Computers and Electronics in Agriculture 238 (2025) 110799

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Evaluating ROCKET and Catch22 features for calf behaviour classification from accelerometer data using Machine Learning models

Oshana Dissanayake [a,c,*], Sarah E. McPherson [c,d,e], Joseph Allyndrée [b], Emer Kennedy [c,d], Pádraig Cunningham [a], Lucile Riaboff [a,c,f]

[a] *School of Computer Science, University College Dublin, Ireland*
[b] *School of Maths and Stats, University College Dublin, Ireland*
[c] *VistaMilk SFI Research Centre, Ireland*
[d] *Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, P61C997, Ireland*
[e] *Animal Production Systems Group, Wageningen University & Research, Wageningen, The Netherlands*
[f] *GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet-Tolosan, France*

ARTICLE INFO

Keywords:
Accelerometers
Behaviour
Features
ROCKET
Catch22
Machine Learning
Dairy calf

ABSTRACT

Monitoring calf behaviour continuously would be beneficial to identify routine practices (e.g., weaning, transport, dehorning, etc.) that impact calf welfare in dairy farms. In that regard, accelerometer data collected from neck collars can be used along with Machine Learning models to classify calf behaviour automatically but further development is needed to classify a broad spectrum of behaviours with good genericity. While the performance of Machine Learning models has been extensively studied, the accelerometer features used as inputs have received limited attention. In this paper, we explored the performance of ROCKET and Catch22 features that were specifically designed for time-series classification. For that purpose, 30 Irish Holstein Friesian and Jersey pre-weaned calves were filmed while equipped with an accelerometer sensor. Behaviours were annotated, allowing for 27.4 h of accelerometer time-series aligned with the corresponding behaviour. ROCKET and Catch22 features were extracted from raw and additional accelerometer time-series, along with commonly used features in the field, referred to as Hand-Crafted features. Each set of features was used to train three Machine Learning models (Random Forest, eXtreme Gradient Boosting, and RidgeClassifierCV) to classify six behaviours (drinking milk, grooming, lying, running, walking and *other*). 10 iterations from a validation set was used to tune each model with ROCKET, Catch22 and Hand-Crafted features extracted from various window sizes [3, 5 seconds] and overlap percentages [0, 25, 50%] between windows. For each feature set, the model achieving the best performance in the validation process was tested with its respective optimal window size and overlap using a test set composed of calves not used for model training. The best results were obtained with RidgeClassifierCV regardless of the features set. The highest performance was achieved with ROCKET (Balanced Accuracy: 0.81), followed by Catch22 (Balanced Accuracy: 0.74), both outperforming Hand-Crafted features (Balanced Accuracy: 0.66). These results highlight that the choice of accelerometer features must be considered as carefully as the models themselves. In particular, ROCKET features could help overcome current limitations, enabling the classification models to be used on farms to improve calf welfare.

1. Introduction

Improving pre-weaned calf welfare is highly important to prevent physiological changes and death that may happen due to prolonged exposure to stress (Koknaroglu and Akunal, 2013) but also to bridge the gap between farming and society (Cardoso et al., 2016). In that respect, monitoring pre-weaned calf behaviour would be promising as any change can signal underlying health concerns or environmental stressors (Mahendran et al., 2023; Nikkhah and Alimirzaei, 2023). Therefore, monitoring calf behaviour individually can be a way to identify stress factors (Dissanayake et al., 2022) and to make recommendations on routine practices that are less stressful for the calves (McPherson et al., 2022). However, behaviour monitoring must be conducted continuously for the targeted applications, which is not compatible with human observations that are time-consuming and labour-intensive (Penning, 1983). Sensors that can monitor livestock behaviour automatically have been increasingly used over the last few

* Corresponding author at: School of Computer Science, University College Dublin, Ireland.
E-mail address: oshana.dissanayake@ucdconnect.ie (O. Dissanayake).

https://doi.org/10.1016/j.compag.2025.110799
Received 19 April 2024; Received in revised form 26 May 2025; Accepted 22 July 2025
Available online 14 August 2025

years (Rushen et al., 2012; Jiang et al., 2023). In particular, onboard accelerometer sensors offer a continuous data stream in the three-dimensional plane from which a diverse range of livestock behaviours can be derived. Accelerometers are also adaptable and versatile, can function for a very long period, and can be integrated into various devices and systems (Martiskainen et al., 2009; Moreau et al., 2009; Tatler et al., 2018; Chakravarty et al., 2019; Yu et al., 2023). However, Riaboff et al. (2022) highlighted in a comprehensive review that the challenge lies in capturing a large spectrum of behaviours, including transitional or maintenance behaviours (Martiskainen et al., 2009; Vázquez Diosdado et al., 2015; Lush et al., 2018; Rodriguez-Baena et al., 2020). This is a major limitation as monitoring should encompass key behaviours to provide an accurate assessment of calf welfare (Duthie et al., 2021; Enríquez et al., 2010; Morisse et al., 1995). Furthermore, Riaboff et al. (2022) highlighted concerns about model genericity, as a substantial drop in performance is usually observed when the model is tested with unseen animals (Rahman et al., 2018). This limitation compromises the deployment in farms, as the models must be robust regardless of the animal.

The prevailing methodology employed in literature for livestock behaviour classification entails the segmentation of the accelerometer signal in fixed-size time windows (Riaboff et al., 2022). A set of features are then calculated in each time window to capture the dynamic properties of the time series, predominantly rooted in the time domain (Lush et al., 2018; Chang et al., 2022; Price et al., 2022; Dissanayake et al., 2022; Balasso et al., 2023). These features, hereafter referred to as Hand-Crafted features, are then used to train a Machine Learning model, usually RandomForest, Linear Discriminant Analysis, Support Vector Machines, Deep Neural Networks, or eXtreme Gradient Boosting (Riaboff et al., 2020; Kleanthous et al., 2022). While the performance of Machine Learning models has often been compared (Dutta et al., 2015; Smith et al., 2016; Hu et al., 2020), the accelerometer features used as inputs have received limited attention. However, it is recognized in applied Machine Learning research that success depends as much on feature engineering, i.e., the way the data is preprocessed and presented, rather than the actual Machine Learning models themselves (Sarker, 2021). In the livestock behaviour literature, Valletta et al. (2017) contend that getting the feature engineering right will almost always result in better predictive accuracy than employing more complex models. In this paper, we introduce two sets of features specifically designed for time-series classification, which have already demonstrated remarkable performance in other applications: the 22 CAnonical Time-series CHaracteristics (Catch22, Lubba et al. (2019)) and the Random Convolutional Kernel Transform features (ROCKET, Dempster et al. (2020)). We investigate the performance of ROCKET and Catch22 features in classifying pre-weaned calf behaviour using accelerometer data compared to the Hand-Crafted features traditionally used in the field, while taking into account the two limitations discussed above. To the best of our knowledge, this is the first study to apply these feature sets to classify livestock behaviour from accelerometer data.

2. Materials and methods

An overview of the methodology applied is described in Fig. 1. Each step is then detailed in the following sections. The Python code utilized for Sections 2.2–2.6 is available through an *open-source Github project*.[1] Data are available through the data paper associated to this article (Dissanayake et al., 2025).

2.1. Data collection

A comprehensive description of the data collection process is provided in the data article associated with this study. We therefore refer the reader to Dissanayake et al. (2025) for further details on data collection.

The experiment was carried out at Teagasc Moorepark Research Farm (Fermoy, Co. Cork, Ireland; 50°07′N; 8°16′W) from January 21 to April 5, 2022. Ethical approval was obtained from the Teagasc Animal Ethics Committee (TAEC; TAEC2021–319). The trial was carried out in accordance with the European Union (Protection of Animals Used for Scientific Purpose) Regulations 2012 (S.I. No. 543 of 2012). 47 Irish Holstein Friesian and Jersey pre-weaned calves housed in a group pen were utilized for the experiment (see Fig. 2), where they were fed using an automatic milk feeder at a rate of 6 L/calf/day with *ad libitum* access to hay, concentrates, and water. Calves were gradually weaned at 56 days using the automatic feeder.

Each calf was equipped with a tri-axis accelerometer data logger (Axivity LTD[2]) fastened to a neck collar starting from one week after birth until 2 months of age (see Fig. 3). The accelerometers were configured with a sensitivity of ±8g and a sampling rate of 25 Hz (battery life: 30 days). A NAND flash memory was used to store the data (memory: 512Mb). The accelerometers were measuring 23 × 32.5 × 7.6 mm and weighing 11g. Each sensor was wrapped in cling film and cotton wool and attached to the collar with vet wrap and insulating tape. The sensors were placed on the left side of the neck in the same orientation for all calves (see Fig. 3). Total weight of the device represents less than 0.01% of calf weight, which should not cause any disruption (Manning et al., 2017). The X-axis detected the top-bottom direction, the Y-axis detected the backwards-forward direction, and the Z-axis detected the left–right direction. The collars were tightly adjusted, and a 13 g metal ring was added to prevent them from moving from the designated side (Fig. 3). For the next ten weeks, collars were taken off every two weeks to retrieve data and replenish the battery. A set of DVRs (4-channel and 8-channel Hikvision, 1080p) and four 8Mp Dome CCTV cameras mounted in each pen were employed in addition to the accelerometer data collection to record videos of calves (see Fig. 2). 2092 h of videos were extracted at the end of the trial.

A subset of 30 calves out of the 47 who demonstrated a wide range of behaviours within the selected videos were targeted to create the dataset for this study. Labelling was done with the Behavioural Observation Research Interactive Software (BORIS) (Friard and Gamba, 2016) using an exhaustive ethogram with 24 behaviours based on Barry et al. (2019) after adaptation to this experiment. Annotations were carried out by three observers, ensuring that all 30 animals were observed for at least 15 min by at least one observer. The three observers annotated a one-hour video for one calf to measure the concordance between annotations every second based on the 24 behaviours (Cohen's Kappa averaged over the 1- video: $0.72+/-0.01$). A total of 27.4 h of observation has been performed over the 30 calves (age: $23.7+/-10.7$ days). Finally, five behaviours indicative of calf welfare each representing more that 1.7% of the whole dataset were kept in the study, i.e., lying (42.1%), drinking milk (9.23%), grooming (4.29%), running (2.24%) and walking (1.71%). All the other behaviours were merged into a class *other* (40.42%) and retained in the rest of the process. The definitions of each behaviour are listed in Table 1 and the amount of data collected for each behaviour is detailed in Table 2. The accelerometer time-series were aligned with the observations after synchronizing the accelerometer timestamps with the video timestamps.

2.2. Accelerometer data pre-processing

The computational workflow was executed on an Intel Xeon E-2378G CPU (2.80 GHz and based on x86_64 architecture) with 16 CPUs, eight cores per socket, and two threads per core with a Matrox G200eW3 GPU.

1 https://github.com/Oshana/calfBehaviourEval.

2 Axivity Ltd; https://axivity.com/product/ax3.

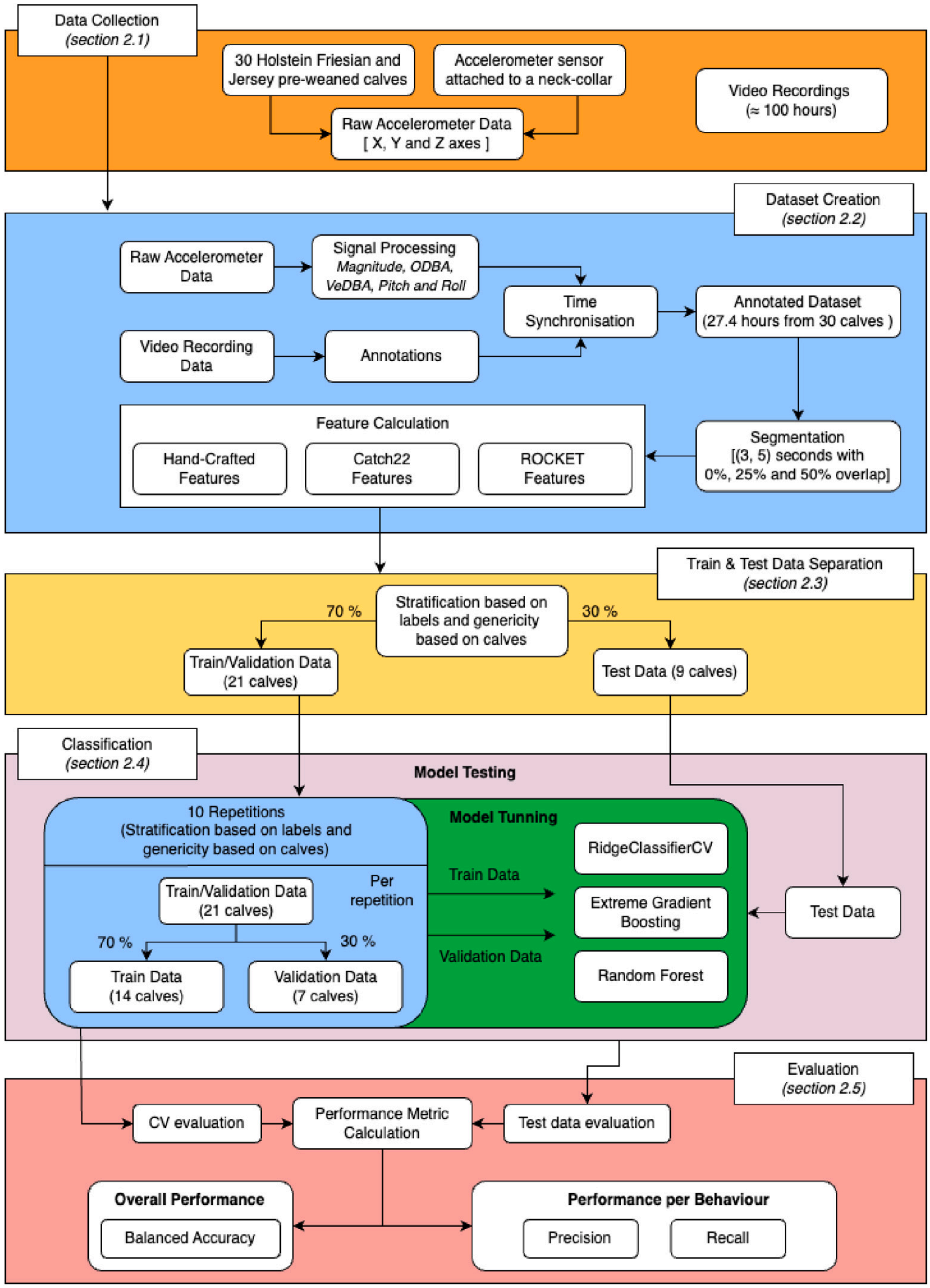


Fig. 1. Overview of the methodology applied to compare classification performance depending on the subset of features (Hand-Crafted, Catch22 and ROCKET) and evaluate model genericity across calves.

2.2.1. *Accelerometer time-series calculation*

The magnitude of the acceleration was computed from the raw X, Y and Z axes readings of the accelerometer, removing the gravity component of the acceleration (van Hees et al., 2013; da Silva et al., 2014). A 6th-order Butterworth high-pass filter with a cut-off frequency of 0.3 Hz (Smith et al., 2016) was applied to get the dynamic acceleration from which Overall Dynamic Body Acceleration (ODBA) (Wilson et al., 2008; Versluijs et al., 2023) and Vectorial Dynamic Body Acceleration (VeDBA) (Walker et al., 2015; Versluijs et al., 2023) were derived. A 6th-order Butterworth low-pass filter with a cut-off frequency of 0.3 Hz was applied to get the static acceleration from which pitch (Walker et al., 2015; Versluijs et al., 2023) and roll (Walker et al., 2015; Versluijs et al., 2023) were computed.

2.2.2. *Time-series segmentation with various configurations*

As the window size and the overlap percentage between windows may have a substantial impact on the classification performance (Riaboff et al., 2019), various pre-processing configurations were firstly investigated. The set of window sizes [3, 5 s] and overlap percentages [0%, 25%, 50%] chosen for this study was based on the recommendations made in Riaboff et al. (2022) to capture short, transient behaviours. This resulted in six pre-processing configurations being tested.

2.2.3. *Feature calculation*

ROCKET and Catch22 features were extracted from the eight time-series introduced in Section 2.2.1 with each of the 6 pre-processing

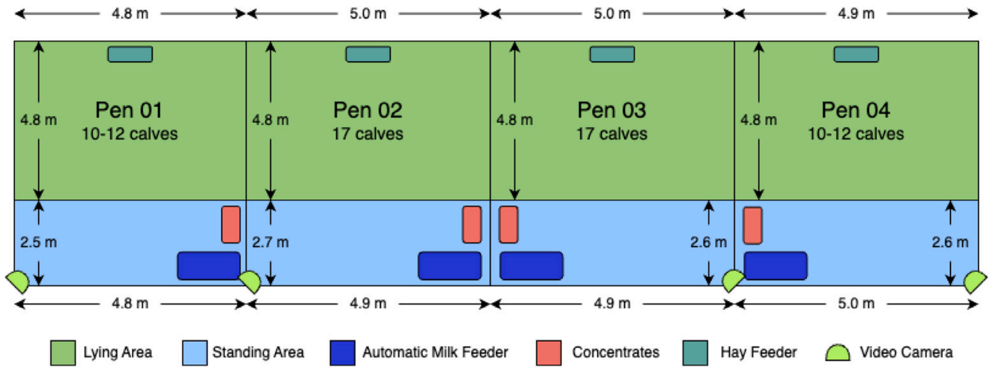


Fig. 2. Design of the group pens used after 3–7 days old of age with cameras placement to film the calves in each pen.

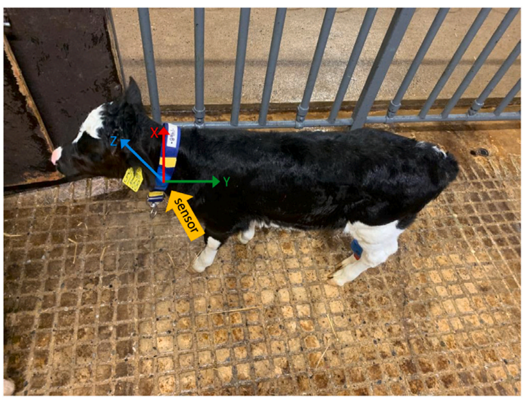


Fig. 3. The collar used to attach the accelerometer sensor and the orientation of the three axes. The X-axis detected the top-bottom direction, the Y-axis detected the backwards-forward direction, and the Z-axis detected the left–right direction.

Table 1
Definition of the different behaviours (Barry et al., 2019) in the ethogram used for that study after merging the least observed behaviours (<1.7%) in the *other* class.

Behaviour	Definition
Drinking milk	Calf is drinking milk from the milk feeder.
Grooming	Calf uses tongue to repeatedly lick own back, side, leg, tail areas.
Lying	Calf is resting either sternally or laterally with all four legs hunched close to body either awake or asleep.
Running	Calf is running (play/not-play).
Walking	Calf is walking or shuffling about.
Other	A collection of 19 other behaviours including rising, lying-down, social interaction, play etc.

Table 2
Summary of data collected: duration of observation, number of sequences associated with each behaviour, and number of calves on which each behaviour was observed.

Behaviour	Total duration (min)	Number of segments	Number of calves
Drinking milk	143.68	169	27
Grooming	75.26	334	29
Lying	650.36	120	27
Running	55.10	608	24
Walking	44.29	561	30
Other	671.63	2636	30

configurations presented in Section 2.2.2. Additionally, a set of features commonly used in the field, hereafter referred to as Hand-Crafted (HC) features, was computed as a baseline for performance comparison.

1. **Hand-Crafted features**
   Hand-Crafted (HC) features calculated in livestock ruminant classification behaviour from accelerometer data are usually extracted in the time domain and frequency domain using the Fourier Transform. Those features provide information on the motion intensity (e.g., median, quartiles, maximum, motion variation from ODBA and VeDBA), the orientation of the animal's body (e.g., pitch and roll), the shape of the signal distribution (e.g., skewness and kurtosis) and the physical properties of the movement (periodic, stochastic, etc.) (e.g., spectral entropy, fundamental frequency, etc.) (Riaboff et al., 2022). In this study, we calculated mean, median, minimum, maximum, standard deviation, first quartile, third quartile, skewness, kurtosis, entropy and motion variation. Across the 8 time-series, this led to a set of 88 HC features typically calculated in the field. This feature set was used as the reference to evaluate the potential additional value of Catch22 and ROCKET. Features, extraction domain, equations and references are displayed in Table 3.
2. **Catch22 features**
   Catch22 are 22 CAnonical Time-series CHaracteristics derived from a large pool of 4791 features that showed exceptional performance across 93 time-series classification datasets. This subset was also selected to minimize redundancy between features, thus providing complementary information on the time-series while reducing computational expenses (Lubba et al., 2019). The Catch22 features cover auto-correlation, value distributions, outliers, and fluctuation scaling properties. This set can be increased to 24 features by including mean and standard deviation to capture the location and spread of the raw time-series distribution in the classification process. As that information may be necessary for the time-windows classification into calf behaviour, the set of 24 features was calculated, leading to a subset of 192 features across the 8 time-series. For the remainder of the paper, we are keeping the name *Catch22* as soon as we refer to that features set. Feature names and their description are displayed in Table 4.
3. **ROCKET features**
   ROCKET (RandOm Convolutional KErnel Transform) is a pioneering technique in time-series classification, known for its efficiency and efficacy in features extraction. ROCKET uses Random Convolutional Kernels, where kernels have random lengths, weights, biases, dilations, and paddings, to transform raw time-series into high-dimensional feature representations. ROCKET features are specifically the maximum and proportion of positive values resulting from each convolution, meaning that for k kernels, $2k$ features per time-series are produced. Random kernels capture diverse temporal patterns without domain-specific knowledge. For this study, we have utilized *miniROCKET* which leverages a more streamlined set of convolutional kernels, focusing on a predefined subset of kernel parameters rather than

Table 3
Hand-Crafted features calculated as the reference: Equations, domain and references.

Feature	Domain	Equation	Reference
Mean	Time	Mean = $\frac{\sum_1^N X_i}{N}$ N = Total number of observations.	Preece et al. (2008) Drover et al. (2017) Tatler et al. (2018)
Median	Time	$M = \begin{cases} X_{\frac{n+1}{2}}, & \text{if } n \text{ is odd} \\ \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}, & \text{if } n \text{ is even} \end{cases}$ X = Data points in the sorted list.	Preece et al. (2008) Figo et al. (2010) Fida et al. (2015)
Minimum	Time	$\text{Min}(X) = \min(X_1, X_2, \ldots, X_n)$ X = Considered set of data points.	Figo et al. (2010) Barwick et al. (2018)
Maximum	Time	$\text{Min}(X) = \max(X_1, X_2, \ldots, X_n)$ X = Considered set of data points.	Figo et al. (2010) Barwick et al. (2018)
Standard deviation	Time	$\sigma = \sqrt{\frac{\sum(x_i-\mu)^2}{N}}$ N = size of the data segment. x_i = Individual observation in the data set. μ = mean of the data segment.	Preece et al. (2008) Figo et al. (2010) Bersch et al. (2014) Drover et al. (2017)
First quartile	Time	$Q1 = \frac{N+1}{4}$ N = Total number of observations in the data set.	Preece et al. (2008) Zdravevski et al. (2017) Fan et al. (2019)
Third quartile	Time	$Q3 = \frac{3}{4}(N+1)$ N = Total number of observations in the data set.	Preece et al. (2008) Zdravevski et al. (2017) Fan et al. (2019)
Spectral entropy	Frequency	$H(s, sf) = \sum_{f=0}^{f_s/2} P(f) log_2[P(f)]$ Where P is the normalized power spectral density, and f_s is the sampling frequency.	Preece et al. (2008) Riaboff et al. (2020) Aziz et al. (2021) Dissanayake et al. (2022)
Motion variation	Time	$MV = \frac{1}{M}(\sum_{i=1}^{M-1} \lvert a_{x,i+1} - a_{x,i}\rvert + \sum_{i=1}^{M-1} \lvert a_{y,i+1} - a_{y,i}\rvert + \sum_{i=1}^{M-1} \lvert a_{z,i+1} - a_{z,i}\rvert)$	Riaboff et al. (2020) Fogarty et al. (2020) Dissanayake et al. (2022)
Skewness	Time	$\gamma = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Y_i-\mu}{\sigma}\right)^3$ γ = skewness N = number of variables in the distribution μ = mean of the distribution σ = standard deviation	AlZubi et al. (2014) Hounslow et al. (2019) Cabezas et al. (2022)
Kurtosis	Time	$\beta = \frac{\frac{1}{N}\sum_{i=1}^{N}(Y_i-\mu)^4}{V^2}$ β = kurtosis N = number of variables in the distribution μ = mean of the distribution V = variance of the dataset	AlZubi et al. (2014) Hounslow et al. (2019) Cabezas et al. (2022)

Table 4
Catch22 features (including mean and standard deviation (Catch24)) for time-series classification problem (Lubba et al., 2019).

Feature name	Description
DN_HistogramMode_5	Mode of z-scored distribution (5-bin histogram)
DN_HistogramMode_10	Mode of z-scored distribution (10-bin histogram)
SB_BinaryStats_mean_longstretch1	Longest period of consecutive values above the mean
DN_OutlierInclude_p_001_mdrmd	Time intervals between successive extreme events above the mean
DN_OutlierInclude_n_001_mdrmd	Time intervals between successive extreme events below the mean
CO_f1ecac	First 1/e crossing of autocorrelation function
CO_FirstMin_ac	First minimum of autocorrelation function
SP_Summaries_welch_rect_area_5_1	Total power in lowest fifth of frequencies in the Fourier power spectrum
SP_Summaries_welch_rect_centroid	Centroid of the Fourier power spectrum
FC_LocalSimple_mean3_stderr	Mean error from a rolling 3-sample mean forecasting
CO_trev_1_num	Time-reversibility statistic, $(xt + 1 - xt)3t$
CO_HistogramAMI_even_2_5	Automutual information, m = 2, τ = 5
IN_AutoMutualInfoStats_40_gaussian_fmmi	First minimum of the automutual information function
MD_hrv_classic_pnn40	Proportion of successive differences exceeding 0.04σ (Mietus et al., 2002)
SB_BinaryStats_diff_longstretch0	Longest period of successive incremental decreases
SB_MotifThree_quantile_hh	Shannon entropy of two successive letters in equiprobable 3-letter symbolization
FC_LocalSimple_mean1_tauresrat	Change in correlation length after iterative differencing
CO_Embed2_Dist_tau_d_expfit_meandiff	Exponential fit to successive distances in 2-d embedding space
SC_FluctAnal_2_dfa_50_1_2_logi_prop_r1	Proportion of slower timescale fluctuations that scale with DFA (50% sampling)
SC_FluctAnal_2_rsrangefit_50_1_logi_prop_r1	Proportion of slower timescale fluctuations that scale with linearly rescaled range fits
SB_TransitionMatrix_3ac_sumdiagcov	Trace of covariance of transition matrix between symbols in 3-letter alphabet
PD_PeriodicityWang_th0_01	Periodicity measure of Wang et al. (2007)
DN_Mean	Mean
DN_Spread_Std	Standard deviation

the exhaustive randomness of ROCKET. This refinement allows miniROCKET to achieve similar or even superior classification performance with a significantly reduced computational footprint (Dempster et al., 2021). The default number of kernels (10 000) was used during the feature generation in that study. ROCKET features were calculated from each of the 8 time-series, leading to a set of 9996 features.

2.3. Partitioning the dataset into training and testing calf-independent sets

The dataset was split into a 70:30 calf ratio (Fig. 1): Out of 30 calves, 21 were chosen for the training set, and 9 were chosen for the test set. This split ensures that the calves used for testing the model have not been used for model training to evaluate the model genericity. Furthermore, stratification has been applied to the annotated behaviours to maintain a consistent proportional distribution between the train and test set. The optimal split used in the rest of the study is the one that minimizes the mean of the differences in the proportions of each of the annotated behaviours between the training and testing sets.

2.4. Modelling with machine learning models

Three Machine Learning models were used with each feature set to (1) evaluate whether certain set of features perform better with a specific model and to (2) compare the performance of the best model respectively got for HC, Catch22, and ROCKET features. EXtreme Gradient Boosting (**XGB**), RandomForest **(RF)** and RidgeClassifierCV **(RCV)** algorithms were applied, considering their high performance in classification problems in various domains.

1. **eXtreme Gradient Boosting**
The eXtreme Gradient Boosting (XGB) algorithm is an ensemble Machine Learning technique that uses gradient boosting techniques to improve model accuracy (Friedman, 2001). It constructs decision trees sequentially, with each tree correcting previous errors to enhance model accuracy. The algorithm uses gradient boosting, where each tree is trained using the gradient of the loss function, minimizing the difference between predicted and actual values. XGB incorporates techniques to prevent overfitting, such as regularization terms in the objective function. Below are a few most prominent hyper-parameters (Chen and Guestrin, 2016). Definitions of the main hyper-parameters and the values tested in this study are shown in Table 5.
2. **Random Forest**
The Random Forest algorithm is a versatile ensemble learning method for classification tasks (Breiman, 2001). It constructs multiple decision trees during training and outputs the mode of classes or mean prediction of the individual trees. Randomness is introduced through bootstrap sampling and a random subset of features at each node, making the model more robust and preventing overfitting. Definitions of the main hyper-parameters and the values tested in this study are shown in Table 5.
3. **RidgeClassifierCV**
RidgeClassifierCV is useful in large-variable scenarios and aims to prevent overfitting while maintaining a balance between bias and variance. It is efficient for high-dimensional datasets and is commonly used in problems requiring interpretability and prediction accuracy. RidgeClassifierCV (Pedregosa et al., 2011) is an extension of the RidgeClassifier based on the theory of Ridge Regression (Hoerl and Kennard, 1970), which is an extension of linear regression that uses cross-validation to determine the optimal regularization parameter. This method splits the dataset into subsets and evaluates the model's performance for different regularization parameter (alpha) values. Definitions of the main hyper-parameters and the values tested in this study are shown in Table 5.

2.5. Optimal pre-processing and tuned model for each set of features and performance evaluation

The Machine Learning models were developed using scikit-learn (v.1.2.2) (Pedregosa et al., 2011), XGBoost (v.1.6.1) (Chen and Guestrin, 2016), and skTime (v.0.24.0) (Löning et al., 2019) available under Python v.3.9.7. For each of the three sets of features and the six pre-processing configurations, the three Machine Learning models were tuned using the training data. A grid search was used to identify the best hyper-parameters (see Table 5). For that purpose, each model was trained with one of the combinations of hyper-parameters in the grid search using 14 calves from the 21 available in the training dataset. The model was then evaluated using a validation dataset made up of the 7 remaining calves in the training set. The process has been iterated 10 times (see Fig. 1). For each of the three feature sets, the pre-processing configuration and the tuned Machine Learning model that achieved the highest average Balanced Accuracy (BA; see Eq. (1)) over 10 iterations were selected for the rest of the study. The optimal pre-processing and tuned model for each set of features was then used to assess the classification performance on the test set. BA was used as a global metric of performance, along with precision and recall as a performance metric per behaviour (see Eqs. (2) and (3), respectively).

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

where **TP** is the number of True Positive, **FN** is the number of False Negative, **FP** is the number of False Positive, and **TN** is the number of True Negative.

2.6. Relative feature importance

Relative Feature Importance was finally explored to find out HC and Catch22 features that were actually most useful in model classification. A RandomForest classifier was trained (excluding the 9 test calves) and the RandomForest feature importance[3] scores were computed to evaluate the contribution of individual features. It should be noted that this analysis was not applied to ROCKET, as it was not appropriate for those features.

3. Results

3.1. Performance from the validation set

BA obtained from the 10 iterations with the validation set for the Hand-Crafted, Catch22 and ROCKET features combined with RF, XGB, and RCV classifiers across the six pre-processing configurations, are displayed in Appendix (Fig. A.9). The highest BA of 0.78 was achieved with ROCKET features combined with RCV, using a 5-s window combined with 0% and 25% overlap. For each set of features, the optimal pre-processing configuration, Machine Learning model, the best hyper-parameters identified from the validation set and the associated BA are presented in Table 6 and in Appendix (Fig. A.10). The optimal pre-processing and tuned Machine Learning model for each feature set were used for performance evaluation on the calf-independent test set in the next section.

3.2. Performance on the test set for each set of features

3.2.1. Overall performance

The highest performance was achieved with ROCKET, followed by Catch22, with BA scores of 0.81 and 0.74, respectively (Fig. 4). Both outperformed HC features, achieving a BA of 0.66.

[3] https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html.

Table 5
Classifiers, hyper-parameters, definitions and values tested.

Classifier	Hyper-parameters	Definition	Values tested
XGB	n_estimators	Number of trees to be built.	100, 200, 300
	eta (learning rate)	Determines the step size at each iteration while moving towards a minimum of the loss function.	0.01, 0.05, 0.1, 0.2
	max_depth	Sets the maximum depth of a tree.	None, 3, 6, 10, 12
RF	n_estimators	Number of trees in the forest	100, 200, 300
	max_depth	Maximum depth of the tree	None, 10, 20, 30
	min_samples_split	Minimum number of samples required to split an internal node	2, 5, 10
	criterion	Function to measure the quality of a split	gini, entropy
	class_weight	Weights associated with classes	None, balanced
RCV	fit_intercept	Determines whether a bias (intercept) term is added to the decision function.	True, False
	class_weight	Weights associated with classes	None, balanced
	alphas	Determines the regularization strength	np.logspace(−3,3,100), np.logspace(−1,10,100)

Table 6
Features set, optimal pre-processing, best classifier, best hyper-parameters and associated balanced accuracy obtained from the validation set.

Features set	Optimal pre-processing [WS (s); overlap (%)]	Best classifier	Best hyper-parameters	Best BA
Hand-Crafted	[5 s; 0%]	RCV	class_weight: balanced fit_intercept: True alphas: np.logspace(−1, 10, 100)	0.66
Catch22	[5 s; 50%]	RCV	class_weight: balanced fit_intercept: True alphas: np.logspace(−1, 10, 100)	0.72
ROCKET	[5 s; 25%]	RCV	class_weight: balanced fit_intercept: True alphas: np.logspace(−1, 10, 100)	0.78

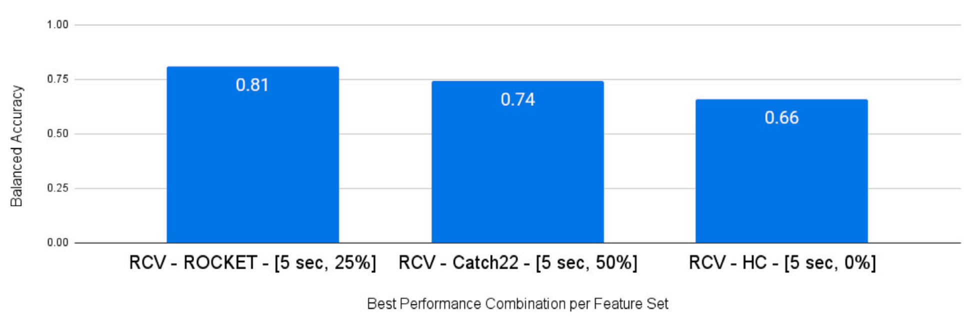


Fig. 4. Balanced accuracy obtained with the best performing pre-processing configuration and Machine Learning model for Hand-Crafted, Catch22 and ROCKET features, using the calf-independent test-set.

3.2.2. Performance per behaviour

Precision and recall obtained from tuned RCV classifier with HC, Catch22 and ROCKET features using their respective optimal pre-processing configuration are displayed for each behaviour in Fig. 5. Best performance were systematically obtained with ROCKET, except for the recall of walking and the recall of running. Notably, ROCKET led to a substantial improvement in detecting the drinking milk behaviour, with precision increasing by 48.9% and 35.6%, and recall improving by 44.7% and 25.5%, compared to HC and Catch22, respectively. The same conclusion can be drawn to a lesser extent with the “other” behaviour as the recall was increased by 22.0% and 18.3% compared to HC and Catch22, respectively. The lowest performance were obtained with HC features in most cases, although the best recall for walking was obtained with this set of features.

Confusion matrices for each set of features are displayed in Fig. 6. A comparison across behaviours further revealed that running and lying were consistently well-predicted across all feature sets, with precision and recall exceeding 85%. On the contrary, grooming is tricky to predict whatever the features set, reaching a maximum precision of 44.1% and maximum recall of 78.7% with ROCKET. Similarly, the precision for walking did not exceed 33.2% with ROCKET while the recall of “other” did not exceed 56.8%. The main confusions for grooming occurred with “other” and drinking milk, as these behaviours were often misclassified among them. For walking, there were frequent misclassifications with “other”, indicating some overlap in movement patterns. The “other” category had significant confusion with lying and grooming, suggesting that some behaviours in this group were not distinct enough from these activities.

3.2.3. Calculation time

Time for feature extraction, training and testing were evaluated for HC, Catch22 and ROCKET using their respective optimal pre-processing configuration and tuned Machine Learning model (Table 7). The shortest feature extraction time was obtained with ROCKET (80.49 s),

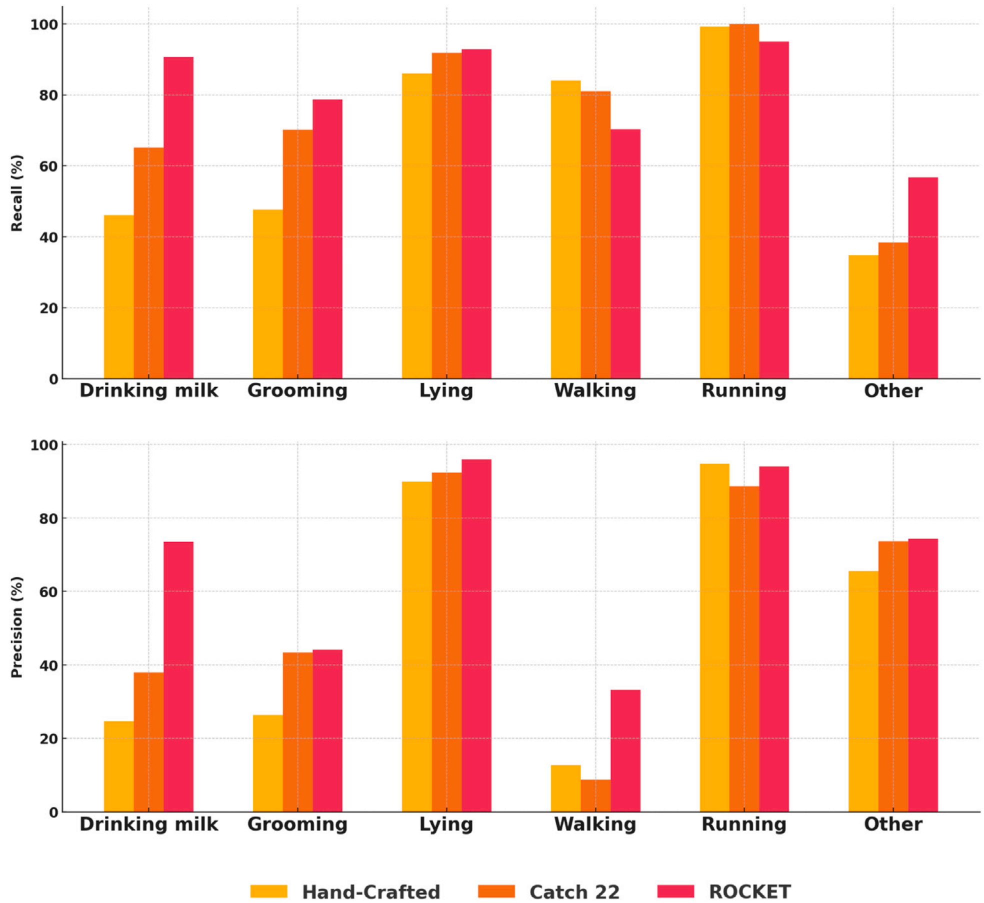


Fig. 5. Recall (top) and Precision (bottom) obtained from the tuned Ridge ClassifierCV models using Hand-Crafted, Catch22 and ROCKET features with their respective optimal pre-processing configurations.

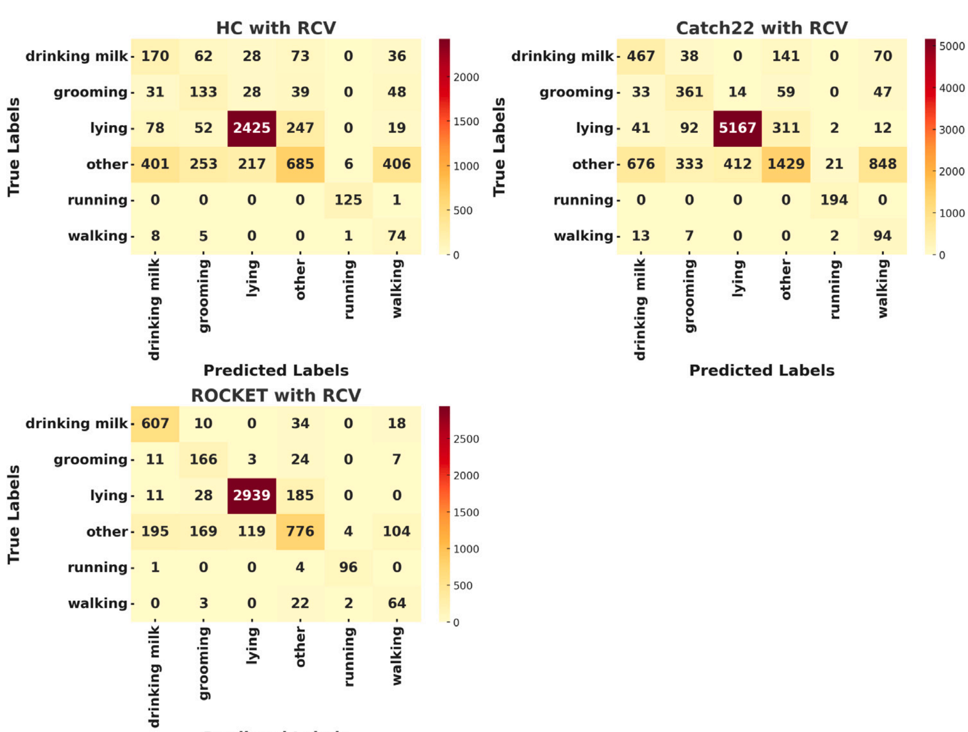
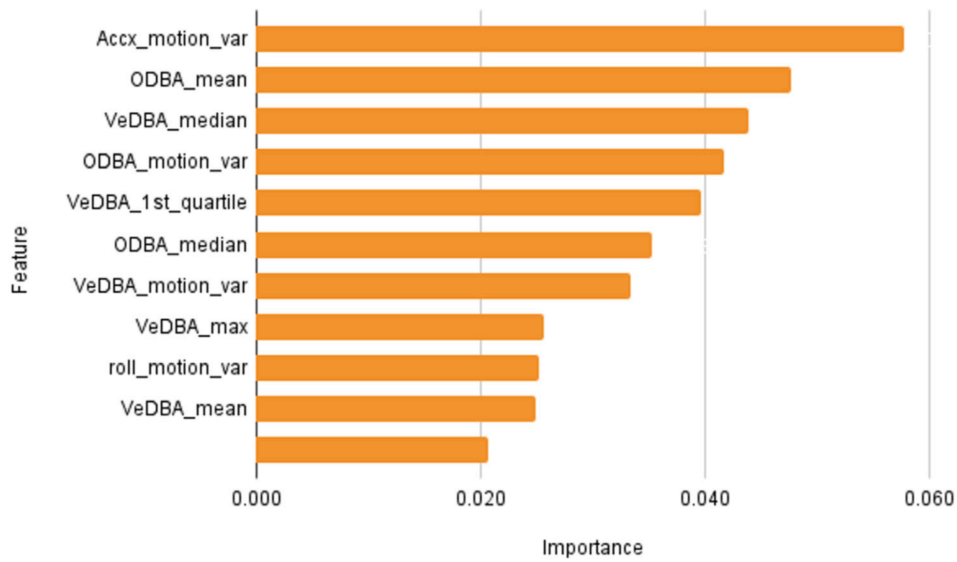


Fig. 6. Confusion matrices obtained for Hand-Crafted (top left), Catch22 (top right) and ROCKET (bottom left) features with the best pre-processing configurations and Machine Learning models, using the calf-independent test.

Table 7
Calculation time (s) for feature extraction, model training and testing for the best configuration for each feature set.

	RidgeClassifier - ROCKET 5 s, 25% overlap	RidgeClassifier - Catch22 5 s, 50% overlap	RidgeClassifier - HC 5 s, 0% overlap
Feature extraction	80.49	290.45	983.38
Training	205.69	0.23	0.06
Testing	0.57	0.01	0.01



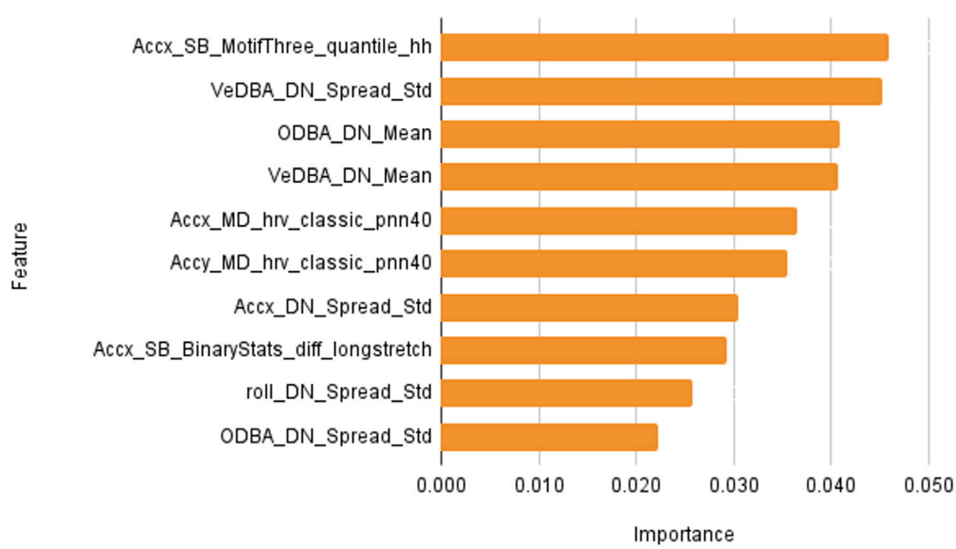


Fig. 7. Top 10 Hand-Crafted features (top) and Catch22 (bottom) with the highest importance in the classification. We refer to Tables 3 and 4 for the description of the Hand-Crafted and Catch22 features, respectively.

following by Catch22 (290.45 s) and HC features (983.38 s). However, training RCV was much longer with ROCKET (205.69 s) compared to Catch22 (0.23 s) and HC (0.06 s) but model testing was still <1 s regardless of the features set. In total, the full procedure for features extraction, model training and testing lasted 286.75 s for ROCKET, 290.69 s for Catch22 and 983.45 s for HC with their respective optimal pre-processing configuration and tuned RCV model.

3.3. Relative feature importance

The top-ranked features based on RandomForest Feature Importance for the HC and Catch22 feature sets are illustrated in Fig. 7.

The X-axis in the raw accelerometer data as well as the dynamic components of the acceleration (OBDA and VeDBA) emerged as the most critical time-series in the classification for both HC and Catch22 features. This result was actually expected as the X-axis detects motion in the top-bottom direction, thus enabling to discriminate behaviours based on the position of the head, while the dynamic components should help to discriminate between active and inactive behaviours.

Among the HC features, motion variation (*Accx_motion_var* (0.058)) is the most important features, indicating that the level of motion fluctuation within the time-series help to distinguish between steady patterns and those that exhibit abrupt changes. This is also observed for Catch22, where the feature indicative of high fluctuation (*MD_hrv_classic_pnn40*) is also highly important for the classification of calf behaviours. Basic statistics such as the mean, median and standard-deviation also belong to the most important features for both HC ($ODBA\_mean$ (0.048), $VeDBA\_median$ (0.044)) and Catch22 (DN_Mean, DN_Spread_Std), suggesting that information about signal distribution in the time domain is highly important for the classification. Interestingly, the most important feature in Catch22 is related to the level of predictability of the time-series ($SB\_MotifThree\_quantile\_hh$), while similar feature in HC (spectral entropy) is not included in the top 10 ranking. This result suggests that symbolic representation of the time-series, as done for the calculation of the $SB\_MotifThree\_quantile\_hh$ feature, may help to discriminate between time-series based on their level of predictability.

4. Discussion

4.1. ROCKET: Boosting classification accuracy compared to Hand-Crafted features

The main objective was to assess the benefits of features specifically designed for time-series to classify calf behaviour from accelerometer data, rather than focusing on the Machine Learning model itself, as this topic has already been extensively explored in the field. In that regard, we introduced ROCKET and Catch22 features and we evaluated their ability to classify 6 pre-weaned calf behaviours. Performance were compared to a subset of features typically used in the field (HC features). As expected, the best performance with an independent calf test-set was obtained with ROCKET (BA: 0.81), well ahead of Hand-Crafted features (BA: 0.66) and, to a lesser extent, Catch22 (BA: 0.74). This result confirms the importance of focusing on model input features as much as the Machine Learning models themselves (Sarker, 2021). Considering that ROCKET can capture local and global patterns in the time-series and achieves better performance than cutting-edge classification techniques in the applied Machine learning domain, this finding is consistent with the literature (Dempster et al., 2020). Notably, compared to HC features, ROCKET substantially enhanced the predictive performance for drinking milk (precision: +48.90%; recall: +44.60%), grooming (precision: +17.80 %; recall: +31.00%) and the other class (precision: +8.70%; recall: +22.00%). Drinking milk is a repetitive sequence starting with milk suckling from the automatic feeder, following by swallowing. Similarly, grooming begins with a movement of the head to reach the area of the body to be licked, a sequence of licking movements and a return movement of the head. There is therefore a temporal structure to the movements of the jaw and head which must be reflected in the accelerometer time-series, as illustrated in Fig. 8. Finding that ROCKET is more adapted than Catch22 or HC in that context is thus coherent with Lubba et al. (2019), who showed that shape-based classifier, such as ROCKET combined with RCV, can accurately capture class differences in the time-series shape. Therefore, our study confirms that ROCKET features must be considered for classifying livestock ruminant behaviour from accelerometer data in further studies to address the current limitations in the field.

4.2. A robust classification framework for pre-weaned calf behaviour from accelerometer data

The second objective was to develop a robust classification framework for pre-weaned calf behaviour based on accelerometer data, which could be applied to calf welfare monitoring on dairy farms. In that regard, ROCKET combined to RCV was able to predict accurately drinking milk (precision: 73.60%; recall: 90.70%), lying (precision: 96.00%; recall: 92.90%) and running (precision: 94.10%; recall: 95.00%), regardless of the individual. This is a major contribution to the field, as few studies have focused on classifying the behaviour of pre-weaned calves, and the genericity of the models from one calf to another has not been tested. Indeed, Carslake et al. (2020) achieved over 90% accuracy in classifying locomotor play, self-grooming, ruminating, non-nutritive suckling, nutritive suckling and active lying using Adaboost. However, their model's performance was not evaluated on unseen calves, as all the 13 calves used for model training were also included in the Cross-Validation procedure applied for performance assessment. Furthermore, it is worth highlighting that the performance obtained for drinking milk are remarkable, as drinking is tricky to classify from accelerometer data in cattle. Indeed, Hosseininoorbin et al. (2021) achieved a maximum F1-score of 19.58% for drinking, despite the robust architecture based on time-frequency data representation combined with Deep Neural Network. This is also a major added value for the dairy calf welfare domain, as drinking milk is one of the most important behaviours for evaluating health and welfare in pre-weaned calves (Duthie et al., 2021). Although the precision for walking (precision: 33.2%) and grooming (precision: 44.10%) is too low for the targeted applications, merging these behaviours into the “other” class resulted to a high overall BA of 88%, with a great recall and precision for each of the 4 behaviours (>75%; data not shown). ROCKET combined to RCV with the optimal pre-processing configuration [5-s window; 25% overlap] therefore formed a robust classification framework to monitor continuously drinking milk, lying, running and other behaviours from accelerometer data in pre-weaned calves. To the best of the author's knowledge, this is the first ready-to-use classification framework for monitoring pre-weaned calf behaviour from accelerometer collars in dairy farms.

4.3. Limitations and perspective for improvement

ROCKET with RCV achieved a BA of 0.81 with reliable predictive performance for drinking milk, lying and running behaviours. However, a substantial number of False Positive was observed for walking and grooming, leading to low recall (33.20% and 44.10%, respectively). This is explained by the confusion between walking, grooming and other (see Fig. 6), likely due to the absence of distinct dynamic characteristics and patterns (see Fig. 8), which is consistent with the literature. Indeed, Hosseininoorbin et al. (2021) also obtained a low predictive performance for walking (F1-score: 32.19) and grooming (F1-score: 15.60) despite time-frequency data representation combined with Deep Neural Network. This is a limitation for calf welfare assessment as a change in these behaviours can be indicative of poor welfare. For example, Enríquez et al. (2010) observed a peak of seeking and walking the day after the separation from the dams. Fewer self-grooming and more scratching were also recorded in the few hours after dehorning (Morisse et al., 1995). It should be noted that walking and grooming represent only 1.71% and 4.29% of the total annotations, respectively (see Table 2), that may not be enough to find out distinctive patterns with ROCKET. The difficulty lies in the fact that walking and grooming are not frequent in pre-weaned calves and the duration of these behaviours is fairly short. For walking behaviour, annotation amount could be boosted by setting up a trial in which calves are forced to walk through a corridor, rather than relying on randomly occurring instances in video recordings. However, this approach may not be applicable to grooming behaviour. In that regard, implementing data augmentation techniques to artificially increase the amount of grooming sequences in the dataset could be investigated (Li et al., 2021). Regarding modelling, feature selection could be applied based on the most important features from our study (see Fig. 7). Another assumption is that a large spectrum of behaviours is difficult to classify as a whole by a single model (Hosseininoorbin et al., 2021). In that respect, a classification framework based on binary classifiers is worth exploring (Arablouei et al., 2021; Smith et al., 2016). Finally, it may be possible that accelerometer data alone are not informative enough to discriminate behaviours with a fine grain. Combining accelerometer data with computer vision could enhance the classification of behaviours influenced by the structure of the barn, such as drinking milk, eating concentrates, eating hay, and walking (Vayssade et al., 2023). The complementary information provided by individual animal tracking could substantially reduce confusion between behaviours.

Fig. 8. Random 3 s time-windows selected from a set of random calves for each behaviour. Drinking milk and grooming behaviour have a subtle shape in the time-domain that is almost phase-aligned.

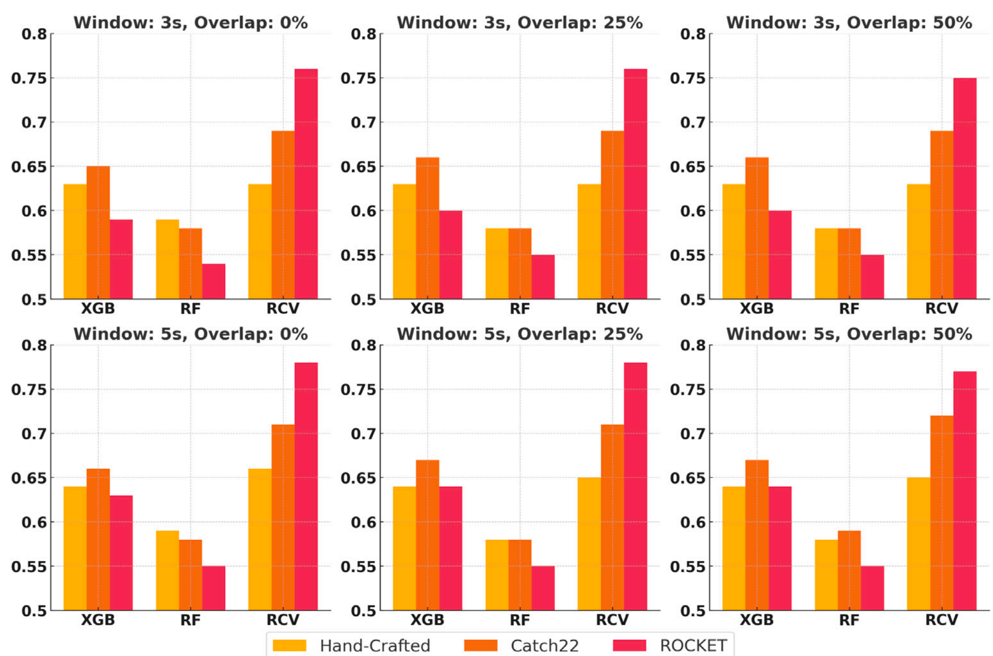


Fig. A.9. Validation stage BA scores from the 10 iterations with the validation set across model, window size, and overlap combinations.

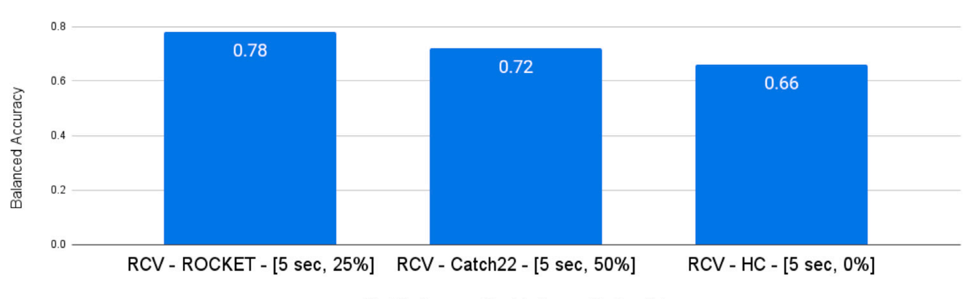


Fig. A.10. Performance obtained with the best performing pre-processing configuration and Machine Learning model for Hand-Crafted, Catch22 and ROCKET features, using the 10 iterations from the validation set.

4.4. Implication for calf welfare assessment

Our classification framework, based on ROCKET features and RCV classifier, enables continuous monitoring of key pre-weaned calf behaviours. It could be used to measure the impact of routine practices (dehorning, transport, separation from the dams, etc.) on calf behaviour to implement new practices that promote calf welfare in dairy farms. The classification framework could also be applied to quantify the deviance in calf behaviour linked to sickness or distress. Furthermore, although accelerometer data retrieval is manual in this study, automatic data transmission could be done when calves are in a designated area, such as the automatic milk feeder. Calculating ROCKET features from 5-s windows and predicting behaviours using tuned RCV classifier is pretty fast (<1.5 min; Table 7). Therefore, predicted behaviours from our classification framework could further be used as inputs of change point detection algorithms to detect distress or sickness in real-time in dairy commercial farms.

5. Conclusion

This study confirms the importance of focusing on model input features as much as the Machine Learning models themselves. Indeed, ROCKET features combined with RidgeClassifierCV resulted on a substantial improvement compared to features commonly used in the field (BA: +0.15). To the best of the author's knowledge, this is the first study in the field to highlight the benefits of incorporating accelerometer features specifically designed for time-series classification in related domains. Our classification framework based on ROCKET with RidgeClassifierCV ensures an accurate and robust prediction of the main pre-weaned calf behaviours (Balanced Accuracy: 0.81), including drinking milk. Although improvement is still required to improve the performance of walking and grooming, it is one of the first ready-to-use system to monitor the main behaviours of pre-weaned calves, regardless of the individual. Therefore, it could be used to evaluate the effect of practices on calf welfare, such as dehorning, transport, weaning from the dams, etc., using accelerometer data. In the longer term, our classification framework could support the development of automated tools for detecting anomalies in calves by measuring changes in behaviour in real-time.

CRediT authorship contribution statement

Oshana Dissanayake: Writing – review & editing, Visualization, Software, Conceptualization, Writing – original draft, Validation, Methodology. **Sarah E. McPherson:** Writing – review & editing, Data curation. **Joseph Allyndrée:** Writing – review & editing, Software. **Emer Kennedy:** Writing – review & editing, Data curation. **Pádraig Cunningham:** Writing – review & editing, Supervision, Validation, Methodology. **Lucile Riaboff:** Validation, Methodology, Writing – review & editing, Supervision, Data curation.

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.

Acknowledgments

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) and the Department of Agriculture, Food and the Marine on behalf of the Government of Ireland to the VistaMilk SFI Research Centre under Grant Number 16/RC/3835.

Appendix

See Figs. A.9 and A.10.

Data availability

We have published an article detailing the data used in this study to Data in Brief journal (https://doi.org/10.1016/j.dib.2025.111462), and have released the code utilized for this work through this article itself.

References

AlZubi, H.S., Gerrard-Longworth, S., Al-Nuaimy, W., Goulermas, Y., Preece, S., 2014. Human activity classification using a single accelerometer. In: 2014 14th UK Workshop on Computational Intelligence. UKCI, IEEE, pp. 1–6.

Arablouei, R., Currie, L., Kusy, B., Ingham, A., Greenwood, P.L., Bishop-Hurley, G., 2021. In-situ classification of cattle behavior using accelerometry data. Comput. Electron. Agric. 183, 106045.

Aziz, W., Hussain, L., Khan, I.R., Alowibdi, J.S., Alkinani, M.H., 2021. Machine learning based classification of normal, slow and fast walking by extracting multimodal features from stride interval time series. Math. Biosci. Eng. 18, 495–517.

Balasso, P., Taccioli, C., Serva, L., Magrin, L., Andrighetto, I., Marchesini, G., 2023. Uncovering patterns in dairy cow behaviour: A deep learning approach with tri-axial accelerometer data. Animals 13 (1886).

Barry, J., Kennedy, E., Sayers, R., de Boer, I.J., Bokkers, E., 2019. Development of a welfare assessment protocol for dairy calves from birth through to weaning. Anim. Welf. 28, 331–344.

Barwick, J., Lamb, D.W., Dobos, R., Welch, M., Trotter, M., 2018. Categorising sheep activity using a tri-axial accelerometer. Comput. Electron. Agric. 145, 289–297.

Bersch, S.D., Azzi, D., Khusainov, R., Achumba, I.E., Ries, J., 2014. Sensor data acquisition and processing parameters for human activity classification. Sensors 14, 4239–4270.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

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).

Cardoso, C.S., Hötzel, M.J., Weary, D.M., Robbins, J.A., von Keyserlingk, M.A., 2016. Imagining the ideal dairy farm. J. Dairy Sci. 99, 1663–1671.

Carslake, C., Vázquez-Diosdado, J.A., Kaler, J., 2020. Machine learning algorithms to classify and quantify multiple behaviours in dairy calves using a sensor: Moving beyond classification in precision livestock. Sensors 21 (88).

Chakravarty, P., Cozzi, G., Ozgul, A., Aminian, K., 2019. A novel biomechanical approach for animal behaviour recognition using accelerometers. Methods Ecol. Evol. 10, 802–814.

Chang, A.Z., Fogarty, E.S., Moraes, L.E., García-Guerra, A., Swain, D.L., Trotter, M.G., 2022. Detection of rumination in cattle using an accelerometer ear-tag: A comparison of analytical methods and individual animal and generic models. Comput. Electron. Agric. 192, 106595.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. pp. 785–794.

Dempster, A., Petitjean, F., Webb, G.I., 2020. Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. Data Min. Knowl. Discov. 34, 1454–1495.

Dempster, A., Schmidt, D.F., Webb, G.I., 2021. Minirocket: A very fast (almost) deterministic transform for time series classification. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 248–257.

Dissanayake, O.I., McPherson, S.E., Allyndrée, J., Kennedy, E., Cunningham, P., Riaboff, L., 2025. Actbecalf: Accelerometer-based multivariate time-series dataset for calf behavior classification. Data Brief 60, 111462.

Dissanayake, O., McPherson, S., Kennedy, E., Sugrue, K., Conneely, M., Shalloo, L., Cunningham, P., Riaboff, L., 2022. Identification of the best accelerometer features and time-scale to detect disturbances in calves. In: International Workshop on Advanced Analytics and Learning on Temporal Data. Springer, pp. 167–180.

Drover, D., Howcroft, J., Kofman, J., Lemaire, E.D., 2017. Faller classification in older adults using wearable sensors based on turn and straight-walking accelerometer-based features. Sensors 17 (1321).

Duthie, C.A., Bowen, J., Bell, D., Miller, G., Mason, C., Haskell, M., 2021. Feeding behaviour and activity as early indicators of disease in pre-weaned dairy calves. Animal 15, 100150.

Dutta, R., Smith, D., Rawnsley, R., Bishop-Hurley, G., Hills, J., Timms, G., Henry, D., 2015. Dynamic cattle behavioural classification using supervised ensemble classifiers. Comput. Electron. Agric. 111, 18–28.

Enríquez, D., Ungerfeld, R., Quintans, G., Guidoni, A., Hötzel, M., 2010. The effects of alternative weaning methods on behaviour in beef calves. Livest. Sci. 128, 20–27.

Fan, S., Jia, Y., Jia, C., 2019. A feature selection and classification method for activity recognition based on an inertial sensing unit. Information 10 (290).

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.

Figo, D., Diniz, P.C., Ferreira, D.R., Cardoso, J.M., 2010. Preprocessing techniques for context recognition from accelerometer data. Pers. Ubiquitous Comput. 14, 645–662.

Fogarty, E.S., Swain, D.L., Cronin, G.M., Moraes, L.E., Trotter, M., 2020. Behaviour classification of extensively grazed sheep using machine learning. Comput. Electron. Agric. 169, 105175.

Friard, O., Gamba, M., 2016. Boris: a free, versatile open-source event-logging software for video/audio coding and live observations. Methods Ecol. Evol. 7, 1325–1330.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 118, 9–1232.

van Hees, V.T., Gorzelniak, L., Dean León, E.C., et al., 2013. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One 8 (e61691), http://dx.doi.org/10.1371/journal.pone.0061691, published 2013 Apr 23.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12, 55–67.

Hosseininoorbin, S., Layeghy, S., Kusy, B., Jurdak, R., Bishop-Hurley, G.J., Greenwood, P.L., Portmann, M., 2021. Deep learning-based cattle behaviour classification using joint time-frequency data representation. Comput. Electron. Agric. 187, 106241.

Hounslow, J., Brewster, L., Lear, K., Guttridge, T., Daly, R., Whitney, N., Gleiss, A., 2019. Assessing the effects of sampling frequency on behavioural classification of accelerometer data. J. Exp. Mar. Biol. Ecol. 512, 22–30.

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. Comput. Electron. Agric. 179, 105857.

Jiang, B., Tang, W., Cui, L., Deng, X., 2023. Precision livestock farming research: A global scientometric review. Animals 13 (2096).

Kleanthous, N., Hussain, A.J., Khan, W., Sneddon, J., Al-Shamma'a, A., Liatsis, P., 2022. A survey of machine learning approaches in animal behaviour. Neurocomputing 491, 442–463.

Koknaroglu, H., Akunal, T., 2013. Animal welfare: An animal science approach. Meat Sci. 95, 821–827.

Li, C., Tokgoz, K.K., Fukawa, M., Bartels, J., Ohashi, T., Takeda, K.i., Ito, H., 2021. Data augmentation for inertial sensor data in cnns for cattle behavior classification. IEEE Sens. Lett. 5, 1–4.

Löning, M., Bagnall, A., Ganesh, S., Kazakov, V., Lines, J., Király, F.J., 2019. Sktime: A unified interface for machine learning with time series. arXiv preprint arXiv:1909.07872.

Lubba, C.H., Sethi, S.S., Knaute, P., Schultz, S.R., Fulcher, B.D., Jones, N.S., 2019. Catch22: Canonical time-series characteristics: Selected through highly comparative time-series analysis. Data Min. Knowl. Discov. 33, 1821–1852.

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. Comput. Electron. Agric. 150, 170–177.

Mahendran, S.A., Wathes, D.C., Blackie, N., Booth, R.E., 2023. Calf health, feeding and social behaviours within groups fed on automatic milk feeders. Ruminants 3, 266–285.

Manning, J.K., Cronin, G.M., González, L.A., Hall, E.J., Merchant, A., Ingram, L.J., 2017. The effects of global navigation satellite system (gnss) collars on cattle (bos taurus) behaviour. Appl. Anim. Behav. Sci. 187, 54–59.

Martiskainen, P., Järvinen, M., Skön, J.P., Tiirikainen, J., Kolehmainen, M., Mononen, J., 2009. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. Appl. Anim. Behav. Sci. 119, 32–38.

McPherson, S., Riaboff, L., Dissanayake, O., Sinnott, A., Cunningham, P., Kennedy, E., 2022. Effect of separation at weaning on the activity of cows and calves reared in a cow-calf contact system measured with accelerometer sensors. In: 10th European Conference on Precision Livestock Farming, Vienna.

Mietus, J., Peng, C., Henry, I., Goldsmith, R., Goldberger, A., 2002. The pnnx files: re-examining a widely used heart rate variability measure. Heart 88 (378).

Moreau, M., Siebert, S., Buerkert, A., Schlecht, E., 2009. Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behaviour. Appl. Anim. Behav. Sci. 119, 158–170.

Morisse, J., Cotte, J., Huonnic, D., 1995. Effect of dehorning on behaviour and plasma cortisol responses in young calves. Appl. Anim. Behav. Sci. 43, 239–247.

Nikkhah, A., Alimirzaei, M., 2023. Understanding calf behavioral responses to environmental changes and challenges: An applied update. Farm Anim. Heal. Nutr. 2, 71–77.

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, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Penning, P., 1983. A technique to record automatically some aspects of grazing and ruminating behaviour in sheep. Grass Forage Sci. 38, 89–96.

Preece, S.J., Goulermas, J.Y., Kenney, L.P., Howard, D., 2008. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE Trans. Biomed. Eng. 56, 871–879.

Price, E., Langford, J., Fawcett, T.W., Wilson, A.J., Croft, D.P., 2022. Classifying the posture and activity of ewes and lambs using accelerometers and machine learning on a commercial flock. Appl. Anim. Behav. Sci. 251, 105630.

Rahman, A., Smith, D., Little, B., Ingham, A., Greenwood, P., Bishop-Hurley, G., 2018. Cattle behaviour classification from collar, halter, and ear tag sensors. Inf. Process. Agric. 5, 124–133.

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. Comput. Electron. Agric. 165, 104961.

Riaboff, L., Poggi, S., Madouasse, A., Couvreur, S., Aubin, S., Bédère, N., Goumand, E., Chauvin, A., Plantier, G., 2020. Development of a methodological framework for a robust prediction of the main behaviours of dairy cows using a combination of machine learning algorithms on accelerometer data. Comput. Electron. Agric. 169, 105179.

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. Comput. Electron. Agric. 192, 106610.

Rodriguez-Baena, D.S., Gomez-Vela, F.A., García-Torres, M., Divina, F., Barranco, C.D., Daz-Diaz, N., Jimenez, M., Montalvo, G., 2020. Identifying livestock behavior patterns based on accelerometer dataset. J. Comput. Sci. 41, 101076.

Rushen, J., Chapinal, N., de Passile, A., 2012. Automated monitoring of behavioural-based animal welfare indicators. Anim. Welf. 21, 339–350.

Sarker, I.H., 2021. Machine learning: Algorithms, real-world applications and research directions. SN Comput. Sci. 2, 160.

da Silva, I.C., van Hees, V.T., Ramires, V.V., Knuth, A.G., Bielemann, R.M., Ekelund, U., Brage, S., Hallal, P.C., 2014. Physical activity levels in three brazilian birth cohorts as assessed with raw triaxial wrist accelerometry. Int. J. Epidemiol. 43, 1959–1968.

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. Comput. Electron. Agric. 131, 40–50.

Tatler, J., Cassey, P., Prowse, T.A., 2018. High accuracy at low frequency: detailed behavioural classification from accelerometer data. J. Exp. Biol. 221, jeb184085.

Valletta, J.J., Torney, C., Kings, M., Thornton, A., Madden, J., 2017. Applications of machine learning in animal behaviour studies. Anim. Behav. 124, 203–220.

Vayssade, J.A., Godard, X., Bonneau, M., 2023. Wizard: Unsupervised goats tracking algorithm. Comput. Electron. Agric. 209, 107831.

Vázquez Diosdado, J.A., Barker, Z.E., Hodges, H.R., Amory, J.R., Croft, D.P., Bell, N.J., Codling, E.A., 2015. Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. Anim. Biotelemetry 3, 1–14.

Versluijs, E., Niccolai, L.J., Spedener, M., Zimmermann, B., Hessle, A., Tofastrud, M., Devineau, O., Evans, A.L., 2023. Classification of behaviors of free-ranging cattle using accelerometry signatures collected by virtual fence collars. Front. Anim. Sci. 4, 1083272.

Walker, J.S., Jones, M.W., Laramee, R.S., Holton, M.D., Shepard, E.L., Williams, H.J., Scantlebury, D.M., Marks, N.J., Magowan, E.A., Maguire, I.E., et al., 2015. Prying into the intimate secrets of animal lives; software beyond hardware for comprehensive annotation in 'daily diary'tags. Mov. Ecol. 3, 1–16.

Wang, X., Wirth, A., Wang, L., 2007. Structure-based statistical features and multivariate time series clustering. In: Seventh IEEE International Conference on Data Mining. ICDM 2007, IEEE, pp. 351–360.

Wilson, R.P., Shepard, E., Liebsch, N., 2008. Prying into the intimate details of animal lives: use of a daily diary on animals. Endanger. Species Res. 4, 123–137.

Yu, H., Muijres, F.T., te Lindert, J.S., Hedenström, A., Henningsson, P., 2023. Accelerometer sampling requirements for animal behaviour classification and estimation of energy expenditure. Anim. Biotelemetry 11 (28).

Zdravevski, E., Risteska Stojkoska, B., Standl, M., Schulz, H., 2017. Automatic machine-learning based identification of jogging periods from accelerometer measurements of adolescents under field conditions. PloS One 12, e0184216.