was used to split the time-windows into training (75%) and validation (25%) sets. A cross-validation technique was employed using four distinct pairs of goats for testing, meaning rotating the pair of goats which was going to be used for the test set. For each pair of goats used for the test set, the performance score of the model was calculated using the AUC score. The mean of the four AUC scores was the final AUC score of the model.

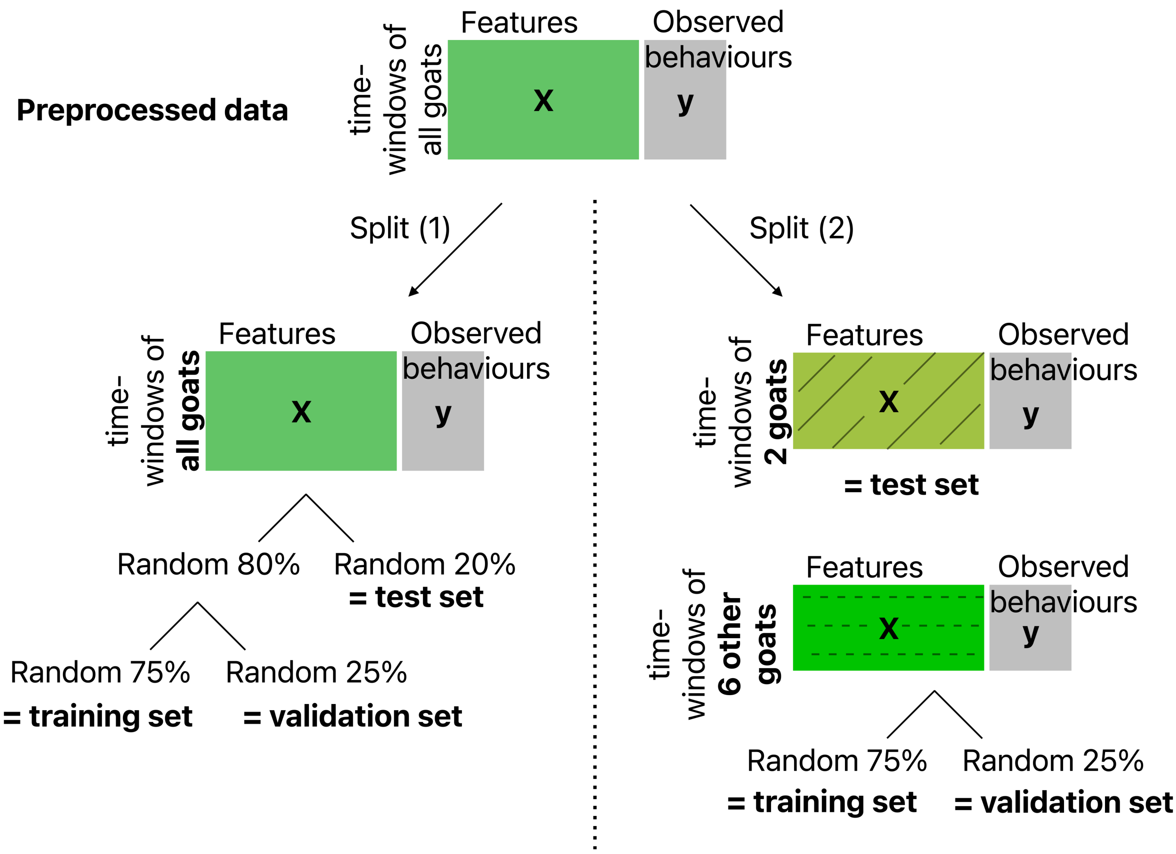


Figure 6 - Dataset splits into training, validation and test sets with a time-windows split approach (1) a goat-based split approach (2).

The final performance of the models was evaluated using usual evaluation metrics: AUC score, accuracy, balanced accuracy, sensitivity, specificity and F1-score, calculated on the test set with the formulas presented in **Appendix 2**.

2.6. Features selection

The features_importances_ function from the scikit-learn package (Pedregosa et al., 2011) was used to evaluate the importance of each feature. It looks at all the decisions made by each tree in the model. It then figures out which features were most helpful on average for making those decisions and calculates a weight, which indicates the impact of that feature on the prediction. The features are then ranked by their weight, allowing us to keep only a limited number of the features with the highest importance. After the features selection, the performance of the models was recalculated with 2000, 1000, 500, 100, 80, 50, 20 and 10 most important features selected with the AUC score.

Without features selection, the models were trained with 777 features multiplied by n (the number of time-series in the dataset). Note that the high dimensionality of the data was not addressed in this work, as Gradient Boosting can handle high-dimensional datasets and train models with thousands of features. But by selecting the most important features, it is possible to improve the performance score of the models by focusing on the most relevant ones, and reduce the preprocessing workload, resulting in smaller data volumes and faster processing times.

2.7. Sensitivity analysis

Variations in the choices of each pre-processing step and features selection were tested on the performance score of the models. To do so, the influence of the value of the cutoff frequency of the high pass filter, presence or absence or additional time series, size of the time-windows and the number of the best selected features were studied:

- Raw acceleration data filtering: cutoff frequencies of 0.05 Hz, 0.01 Hz, 0.1 Hz, 0.2 Hz, 0.3 Hz, 0.4 Hz
- Additional time-series calculation (combined or not): pitch, roll, rotated acceleration data (mean/median)
- Size of the time-window: 10, 20, 30, 40, 50, 60, 80, 120 seconds
- Features selection: no selection, 2000, 1000, 500, 100, 80, 50, 20, 10 best features selected

We used the AUC score as a key performance metric to assess the effectiveness of each pre-processing option. We tested how variations in the pre-processing steps affected the AUC score. It is important to note that while commonly used metrics such as accuracy, sensitivity, specificity and F1 score provide valuable insights into model performance, they are threshold-dependent. When working with binary classification problems, the predicting models gives prediction probabilities in output. To calculate the accuracy for example, a fixed threshold set at 0.5 classifies the prediction probabilities into 0 when lower than 0.5 and 1 when higher than 0.5. Then, it compares the true labels of the samples with the predicted labels and calculates the accuracy. The AUC score measures the area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive (TP) rate (sensitivity) against the False Positive (FP) rate across various threshold values. Accordingly, the AUC score provides a single measure to evaluate model performances that can be used to determine which model is better on average. A value above 0.70 is considered to be a “strong model”, while a value below 0.60 is considered to be a “weak model” (Kelleher et al., 2020). A value of 1 would mean a perfect classification.

The different steps of model development are represented in **Figure 7**.

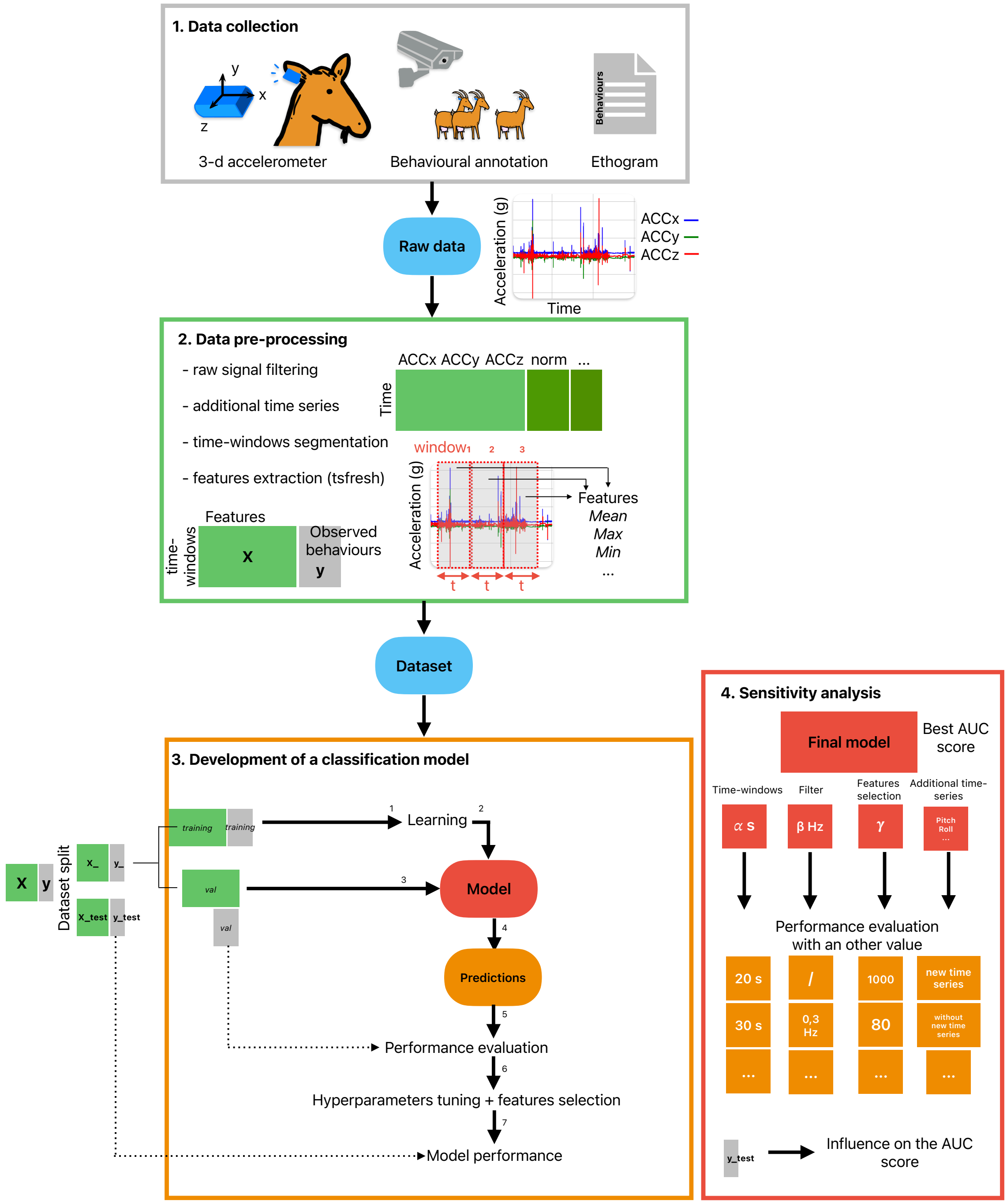


Figure 7 - Flowchart representing the pipeline of the process of raw acceleration data and associated behaviours processing to automatic predictions and sensitivity analysis of the pre-processing steps. The data collection (1.) is described in the sections **2.1**, **2.2** & **2.3**. The data pre-processing (2.) is presented in section **2.4**. The development of the model (3.) is explained in section **2.5**. The sensitivity analysis (4.) methodology is presented in section **2.7**.

3. Results

3.1. Sensitivity analysis

The behaviour "ruminating" is used as an illustrative example to present the results in detail. The results of “head in the feeder”, “lying” and “standing” behaviours are presented in **Appendices 3, 4 and 5**.

For rumination, the best AUC score (0.800) was obtained with a 10 and 20 seconds time window (**Table 2**), no filtering (**Table 3**), three additional time-series (median rotated data, pitch and roll series (**Table 4**) and a selection of the 100 first-best features (**Table 5**).

Table 2 Classification performance for the behaviour “ruminating” using the algorithm with different selected time-window sizes.

Time-window sizes (s)	AUC
10	0.800
20	0.800
30	0.792
40	0.790
50	0.773
60	0.746
80	0.795
120	0.782

Table 3 Classification performance for the behaviour “ruminating” using the algorithm with different cut-off frequencies.

Cut-off frequencies (Hz)	AUC
No filtering	0.800
0,01	0.749
0,05	0.741
0,1	0.734
0,2	0.731
0,3	0.747
0,4	0.762

Table 4 Classification performance for the behaviour “ruminating” using the algorithm with and without additional time-series.

Additional time-series	AUC
None	0.769
Euler angles	0.774
Rotated acceleration data (mean)	0.797

Rotated acceleration data (median)	0.798
Rotated acceleration data (mean) + Euler angles	0.781
Rotated acceleration data (median) + Euler angles	0.800

Table 5 Classification performance for the behaviour "ruminating" using the algorithm with features selection.

Number of the best features selected	AUC
6993 (no features selection)	0.791
2000	0.792
1000	0.796
500	0.799
100	0.800
80	0.799
50	0.792
20	0.797
10	0.764

The optimal pre-processing steps and features selection for each behaviour and their scores are presented in **Table 6**. The pre-processing treatments listed in **Table 6** are those who give the best AUC score for each behaviour. Accuracy, balanced accuracy, F1-score, sensitivity, and specificity of each model are also mentioned.

Table 6 Optimal processing steps for the four chosen behaviours with performance scores

	Ruminating	Head in the feeder	Lying	Standing
Time-windows size (s)	20	60	50	60
Raw acceleration filtering (Hz)	/	/	/	/
Additional time-series	Median + Euler angles	Mean + Euler angles	Mean + Euler angles	Euler angles
Features selection	100 best features	80 best features	80 best features	80 best features
AUC score	0.800	0.819	0.829	0.823
Accuracy (%)	78.7	73.7	78.1	74.6
Balanced accuracy (%)	70.6	73.8	76.5	70.3
F1-score (%)	56.3	67.7	70.6	81.5
Sensitivity (%)	54.2	74.3	69.8	95.4
Specificity (%)	58.6	62.2	71.3	71.2

The highest AUC score was obtained with 20 seconds time-windows for “ruminating”, 60 seconds for “head in the feeder” and for “standing”, and 50 seconds for “lying”. The acceleration data filtering did not improve the AUC score of the four behaviours. The Euler angles improved the AUC scores of the four behaviours. The rotated acceleration data improved the AUC scores of “ruminating”, “head in the feeder” and “lying” behaviours. For “head in the feeder”, “lying” and “standing” behaviours, the 80 best features were sufficient to achieve the best predictive AUC score while the 100 best features were necessary for “ruminating”.

3.2. Feature importance

The most important feature of each model for the prediction of the behaviour is the sum of reoccurring values calculated on the rotated ACCx (acceleration on the x-axis) for “rumination”. The sum of reoccurring values identifies values that recur within the dataset, like a pattern.

The permutation entropy calculated on the rotated ACCz (acceleration on the z-axis) is the most important feature for “head in the feeder” and the permutation entropy calculated on the norm is the most important feature for “lying”. Permutation entropy is a measure of the complexity or unpredictability of a time series. This feature belongs to the pattern measures category as it quantifies the amount of randomness in the sequence of observations.

The complexity estimate based on the Lempel-Ziv compression algorithm calculated on ACCz is the most important feature for “standing”. It converts the acceleration values of each time-window into binary values and measures the repetitiveness of binary sequences to detect a repetitive pattern.

3.3. Behaviour prediction with goat-based split approach

A decrease in the AUC scores for the four behaviours was observed when training each model with a goat-based split approach (**Table 7**). The most important decrease was obtained for “rumination” (-19.5%), dropping from 0.800 to 0.644. “Head in the feeder” and “lying” decreased by 10.5% and 10.6% respectively and standing by -9.0%.

Table 7 Performance score of the four behaviours with a goat-based split approach and cross-validation

	Rumination	Head in the feeder	Lying	Standing
AUC score with a goat-split approach	0.644	0.733	0.741	0.749
AUC score with a time-window-split approach	0.800	0.819	0.829	0.823
Percentage change	-19.5%	-10.5%	-10.6%	-9.0%

4. Discussion

This paper describes a supervised classification algorithm for automatically characterising four goat behaviours using accelerometer data. These behaviours were well represented in the dataset and are often well predicted in the literature in small ruminants like sheep with accuracies above 90% and/or f1-scores above 75% for standing and lying (Mansbridge et al., 2018; Walton et al., 2018; Hu et al., 2020) and ruminating (Hu et al., 2020; Decandia et al., 2021). Transitional behaviours or rare behaviours were observed but their short duration and low frequencies of occurrence made them impossible for supervised learning models to predict in the present dataset. Moreover, applying the presented pipeline in similar contexts during different periods or in different farms could be useful to detect similarities in the key factors for predicting certain behaviours. For instance, the present experiment was carried out on slatted-floor (commonly used in feeding trials) but other housing environments exist such as straw-bedded floor. The four behaviours could be expressed with a different acceleration pattern on straw and other behaviours might have been restrained on slatted-floor. This highlights the importance of the farm context that can influence animal behaviour.

The use of a goat-based split approach allows for evaluating the robustness of the model (Rahman et al., 2018). When training was done using different animals, a decrease in the AUC score was observed, particularly for rumination; AUC decreased from 0.800 to 0.644. This indicates a limited generalizability in predicting those behaviours on goats that did not participate in the training, which is the biggest challenge for developing long-term behavioural monitoring systems because the goal is to avoid using behavioural observation. The difficulty lies in the high individual variability of the expressed behaviours, the available training data was not sufficient to train a strong model that can be used on goats that did not participate in the training in this farm context; training and testing and on a small dataset gathered in the same context is not enough to develop strong models. However, our work establishes a practical pipeline to pre-pre-processed data and generate models with data collected in several contexts. Using the pipeline to pre-process new datasets could allow to train afterwards stronger models from larger training datasets. The developed models show the potential of accelerometer data to reliably predict certain goat behaviours.

The analysis of the impact of using different selected values at each pre-processing step helps to identify the optimal parameters for each model. In the literature, the effect of the window size varies greatly from one study to another, from a few seconds (Robert et al., 2009) to several minutes (Vazquez-Diosdado et al., 2019). In our work, the best AUC scores were obtained with different time-window sizes depending on the considered behaviour (10 or 20 seconds for "ruminating"; 60 seconds for "head in the feeder" and "standing"; 50 seconds for "lying"). The 20 seconds time-window was selected for "ruminating" for faster computation. The size of the time-window can be explained as a trade-off between the greater amount of information that can be encompassed by a large time-window and the risk that a too large time-window would contain more than one behaviour being expressed (Banos et al., 2014). The challenge is to find the ideal window size, which the average duration of the predicted behaviour could also explain. The shorter the duration of the bout of behaviour, the smaller the window should be, and this relationship is influenced by the data acquisition frequency. A higher data acquisition frequency should allow to capture more details in acceleration patterns, which may result in better performance scores. Especially that the experimental context can interfere on the apparition and duration of a behaviour and have repercussions on the ideal size of the time-windows. Results show that a large time window is better for "head in the feeder" while a small time window seems to lead to the best score for "rumination". This could partly be explained by the fact that animals spend a longer time at the feeder after the feed delivery, while rumination corresponds to shorter events, often interrupted.

Filtering the data seems to decrease the model performance. Although it can appear that ear-mounted accelerometer signals are likely to be noisy (Chapa et al., 2020), filtering the signal does not seem to improve the performance of the prediction model in the case of animal behaviour prediction (Riaboff et al., 2019). In our study, and for the four behaviours, data filtering systematically decreased the AUC scores. High-pass filter excluding the low frequencies of the signal removes information from the signal that seems useful to predict the four behaviours. We conclude that the orientation of the static acceleration provided by raw acceleration data is useful to predict rumination, and lying and standing positions. Moreover, for each tested cut-off frequency, very low frequencies are filtered and thus, there are no major effects of the tested filters on the AUC scores. In several papers, calculating additional time-series like pitch and roll improved the performance prediction as it represents the rotation around two axes due to changes in position (Lush et al., 2018 during urination events in sheep; Vazquez-Diosdado et al., 2019 for walking, lying and standing behaviours in sheep). It approximates the head/body inclination during the expressed behaviour and thus is useful for predicting behaviours involving different postures and body/head tilt (Zobel et al., 2015, Lush et al., 2018). Pitch and roll enhanced the AUC score of the four behaviours in the present study, which often induce a unique head position. During rumination and standing, goats keep their head still and upright. When they have their head in the feeder, the head is mostly down, while they eat the food. During lying behaviour, goats can keep their head upright when ruminating or just observing their environment. They can also put their head down on another goat to rest or under their own body, inducing a few head movements. Moreover, additional time series enriches the features set, as tsfresh calculates several features on each time-series of the data. These features effectively decompose the original time-series into components that are not simply linear transformations but which capture more complex facets like non-linear entropic measures to quantify the complexity, unpredictability, or irregularity of a time series for example. The most important features for “head in the feeder” and “lying” were features that calculate the permutation entropy on the z-axis rotated acceleration and on the norm respectively. This feature captures the complexity of a dynamic system by characterising the order relations between values of a time series (Henry & Judge, 2019). For instance, when comparing the means of the permutation entropy calculated on the norm when the goat was lying or not, the mean value for lying events (4.3±0.3) was lower than the one for non-lying events (4.5±0.1) (**Appendix 6**). This suggests that during lying events, the pattern in the time series data tends to be more predictable, it has higher order relations compared to periods without lying events, and it fits with the common intuition that lying is a mostly static behaviour. Generating a large set of features shows the potential for reliable predictors of these behaviours.

In this study, the AUC score was selected as the primary evaluation metric over accuracy, balanced accuracy, F1-score, specificity, and sensitivity. According to López et al. (2013), the AUC score is a comprehensive measure of classifier performance, encompassing the other metrics cited and providing a reliable measure to identify which model is better on average. AUC score is recommended when dealing with imbalanced data and when aiming to achieve good quality results for both classes, i.e., the presence of the behaviour or not (López et al., 2013), without being threshold-dependent. Therefore, the search for the best pre-processing steps focuses on improving the AUC score. Few studies use AUC score as the referential performance metric. Strong results have been obtained in cows, for example Cabezas et al. (2022) for rumination (AUC=0.967) and lying (AUC=0.894); Shen et al. (2020) obtained a score of 0.959 for feeding and rumination. In sows, Escalante et al. (2013) obtained an AUC score of 0.872 for feeding behaviour and 0.787 for lying. No studies on small ruminants used this metric to evaluate the performance of the model. The AUC scores of the models developed with a split based on the time-windows of all goats when predicting “ruminating”, “head in the feeder”, “standing” and “lying” were 0.800, 0.819, 0.829 and 0.823 respectively.

5. Conclusion

This study aimed to develop a supervised classification algorithm to characterise goat behaviour automatically using accelerometer data. It involved a comprehensive analysis of various pre-processing techniques, parameter values, and feature sets to identify optimal configurations for behaviour prediction. Segmenting data into time windows and the computation of features based on the time series enriched the dataset, contributing to better describing the signal and improving the predictive capabilities of the model. While the model showed good performance in predicting “ruminating”, “head in the feeder” and posture behaviours, challenges were found in achieving acceptable scores when training involved different animals than the animals used for prediction. Collecting data on a small number of animals limits the prediction capacity of the model. The developed pipeline in this paper is freely available online (https://doi.org/10.5281/zenodo.12624785) (Mauny et al, 2024b) and lays a foundation to gather supplementary data for automated behaviour recognition using accelerometer data with automatic features calculation and selection. Its use is described in a datapaper (Mauny et al., in prep).

6. Appendices

Appendix 1 Ethogram of the observed activities in the experimental setup.

Behaviour	Category	Description
Standing	Position	*The animal is standing still, all four limbs are on the ground. Occasionally, one limb may be raised, and the animal occasionally moves its head.*
Walking	Position	*At least two limbs are in motion.*
Lying down	Position	*The goat first places its two front legs, knees on the ground, then the hindquarters fall down.*
Lying	Position	*The animal is lying down and still, occasionally moving its head.*
Standing up	Position	*The goat positions itself on its knees, forearms on the ground, and then lifts its hindquarters to stand up.*
Climbing	Position	*The goat places its front legs on a surface or an object.*
Milking	Position	*The animal has left the pen to go for milking.*
Other position	Position	*Position or movement not mentioned above.*
Non-visible	Position	*The movement/position of the animal cannot be determined with certainty or is not visible, incomplete…*
Head in the feeder	Feeding	*The animal opens its gate by placing its head in it.*
Drinking	Feeding	*The animal has its muzzle in the water trough.*
Ruminating	Feeding	*The animal chews for a while. The regular jaw movements accompanied by ear movement can be observed, periodically interrupted by the swallowing of the bolus.*
Salt block	Feeding	*The animal licks the salt block.*
None	Feeding	*The animal is not eating, does not have its muzzle in the water trough, and is not ruminating.*
Other feeding behaviour	Feeding	*Feeding behaviour not mentioned above.*

Behaviour	Category	Description
Kicked	Social	*The animal is targeted by another animal. It receives a headbutt to its head or another part of the body.*
Fighting	Social	*The animal fights with another individual. The animal headbutts the head or body of the other animal. The animal may rise on its hind legs and then fall back before headbutting another individual.*
Allogrooming	Social	*The animal's muzzle makes contact with the body of another animal.*
Other social interaction	Social	*Social interaction not mentioned above.*
None	Social	*No social interaction.*
Shaking	Other	*The animal shakes its entire body vigorously and then often briefly shakes its head as a second step.*
Head shaking	Other	*The animal shakes its head only.*
Rubbing	Other	*The animal rubs its body against the barrier and moves forward while continuing to press its body against it.*
Chewing	Other	*Nibbles on something (anything that is not food or another goat).*
Grooming	Other	*The animal scratches or licks itself.*
Scratching	Other	*The goat uses one of its hind legs to scratch the front part of its body.*
Other behaviour	Other	*Other behaviour not mentioned above.*
None	Other	*No other behaviour.*
None	Other	*No other behaviour.*
Disturbance yes	Disturbance	*An unexpected/abnormal event occurs; an individual enters the device, noise, camera obstructed…*
No disturbance	Disturbance	*Nothing abnormal/unexpected to report.*

Appendix 2 Confusion matrix (a) and performance model evaluation metrics (b)

(a)

	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

(b)

Metric	*Formula*
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Sensitivity (recall)	$\frac{TP}{FN + TP}$
Specificity	$\frac{TN}{TN + FP}$
Balanced accuracy	$\frac{(sensitivity + specificity)}{2}$
Precision	$\frac{TP}{TP + FP}$
F1-score	$2 * \frac{recall * precision}{recall + precision}$
AUC score	Area under the Receiver Operating Characteristic (ROC) curve

Appendix 3 Classification performance for "head in the feeder" using the Catboost algorithm with different selected time-window sizes (i), cut-off frequencies (ii), additional time-series (iii), numbers of selected features (iv).

Time-window sizes (s)	**10**	**20**	**30**	**40**	**50**	**60**	**80**	**120**
AUC score	0.772	0.779	0.775	0.779	0.784	0.819	0.770	0.773

Cut-off frequencies (Hz)	**No filtering**	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**
AUC score	0.819	0.746	0.784	0.782	0.775	0.785	0.793

Additional time-series	None	Euler angles	Rotated acc data (mean)	Rotated acc data (median)	Rotated acc data (mean) + Euler angles	Rotated acc data (median) + Euler angles
AUC score	0.797	0.795	0.815	0.775	0.819	0.774

Number of the best features selected	6993 (no selection)	2000	1000	500	100	80	50	20	10
AUC score	0.816	0.817	0.819	0.818	0.818	0.819	0.817	0.808	0.792

Appendix 4 Classification performance for “lying” using the Catboost algorithm with different selected time-window sizes (i), cut-off frequencies (ii), additional time-series (iii), numbers of selected features (iv).

Time-window sizes (s)	10	20	30	40	50	60	80	120
AUC score	0.809	0.789	0.780	0.787	0.829	0.817	0.801	0.775

Cut-off frequencies (Hz)	No filtering	0.01	0.05	0.1	0.2	0.3	0.4
AUC score	0.829	0.764	0.767	0.774	0.782	0.776	0.803

Additional time-series	None	Euler angles	Rotated acc data (mean)	Rotated acc data (median)	Rotated acc data (mean) + Euler angles	Rotated acc data (median) + Euler angles
AUC score	0.829	0.764	0.784	0.792	0.829	0.790

Number of the best features selected	6993 (no feature selection)	2000	1000	500	100	80	50	20	10
AUC score	0.822	0.822	0.826	0.826	0.825	0.829	0.823	0.811	0.810

Appendix 5 Classification performance for “standing” using the Catboost algorithm with different selected time-window sizes (i), cut-off frequencies (ii), additional time-series (iii), numbers of selected features (iv).

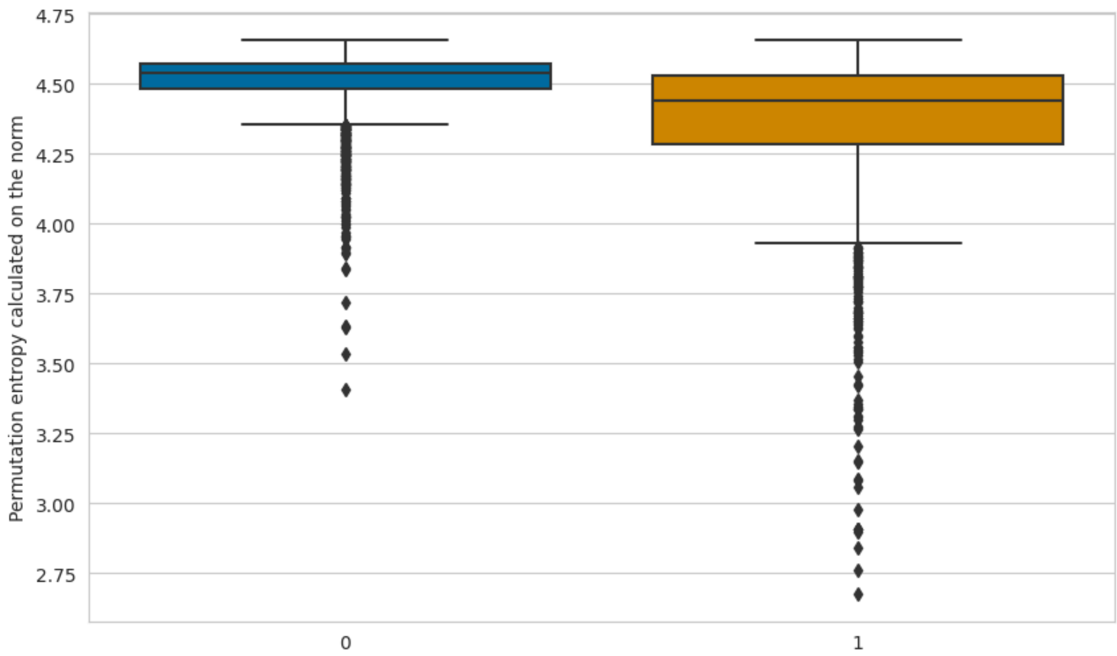
Time-window sizes (s)	10	20	30	40	50	60	80	120
AUC score	0.790	0.799	0.820	0.774	0.804	0.823	0.802	0.801

Cut-off frequencies (Hz)	No filtering	0.01	0.05	0.1	0.2	0.3	0.4
AUC score	0.823	0.808	0.797	0.772	0.780	0.792	0.798

Additional time-series	None	Euler angles	Rotated acc data (mean)	Rotated acc data (median)	Rotated acc data (mean) + Euler angles	Rotated acc data (median) + Euler angles
AUC score	0.819	0.823	0.815	0.766	0.819	0.776

Number of the best features selected	4662 (no feature selection)	2000	1000	500	100	80	50	20
AUC score	0.819	0.818	0.819	0.819	0.819	0.820	0.816	0.801

Appendix 6 Permutation entropy calculated on the norm distribution during non-lying (0) and lying events (1)



Acknowledgements

We sincerely thank the staff at the experimental goat unit of INRAE in Thiverval-Grignon and the ADAPT-HERD project for allowing us to collect acceleration and behavioural data on the goats during the experiment.

Data, scripts, code, and supplementary information availability

Data is available online: doi:10.57745/LGZBM1 https://doi.org/10.57745/LGZBM1 (Mauny et al, 2024a)

Scripts and code are available online: doi: 10.5281/zenodo.12624785 https://doi.org/10.5281/zenodo.12624785 (Mauny et al, 2024b)

Description of the data and code is currently in preparation (Mauny et al, in prep).

[The references of the datasets, scripts and codes should also be present in the reference list and cited in the text.]

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Funding

This work benefited from state aid managed by the National Research Agency under France2030 under the reference ANR-22-PEAE-0008.

References

Bailey, D.W., Trotter, M.G., Tobin, C., Thomas, M.G., 2021. Opportunities to Apply Precision Livestock Management on Rangelands. Frontiers in Sustainable Food Systems 5. https://doi.org/10.3389/fsufs.2021.611915

Banos, O., Galvez, J.-M., Damas, M., Pomares, H., Rojas, I., 2014. Window Size Impact in Human Activity Recognition. Sensors 14, 6474–6499. https://doi.org/10.3390/s140406474

Berckmans, D., 2014. Precision livestock farming technologies for welfare management in intensive livestock systems. Rev. Sci. Tech. OIE 33, 189–196. https://doi.org/10.20506/rst.33.1.2273

Bersch, S.D., Azzi, D., Khusainov, R., Achumba, I.E., Ries, J., 2014. Sensor data acquisition and processing parameters for human activity classification. Sensors (Basel) 14, 4239–4270. https://doi.org/10.3390/s140304239

Borchers, M.R., Chang, Y.M., Tsai, I.C., Wadsworth, B.A., Bewley, J.M., 2016. A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors. Journal of Dairy Science 99, 7458–7466. https://doi.org/10.3168/jds.2015-10843

Brown, D.D., Kays, R., Wikelski, M., Wilson, R., Klimley, A.P., 2013. Observing the unwatchable through acceleration logging of animal behavior. Animal Biotelemetry 1, 20. https://doi.org/10.1186/2050-3385-1-20

Buller, H., Blokhuis, H., Lokhorst, K., Silberberg, M., Veissier, I., 2020. Animal Welfare Management in a Digital World. Animals 10, 1779. https://doi.org/10.3390/ani10101779

Cabezas, J., Yubero, R., Visitación, B., Navarro-García, J., Algar, M.J., Cano, E.L., Ortega, F., 2022. Analysis of Accelerometer and GPS Data for Cattle Behaviour Identification and Anomalous Events Detection. Entropy 24, 336. https://doi.org/10.3390/e24030336

Chapa, J.M., Maschat, K., Iwersen, M., Baumgartner, J., Drillich, M., 2020. Accelerometer systems as tools for health and welfare assessment in cattle and pigs – A review. Behavioural Processes 181, 104262. https://doi.org/10.1016/j.beproc.2020.104262

Christ, M., Braun, N., Neuffer, J., Kempa-Liehr, A.W., 2018. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). Neurocomputing 307, 72–77. https://doi.org/10.1016/j.neucom.2018.03.067

Decandia, M., Rassu, S.P.G., Psiroukis, V., Hadjigeorgiou, I., Fountas, S., Molle, G., Acciaro, M., Cabiddu, A., Mameli, M., Dimauro, C., Giovanetti, V., 2021. Evaluation of proper sensor position for classification of sheep behaviour through accelerometers. Small Ruminant Research 201, 106445. https://doi.org/10.1016/j.smallrumres.2021.106445

Escalante, H.J., Rodriguez, S.V., Cordero, J., Kristensen, A.R., Cornou, C., 2013. Sow-activity classification from acceleration patterns: A machine learning approach. Computers and Electronics in Agriculture 93, 17–26. https://doi.org/10.1016/j.compag.2013.01.003

Fida, B., Bernabucci, I., Bibbo, D., Conforto, S., Schmid, M., 2015. Pre-Processing Effect on the Accuracy of Event-Based Activity Segmentation and Classification through Inertial Sensors. Sensors 15, 23095–23109. https://doi.org/10.3390/s150923095

Frost, A.R., Schofield, C.P., Beaulah, S.A., Mottram, T.T., Lines, J.A., Wathes, C.M., 1997. A review of livestock monitoring and the need for integrated systems. Computers and Electronics in Agriculture, Livestock Monitoring 17, 139–159. https://doi.org/10.1016/S0168-1699(96)01301-4

González, L.A., Tolkamp, B.J., Coffey, M.P., Ferret, A., Kyriazakis, I., 2008. Changes in Feeding Behavior as Possible Indicators for the Automatic Monitoring of Health Disorders in Dairy Cows. Journal of Dairy Science 91, 1017–1028. https://doi.org/10.3168/jds.2007-0530

Hansen, I., 2015. Behavioural indicators of sheep and goat welfare in organic and conventional Norwegian farms. Acta Agriculturae Scandinavica, Section A — Animal Science 65, 55–61. https://doi.org/10.1080/09064702.2015.1050447

Henry, M., Judge, G., 2019. Permutation Entropy and Information Recovery in Nonlinear Dynamic Economic Time Series. Econometrics 7, 10. https://doi.org/10.3390/econometrics7010010

Hu, S., Ingham, A., Schmoelzl, S., McNally, J., Little, B., Smith, D., Bishop-Hurley, G., Wang, Y.-G., Li, Y., 2020. Inclusion of features derived from a mixture of time window sizes improved classification accuracy of machine learning algorithms for sheep grazing behaviours. Computers and Electronics in Agriculture 179, 105857. https://doi.org/10.1016/j.compag.2020.105857

Kamminga, J.W., Le, D.V., Meijers, J.P., Bisby, H., Meratnia, N., Havinga, P.J.M., 2018. Robust Sensor-Orientation-Independent Feature Selection for Animal Activity Recognition on Collar Tags. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2, 15:1-15:27. https://doi.org/10.1145/3191747

Kelleher, J.D., Namee, B.M., D'Arcy, A., 2020. Fundamentals of Machine Learning for Predictive Data Analytics, second edition: Algorithms, Worked Examples, and Case Studies. MIT Press.

Laguna, J.O., Olaya, A.G., Borrajo, D., 2011. A dynamic sliding window approach for activity recognition, in: Proceedings of the 19th International Conference on User Modeling, Adaption, and Personalization, UMAP'11. Springer-Verlag, Berlin, Heidelberg, pp. 219–230. https://doi.org/10.1007/978-3-642-22362-4_19

López, V., Fernández, A., García, S., Palade, V., Herrera, F., 2013. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences 250, 113–141. https://doi.org/10.1016/j.ins.2013.07.007

Lush, L., Wilson, R.P., Holton, M.D., Hopkins, P., Marsden, K.A., Chadwick, D.R., King, A.J., 2018. Classification of sheep urination events using accelerometers to aid improved measurements of livestock contributions to nitrous oxide emissions. Computers and Electronics in Agriculture 150, 170–177. https://doi.org/10.1016/j.compag.2018.04.018

Madsen, T.N., Kristensen, A.R., 2005. A model for monitoring the condition of young pigs by their drinking behaviour. Computers and Electronics in Agriculture 48, 138–154. https://doi.org/10.1016/j.compag.2005.02.014

Mansbridge, N., Mitsch, J., Bollard, N., Ellis, K., Miguel-Pacheco, G.G., Dottorini, T., Kaler, J., 2018. Feature Selection and Comparison of Machine Learning Algorithms in Classification of Grazing and Rumination Behaviour in Sheep. Sensors 18, 3532. https://doi.org/10.3390/s18103532

Mauny, S., Kwon J., Friggens, N.C, Duvaux-Ponter, C., Taghipoor, M., 2024a. Accelerometers and annotated video recordings of behaviours: a dataset for training behaviour classification models in goats. Recherche Data Gouv, V1. https://doi.org/10.57745/LGZBM1

Mauny, S., Kwon J., Friggens, N.C, Duvaux-Ponter, C., Taghipoor, M., 2024b. ACT4Behav (Accelerometer-based Classification Tool for identifying Behaviours): a Machine Learning pipeline with extensive pre-processing and feature creation options. Zenodo. https://doi.org/10.5281/zenodo.12624796

Müller, R., Schrader, L., 2003. A new method to measure behavioural activity levels in dairy cows. Applied Animal Behaviour Science 83, 247–258. https://doi.org/10.1016/S0168-1591(03)00141-2

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Pereira, G.M., Heins, B.J., Endres, M.I., 2018. Technical note: Validation of an ear-tag accelerometer sensor to determine rumination, eating, and activity behaviors of grazing dairy cattle. Journal of Dairy Science 101, 2492–2495. https://doi.org/10.3168/jds.2016-12534

Pietka, E., 1988. Expert systems in parameter extraction of the ECG signal, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Presented at the Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 165–166 vol.1. https://doi.org/10.1109/IEMBS.1988.94460

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2019. CatBoost: unbiased boosting with categorical features. https://doi.org/10.48550/arXiv.1706.09516

Rahman, A., Smith, D.V., Little, B., Ingham, A.B., Greenwood, P.L., Bishop-Hurley, G.J., 2018. Cattle behaviour classification from collar, halter, and ear tag sensors. Information Processing in Agriculture 5, 124–133. https://doi.org/10.1016/j.inpa.2017.10.001

Riaboff, L., Aubin, S., Bédère, N., Couvreur, S., Madouasse, A., Goumand, E., Chauvin, A., Plantier, G., 2019. Evaluation of pre-processing methods for the prediction of cattle behaviour from accelerometer data. Computers and Electronics in Agriculture 165, 104961. https://doi.org/10.1016/j.compag.2019.104961

Riaboff, L., Shalloo, L., Smeaton, A.F., Couvreur, S., Madouasse, A., Keane, M.T., 2022. Predicting livestock behaviour using accelerometers: A systematic review of processing techniques for ruminant behaviour prediction from raw accelerometer data. Computers and Electronics in Agriculture 192, 106610. https://doi.org/10.1016/j.compag.2021.106610

Robert, B., White, B.J., Renter, D.G., Larson, R.L., 2009. Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. Computers and Electronics in Agriculture 67, 80–84. https://doi.org/10.1016/j.compag.2009.03.002

Sakai, K., Oishi, K., Miwa, M., Kumagai, H., Hirooka, H., 2019. Behavior classification of goats using 9-axis multi sensors: The effect of imbalanced datasets on classification performance. Computers and Electronics in Agriculture 166, 105027. https://doi.org/10.1016/j.compag.2019.105027

Schweinzer, V., Gusterer, E., Kanz, P., Krieger, S., Süss, D., Lidauer, L., Berger, A., Kickinger, F., Öhlschuster, M., Auer, W., Drillich, M., Iwersen, M., 2019. Evaluation of an ear-attached accelerometer for detecting estrus events in indoor housed dairy cows. Theriogenology 130, 19–25. https://doi.org/10.1016/j.theriogenology.2019.02.038

Shen, W., Cheng, F., Zhang, Yu, Wei, X., Fu, Q., Zhang, Yonggen, 2020. Automatic recognition of ingestive-related behaviors of dairy cows based on triaxial acceleration. Information Processing in Agriculture 7, 427–443. https://doi.org/10.1016/j.inpa.2019.10.004

Smith, D., Rahman, A., Bishop-Hurley, G.J., Hills, J., Shahriar, S., Henry, D., Rawnsley, R., 2016. Behavior classification of cows fitted with motion collars: Decomposing multi-class classification into a set of binary problems. Computers and Electronics in Agriculture 131, 40–50. https://doi.org/10.1016/j.compag.2016.10.006

Thorup, V.M., Munksgaard, L., Robert, P.-E., Erhard, H.W., Thomsen, P.T., Friggens, N.C., 2015. Lameness detection via leg-mounted accelerometers on dairy cows on four commercial farms. Animal 9, 1704–1712. https://doi.org/10.1017/S1751731115000890

Thorup, V.M., Nielsen, B.L., Robert, P.-E., Giger-Reverdin, S., Konka, J., Michie, C., Friggens, N.C., 2016. Lameness Affects Cow Feeding But Not Rumination Behavior as Characterized from Sensor Data. Front. Vet. Sci. 3. https://doi.org/10.3389/fvets.2016.00037

Vázquez-Diosdado, J.A., Paul, V., Ellis, K.A., Coates, D., Loomba, R., Kaler, J., 2019. A Combined Offline and Online Algorithm for Real-Time and Long-Term Classification of Sheep Behaviour: Novel Approach for Precision Livestock Farming. Sensors 19, 3201. https://doi.org/10.3390/s19143201

Vogt, M., 2019. Sustainability Certification Schemes in the Agricultural and Natural Resource Sectors: Outcomes for Society and the Environment. Routledge.

Walton, E., Casey, C., Mitsch, J., Vázquez Diosdado, J.A., Yan, J., Dottorini, T., Ellis, K.A., Winterlich, A., Kaler, J., 2018. Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour. This Issue 5, 171442–171442. https://doi.org/10.1098/rsos.171442

Winckler, C., 2019. Assessing animal welfare at the farm level: do we care sufficiently about the individual? Animal Welfare 28, 77–82. https://doi.org/10.7120/09627286.28.1.077

Zobel, G., Weary, D.M., Leslie, K., Chapinal, N., von Keyserlingk, M.A.G., 2015. Technical note: Validation of data loggers for recording lying behavior in dairy goats. Journal of Dairy Science 98, 1082–1089. https://doi.org/10.3168/jds.2014-8635