Computers and Electronics in Agriculture 192 (2022) 106610

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Predicting livestock behaviour using accelerometers: A systematic review of processing techniques for ruminant behaviour prediction from raw accelerometer data

L. Riaboff[a,c,*], L. Shalloo[c,d], A.F. Smeaton[e], S. Couvreur[f], A. Madouasse[g], M.T. Keane[a,b,c]

[a] *School of Computer Science, University College Dublin, Dublin, Ireland*
[b] *Insight SFI Centre for Data Analytics, Dublin City University, Dublin, Ireland*
[c] *VistaMilk SFI Research Centre, Ireland*
[d] *Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork P61C997, Ireland*
[e] *Insight SFI Centre for Data Analytics, Dublin City University, Glasnevin, Dublin 9, Ireland*
[f] *USC ESA-INRAE 1481 URSE, Ecole Supérieure d'Agricultures, University Bretagne Loire, 49007 Angers, France*
[g] *INRAE, Oniris, BIOEPAR, 44300 Nantes, France*

ARTICLE INFO

Keywords:
Behaviour
Ruminant
Prediction
Accelerometer sensor
Raw data
Methodology
Processing techniques

ABSTRACT

Precision Technologies are emerging in the context of livestock farming to improve management practices and the health and welfare of livestock through monitoring individual animal behaviour. Continuously collecting information about livestock behaviour is a promising way to address several of these target areas. Wearable accelerometer sensors are currently the most promising system to capture livestock behaviour. Accelerometer data should be analysed properly to obtain reliable information on livestock behaviour. Many studies are emerging on this subject, but none to date has highlighted which techniques to recommend or avoid. In this paper, we systematically review the literature on the prediction of livestock behaviour from raw accelerometer data, with a specific focus on livestock ruminants. Our review is based on 66 surveyed articles, providing reliable evidence of a 3-step methodology common to all studies, namely (1) *Data Collection*, (2) *Data Pre-Processing* and (3) *Model Development*, with different techniques used at each of the 3 steps. The aim of this review is thus to (i) summarise the predictive performance of models and point out the main limitations of the 3-step methodology, (ii) make recommendations on a methodological blueprint for future studies and (iii) propose lines to explore in order to address the limitations outlined. This review shows that the 3-step methodology ensures that several major ruminant behaviours can be reliably predicted, such as grazing/eating, ruminating, moving, lying or standing. However, the areas faces two main limitations: (i) Most models are less accurate on rarely observed or transitional behaviours, behaviours may be important for assessing health, welfare and environmental issues and (ii) many models exhibit poor generalisation, that can compromise their commercial use. To overcome these limitations we recommend maximising variability in the data collected, selecting pre-processing methods that are appropriate to target behaviours being studied, and using classifiers that avoid over-fitting to improve generalisability. This review presents the current situation involving the use of sensors as valuable tools in the field of behaviour recording and contributes to the improvement of existing tools for automatically monitoring ruminant behaviour in order to address some of the issues faced by livestock farming.

Abbreviations: Accx, X-axis acceleration; Accy, Y-axis acceleration; Accz, Z-axis acceleration; Ada, Adaptative Boosting; ANFIS, Adaptive Neuro Fuzzy Inference System; CNN, Convolutional Neural Network; DL, Deep Learning; DT, Decision tree; FD, Frequency Domain; FDR, False Discovery Rate; GPS, Global Positioning System; HMM, Hidden Markov Model; kNN, k-Nearest Neighbours; LDA, Linear Discriminant Analysis; LOOV, Leave-One-Out Validation; LR, Logistic Regression; LSTM, Long Short-Time Memory; MI, Motion intensity; MLP, Multilayer Perceptron; MT, Manual Thresholding; NB, Naive Bayes; OBDA, Overall Body Dynamic Acceleration; PCA, Principal Component Analysis; PNN, Probabilistic Neural Network; QDA, Quadratic Discriminant Analysis; RF, Random Forest; RS, Random Subspace; RFID, Radio-Frequency Identification; RNN, Recurrent Neural Network; RSS, Received Signal Strength; SEML, Supervised Ensemble Machine Learning; SM, Statistical Models; SML, Supervised Machine Learning; SVM, Support Vector Machine; TD, Time Domain; TFD, Time Frequency Domain; UML, Unsupervised Machine Learning; USB, Universal Serial Bus; VeDBA, Vector Dynamic Body Acceleration; WS, Window Size; XGB, eXtreme Gradient Boosting.

* Corresponding author at: UCD, School of Computer Science, Belfield, Dublin 4, Ireland.
E-mail address: lucile.riaboff@ucd.ie (L. Riaboff).

https://doi.org/10.1016/j.compag.2021.106610
Received 4 August 2021; Received in revised form 12 November 2021; Accepted 30 November 2021
Available online 17 December 2021

1. Introduction

Technologies are emerging in the context of precision livestock farming to improve the efficiency of livestock management (Rutten et al., 2016; Shalloo et al., 2021) and the monitoring of health and welfare (Medria Solutions, 2020). In the context of these developments, the collection of continuous information about the behaviour of livestock offers considerable promise. For example, the continuous monitoring of grazing behaviour may ensure a better understanding of feeding strategies in order to adapt management practices for greater efficiency (Carvalho, 2013). Continuous information about resting and lying behaviours can also help to detect stressful situations, as a change is observed when conditions are sub-optimal (Heinicke et al., 2019; Silberberg et al., 2017). Monitoring of feeding, drinking or lying behaviour can also help to detect reproductive events such as calving (Jensen, 2012), thus contributing to better reproductive performance on the farm.

Several sensor types have been used in the literature to monitor animal behaviour (Delagarde et al., 1999; Rutter et al., 1997; Ruuska et al., 2016) but wearable 3-Dimensional accelerometer sensors seem currently the most promising of these sensing systems (Bailey et al., 2018). In this regard, commercial systems based on 3-Dimensional accelerometers are now available to automatically quantify livestock behaviour (Borchers et al., 2016; Hendriks et al., 2020; Shalloo et al., 2021), especially in cattle, such as the CowManager (Agis, Harmelen, the Netherlands), the HOBO Data Logger (HOBO Pendant G Acceleration Data Logger, Onset Computer Corporation, Pocasset, MA), the MooMonitor + collar (Dairymaster, Tralee, Ireland), the AfiAct Pedometer Plus (Afimilk, S.A.E. Afikim, Kibbutz Afikim, Israel) or the IceTag (IceRobotics Ltd., Edinburgh, Scotland). However, as shown in Borchers et al. (2016), these systems focus either on feeding behaviours, such as grazing or ruminating (e.g., CowManager), or on the lying position (e.g., HOBO Data Logger). None of these systems covers a broad spectrum of behaviours (e.g., predicting ruminating, grazing, running, walking, grooming, drinking, lying down and standing up), unless several sensors are combined on different positions on the animal (RumiWatchSystem, Itin + Hoch GmbH, Liestal, Switzerland). This constraint, together with the high cost and difficult data management, makes the RumiWatchSystem not compatible in its current form with use in commercial farms. In addition, none of these systems provides information about behaviours that are less frequently expressed in animals, such as grooming, scratching, urinating or drinking, but that are still relevant to address certain issues (e.g., health; Galindo and Broom, 2002 or environmental; Lush et al., 2018). Finally, the validation of these systems to assess their reliability is not always reported (Borchers et al., 2016). Therefore, the limitations of current commercial systems justifies further development of new methods to predict livestock behaviours from wearable 3-Dimensional accelerometers (Pavlovic et al., 2021).

Using 3-Dimensional accelerometer sensors requires appropriate analytical methods for robust and reliable identification of behaviours from raw data (e.g., Barwick et al., 2018). Many different processing techniques have been used in the literature to predict livestock behaviour from accelerometer data. Although some studies have already compared several methods (Barwick et al., 2018; Dutta et al., 2015; Hu et al., 2020; Smith et al., 2016), there are currently no clear recommendations on those to be adopted or avoided, either for the data collection or analysis purposes. To begin to tackle this challenge, this paper reviews different processing techniques used to predict livestock behaviour continuously from raw accelerometer data, to identify the most effective ones and to highlight their limitations. To the best of the authors' knowledge, no such review of this topic has yet been published in the literature.

In this review, we focus specifically on the state-of-the-art in processing techniques used to predict ruminant behaviour from raw accelerometer data. Furthermore, our aim in this research is not just to record what has been done, but also to identify the most promising methods and techniques from a methodological standpoint in this field. In the remainder of this introduction, we outline (i) the keywords and searches used in our systematic survey (Section 1.1) and (ii) the structure of the review (Section 1.2).

Table 1
Three systematic searches using keyword sets per ruminant livestock species and the numbers of non-unique citations found.

Search #	Keywords	# Citations
#1 (cow)	accelerometer, behaviour, classification, prediction, monitoring, processing, calf, cow, heifer, beef, bull	307
#2 (sheep)	accelerometer, behaviour, classification, prediction, monitoring, processing, sheep, ewes, lamb	311
#3 (goat)	accelerometer, behaviour, classification, prediction, monitoring, processing, goat	969

1.1. Methodology used to select the papers in the systematic review

Our systematic review of the literature on processing techniques for predicting ruminant behaviour arises from several literature searches carried out using Google Scholar (https://scholar.google.com/) during a period spanning 24th December 2020 to 3rd January 2021. The overall trawl was split into three searches related to the three main ruminant species, anchored by the keywords "cow" (307 citations), "sheep" (301 citations) and "goat" (969 citations) (Table 1).

As displayed in PRISMA flow diagram (Fig. 1), for each search, the initial set of papers were further filtered based on a close reading of the title and abstract using the following three criteria:

- Criterion 1: The citation is an *original paper* resulting from a peer-review publication process.
- Criterion 2: The target is to *predict state behaviours* of ruminants *per se*, as defined by Kilgour (2012) on pasture or by Zambelis (2019) in the barn, including maintenance (e.g., feeding, ruminating, drinking), self-expression (e.g., scratching) or social behaviours (e.g., grooming). In this respect, it is necessary to note that citations relating to the prediction of reproductive behaviours, such as *estrus* or calving, were excluded from the selection process because the methodology used to predict them is substantially different from that used to predict the basic behaviours of interest (Benaissa et al., 2020). In the same way, behaviours related to health issues, like lameness, were considered outside the scope of the review.
- Criterion 3: The accelerometer data used to predict behaviours are *raw data*, as this is the most promising way to improve current findings in the field.

After applying these criteria the initial trawl of 1577 articles was reduced to 66 key citations. These 66 papers were subsequently read in full and their methodologies for predicting ruminant behaviour were noted, analysed and collated for the review.

1.2. Structure of the paper

In the remainder of this paper, we present a systematic review of the literature on processing techniques for predicting ruminant behaviour from raw accelerometer data. In the processing of raw accelerometer data to predict animal behaviour, most researchers use a methodology involving the following three stages, as illustrated in Fig. 2:

1) *Data Collection*: Collection of raw accelerometer data from wearable sensors and behavioural observations in the field.
2) *Data Pre-processing*: Pre-processing of accelerometer data to get a suitable dataset.

3) *Development of a Behavioural Classification Model:* Development of classification model to predict ruminant behaviours.

This 3-step methodology is used to structure Sections 2–4. Section 2 reviews the different approaches for the *Data Collection*, Section 3 recounts methods used for the *Data Pre-processing* and Section 4 explores the dominant techniques used for the *Development of a Behavioural Classification Model.* Section 5 reviews the predictive performance found across the 66 selected studies. Sections 2-5 are designed using the same scheme: techniques used in the 66 papers are described in the core text and summarized in figures. Details of 30 articles of the 66 read in full are provided in appendices. These 30 core articles were selected to avoid large tables based on their high and complete methodological practices; that is, there is no main missing elements in the description of the material and methods used, the results are reported in a complete and understandable way and at least two distinct behaviours are predicted in order to display the performance obtained in a challenging framework. The details of the remaining 36 articles are reported in Supplementary Materials. Section 6 presents a set of recommendations that are more likely to lead to successful results for future studies. It is our hope that these recommendations will put the field on a better footing and contribute to the improvement of existing tools for automatically monitoring ruminant behaviour. A conclusion is finally proposed in Section 7.

2. Data collection: Animal, sensor and observational options

The purpose of this first step is to collect the data needed to develop a classification model in the final stage of the methodology (Fig. 2), that is:

- Raw accelerometer data, collected at regular time-intervals (also called *sampling rate*; units: Hertz) using one or more 2-Dimensional or 3-Dimensional accelerometer sensors attached to animals, positioned on varied body-parts.

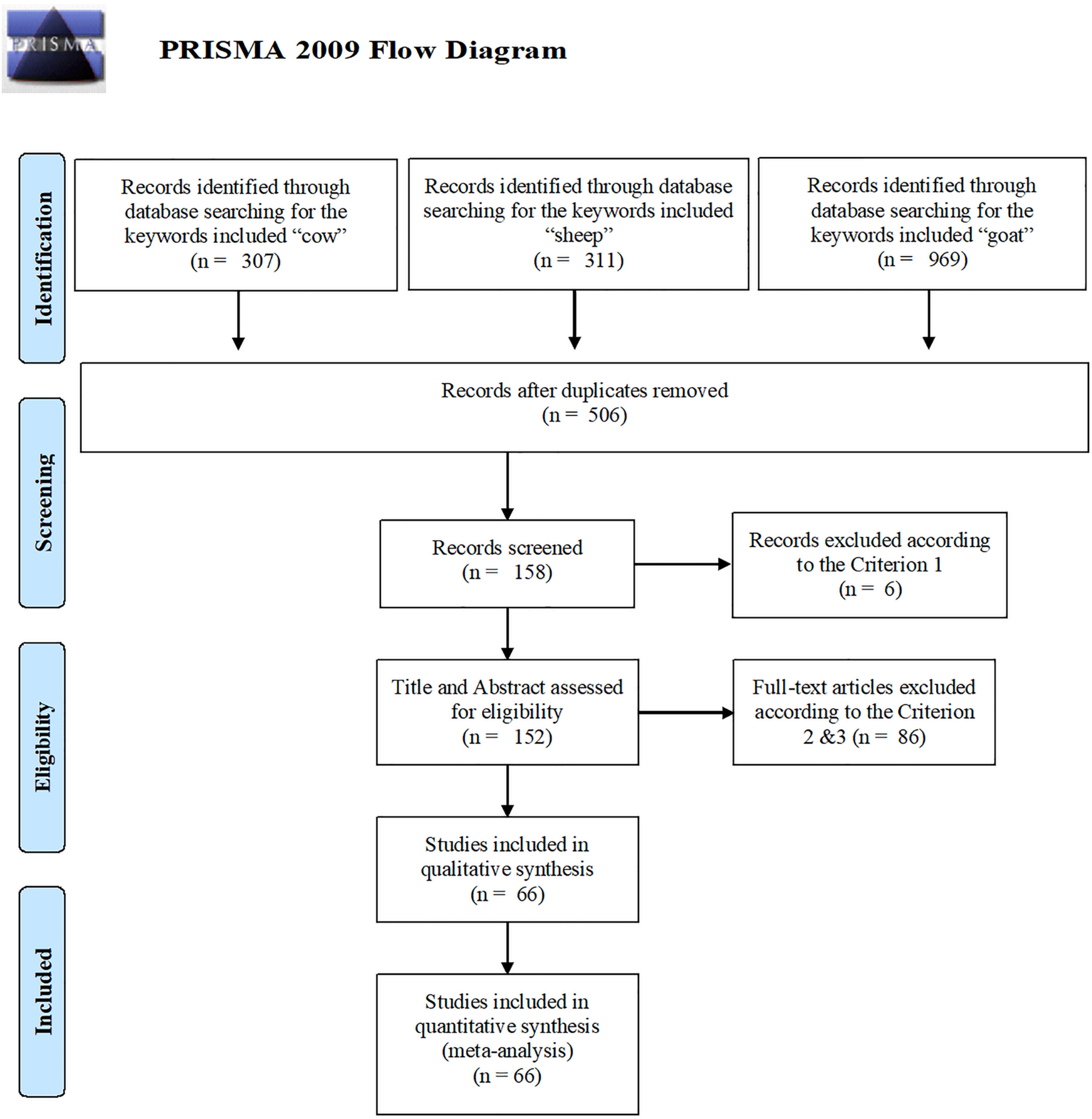
PRISMA
PRISMA 2009 Flow Diagram
Identification
Screening
Eligibility
Included
Records identified through database searching for the keywords included "cow" (n = 307)
Records identified through database searching for the keywords included "sheep" (n = 311)
Records identified through database searching for the keywords included "goat" (n = 969)
Records after duplicates removed (n = 506)
Records screened (n = 158)
Records excluded according to the Criterion 1 (n = 6)
Title and Abstract assessed for eligibility (n = 152)
Full-text articles excluded according to the Criterion 2 &3 (n = 86)
Studies included in qualitative synthesis (n = 66)
Studies included in quantitative synthesis (meta-analysis) (n = 66)

Fig. 1. PRISMA flow diagram for the systematic review of the literature on the prediction of ruminant behaviours using raw accelerometer data. The keywords associated with each species are provided in Table 1. The criteria mentioned in the diagram are those introduced above.

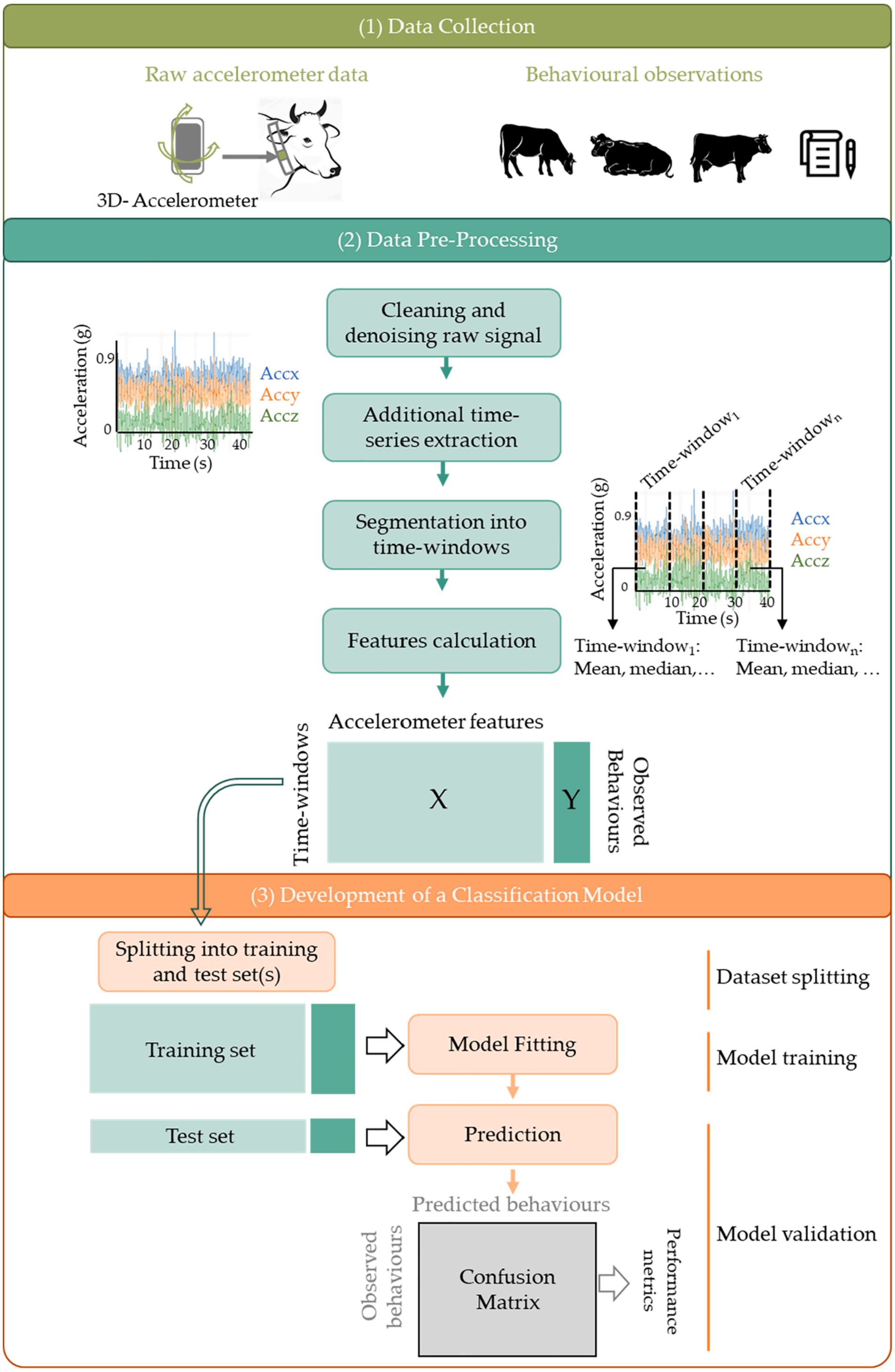

Fig. 2. Overview of the 3-steps methodology used in the literature to predict ruminant behaviour from raw accelerometer data with (1) the *Data Collection* step, (2) the *Data Pre-Processing* step and (3) the *Development of a Classification Model* step which involves the sub-steps of Dataset splitting, Model training and Model Validation.

- Observations of the equipped animals expressing different behaviours. The behaviours observed and the associated time are recorded.

The approach adopted in the studies varies according to three main sets of choices about (i) the animals studied, (ii) what sensors were used and how they were deployed and (iii) how the equipped animals are observed. Fig. 3 provides the summary statistics for the options adopted in this literature. Appendix 1 presents detailed profiles of the 30 core papers (see Supplementary Materials A1 for the details of the remaining 36 papers). In the following sub-sections, we elaborate on the specifics of the three sets-of-choices made.

2.1. *Animals*

2.1.1. *Species and breeds*

All the main species of livestock ruminants (cattle, sheep and goat) have been studied over the 66 papers reported, though almost two thirds were on cattle and one third on sheep (Fig. 3a). Of the studies using

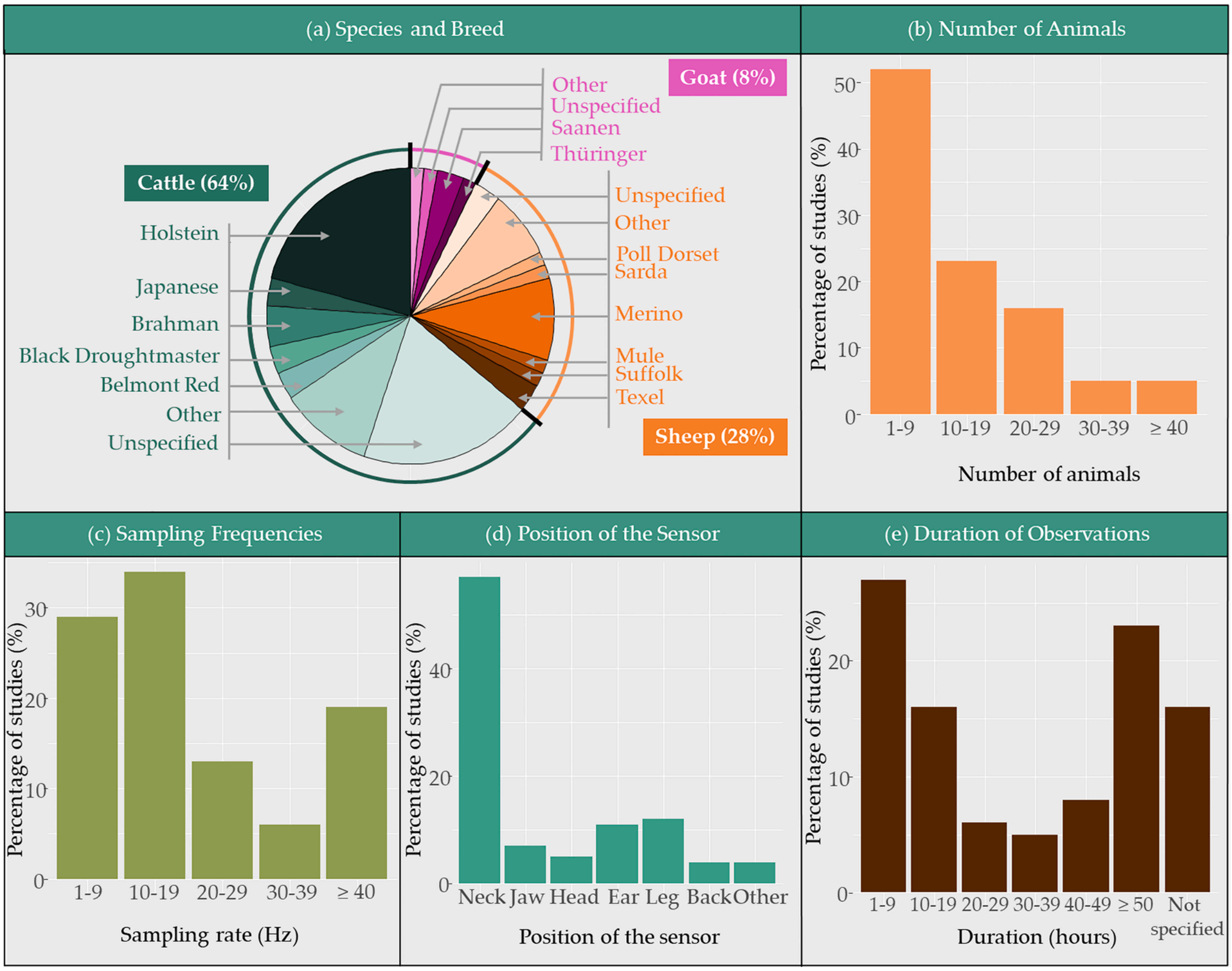

Fig. 3. In *Data Collection*: Distribution of the different techniques used in the surveyed papers (N = 66) for (a) species and breeds, (b) number of animals, (c) sampling frequencies, (d) positioning of sensors and (e) duration of animal observation. Pie chart (a) shows the percentage of studies using cattle, sheep and goats with the distribution of breeds for each species. The bar charts (b)-(e) show the percentage of studies using the different techniques that we grouped into classes/categories.

cattle, 33% were done with Holstein and 20% were done with Japanese Black, Droughtmaster, Belmont Red, Brahman or crossings of these species. Cattle breed was unspecified in 31% of the cattle studies. In sheep, 30% of the studies were done on Merino or Merino crosses, 39% on Suffolk, Mule, Sarda, Poll Dorset or crossings of these species and 11% on Texel or crosses Texel. Sheep breed was unspecified in 7% of the studies. Few studies focused on goat but the Saanen and Saanen crosses seem the most used in the goat studies (40%; Fig. 3a).

2.1.2. *Number of equipped animals*

The median number of animals used in the studies is 8 animals but the number of equipped animals varied considerably from one study to another, ranging from a minimum of one animal used per experiment (Watanabe et al., 2005) to a maximum of 86 animals (Riaboff et al., 2020) but studies using more than 30 animals represent less than 10% of all studies (Fig. 3b). Suggestions on the optimal number of animals for studies are proposed in Section 6.

2.2. *Sensors and devices*

2.2.1. *Accelerometer sensor*

Of the 66 studies using a wearable accelerometer to predict behaviours, 97% used 3D-accelerometer sensors while the remaining used 2D-accelerometers (e.g. Nadimi et al., 2008). Accelerometer data were collected either using a study-specific device including a datalogger (e.g. Axivity AX3; Fogarty et al., 2020), a smartphone (e.g. iPhone 4S; Andriamandroso et al., 2017) or a commercial system for ruminants in which raw data is available (e.g. HOBO® loggers; Benaissa et al., 2018). It should be noted that additional sensors were sometimes attached to animals and used for behaviour prediction in addition to accelerometer data. The added sensors were most often a 3D-gyroscope (Walton et al., 2018), a 3D-magnetometer (Dutta et al., 2015) or a GPS (González et al., 2015). The value of combining the accelerometers with other sensors is discussed in Section 6.

2.2.2. *Sampling frequencies*

The median frequency used to sample raw accelerometer data is 12 Hz. The sampling rate varied considerably from one study to another, ranging from a minimum of 0.02 Hz (Zobel et al., 2015) to a maximum of 200 Hz (Kleanthous et al., 2018). The signal was sampled at less than 20 Hz in 70% of the studies and 17% of the studies used sampling frequencies higher than 40 Hz (Fig. 3c). It should be noted that several sampling frequencies have been used in some studies to explore their impact on behaviour prediction, either configuring different sampling frequencies during experimentation (Walton et al., 2018) or down sampling accelerometer signal at different frequencies from the original sampling rate (Benaissa et al., 2018). A recommendation on the sampling rate to be used is proposed in Section 6.

2.2.3. *Sensor positioning on animals*

Accelerometer devices have been attached to different positions on ruminants' bodies (Fig. 3d), mostly around the neck using a collar (57%;

e.g. Smith et al., 2016), on the leg using a tape (12%; e.g. Robert et al., 2009) or on the ear using a tag (11%; e.g. Fogarty et al., 2020). The other main positions are under the jaw (7%; e.g. Watanabe et al., 2008) or the side of the head using a halter (5%; Kour et al., 2018) or on the back using a harness or adhesive (4%; Lush et al., 2018). Remaining positions represented only 4% of all the studies and involved the withers using a belt or a patch (e.g. Abell et al., 2017) or reticulum using a motion-sensitive bolus-sensor (Hamilton et al., 2019). See Section 6 for a discussion on where best to position sensors based to the target behaviours being studied.

2.3. Observations of animals

2.3.1. Behaviours encoded during the observation period

A wide range of ruminant state behaviours encoded during the observation period have been reported across the studies, including ruminating, resting, lying, standing, walking, running, chewing and feeding or grazing depending on whether observations were carried out in a barn or in a pasture, respectively (Martiskainen et al., 2009; Riaboff et al., 2020). Transitional behaviours, such as standing up or lying down, were recorded in some studies (Vázquez Diosdado et al., 2015). Social and welfare behaviours have already been encoded such as interaction between cows (Riaboff et al., 2020), grooming (Smith et al., 2015), social licking (Peng et al., 2019) or scratching (Lush et al., 2018). Behaviours contributing to the maintenance of integrity have also been recorded, such as drinking, urinating or defecating (Smith et al., 2015). It should be noted that the spectrum of behaviours was often broader in sheep and goat, as studies also include other specific behaviours, such as climbing, rubbing, perching or trotting (Kamminga et al., 2018; Radeski and Ilieski, 2017; Zobel et al., 2015). In general, the behaviours encoded have been defined in an ethogram, with slight differences in the definition used from one study to another; for example, grazing behaviour may include chewing head up (Barwick et al., 2018) or chewing head down (Riaboff et al., 2020).

2.3.2. Encoding techniques to observe animals

In most studies, equipped animals were observed successively by one or more experimenters during a continuous period using either direct observations (Dutta et al., 2015) or video recordings (Barwick et al., 2018). During these periods, both the observed behaviours and the corresponding times were recorded. Scan-sampling has been used occasionally so that each animal has been monitored regularly for a given period of time and the behaviour encoded is that expressed by the animal each time it has been observed (Tamura et al., 2019). Encoding of behaviours was performed using either manual annotations (Decandia et al., 2018), custom-designed applications (Hu et al., 2020) or software designed for behaviour encoding (e.g. Noldus Observer XT 11, Noldus; Mansbridge et al., 2018). The value of using some of these approaches over others is discussed in Section 6.

2.3.3. Duration of observation

Duration of observation was unspecified in 16% of the studies and is sometimes reported as an approximation (e.g. Walton et al., 2018) but we can still estimate that the average total duration of observation is 47 h ± 69 (mean ± standard deviation). Duration of observations also varied considerably from one study to another, ranging from <2 h (e.g. Radeski and Ilieski, 2017) to >200 h (e.g. Wang et al., 2018). A more detailed analysis also highlights two distributions for the 66 studies (Fig. 3e), with a duration of observation <20 h in 42% of cases but also >50 h in 25% of studies. A recommendation on the minimum observation time of animals to ensure reliable model development and validation is proposed in Section 6.

3. Pre-processing: Data cleaning, time series calculation, segmentation & feature extraction

The raw accelerometer signal measured on three axes – the x-axis, y-axis, z-axis – is usually represented as a *time-series*; that is, acceleration measurements associated with the corresponding recording time. The main target of this step is to compute features from the time-series that best represent the information provided in the accelerometer signal, to best predict the targeted behaviours being studied. The pre-processing step is usually carried out in 4 sub-steps (Fig. 2) that include: *cleaning and removing noise* in the raw time-series, *calculating additional time-series*, *segmenting* the time-series into *time-windows* and, finally, *calculating features* from each time-window. The main differences identified in the 66 papers are related to (i) how the initial raw data is cleaned and noise is removed, (ii) how the filtered time-series is segmented and (iii) what different types of features are calculated. Fig. 4 provides the summary statistics of the options adopted in the literature. Appendix 2 presents detailed profiles of the 30 core papers (see Supplementary Materials A2 for the details of the remaining 36 papers). In the following sub-sections, we elaborate on the specifics of the three sets-of-choices made.

3.1. Cleaning and removing noise in the raw time-series

Poor data transmission, or a problem during data acquisition, may result in missing samples, especially when a wireless data-acquisition system was used (e.g. up to 10% of the acceleration measurements may not be received; Martiskainen et al., 2009). In the 66 surveyed papers, 15% of the studies reported removing missing data (Fig. 4a) due to sensor malfunction or other data retrieving issues (e.g. Walton et al., 2018). Typically, when the timestamps of the records were not associated with their corresponding accelerometer values, these records were most often removed (Peng et al., 2020; Wang et al., 2019) before continuing with signal processing. More rarely, missing records have been replaced using linear interpolation (Kleanthous et al., 2018).

In addition to the missing data, raw accelerometer time-series may be noisy or have outliers (e.g. sensor impact, measurement error; see Supplementary Data B). For that reason, removing noise from the raw time-series has been reported in 15% of the studies (Fig. 4a). Outliers have often been removed arbitrarily using a threshold based on quantiles (95th and 5th quantile; Williams et al., 2020) or standard deviations (mean ± 2 standard deviations; Giovanetti et al., 2017). In some studies, artefacts were corrected using unwrap correction (e.g. jumps in orientation coordinates due to spherical coordinate representation; Smith et al., 2015). Moving average filtering (Hamalainen et al., 2011) or low-pass filtering (cut-off frequency: 10 Hz; le Roux et al., 2017) have also often been used to filter the noisy time-series.

Recommendations regarding the cleaning and procedure for removing noise in the raw time-series are suggested in Section 6.

3.2. Calculating additional time-series

Additional time-series are calculated (Fig. 2) from the cleaned or raw time-series in 68% of the studies (Fig. 4b). Additional time-series can be calculated to be used on their own (23%) or in combination with additional time-series (44%). Based on the 66 papers, there are three main reasons to calculate additional time-series.

3.2.1. Time-series independent of the orientation of the sensor

As illustrated in Supplementary Data B1, the influence of the gravity force is more or less important depending on the direction of the axis. Therefore, a minor rotation of the sensor around the neck or the leg may result in different acceleration measurements for the same expressed behaviour by the same animal and from the same sensor (Hamalainen et al., 2011). An additional time-series independent of the sensor orientation is thus usually calculated (Fida et al., 2015), such as the magnitude of the acceleration, which can be called the *Signal Vector*

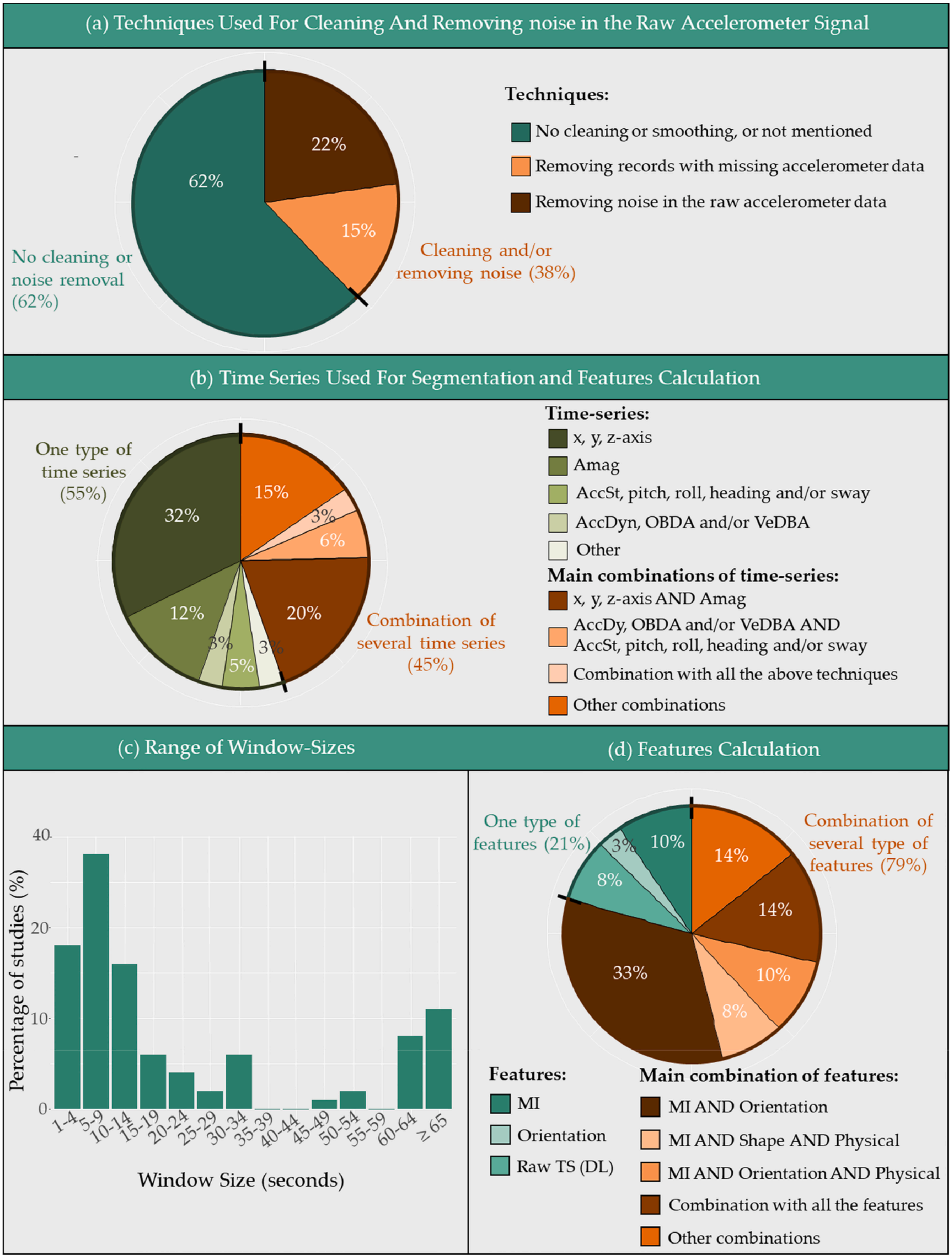

Fig. 4. In *Pre-Processing*: Distribution of the techniques used in the surveyed papers (N = 66), including (a) the cleaning and filtering techniques applied on the raw accelerometer signal, (b) the time-series extracted from the pre-processed signal to segment it and then calculate features, (c) the window-size chosen for the segmentation and (d) the type of features calculated in each time-window. Note: Abbreviations used in the Figure: AccSt: Static Acceleration; AccDy: Dynamic Acceleration; Amag: Magnitude of the acceleration; OBDA: Overall Body Dynamic Acceration; VeDBA: Vector Dynamic Body Acceleration; TS: Time-Series; MI: Motion Intensity; DL: Deep Learning.

Magnitude (Robert et al., 2009), *Asum* (Benaissa et al., 2017), *resultant* (Watanabe et al., 2008) or *Amag* (Riaboff et al., 2019) time-series, depending on the paper, and calculated as follows for a 3D-accelerometer sensor:

$$\text{Amag} = \sqrt{\text{x\_axis}^2 + \text{y\_axis}^2 + \text{z\_axis}^2} \tag{1}$$

As illustrated in Fig. 4b, Amag is used as a single time-series in the further steps in 12% of the 66 surveyed papers, and in addition to the 3 time-series from the x-, y-, and z-axes in 20% of them.

3.2.2. Time-series related to the two main components of the acceleration

Some additional time-series are calculated because they give an approximation for the energy expended during animal movement (Qasem et al., 2012), such as the Overall Body Dynamic Acceleration (OBDA) or the Vectorial Dynamic Body Acceleration (VeDBA) (Benaissa et al., 2017; Khanh et al., 2016). Both of these time-series are derived from the *dynamic acceleration* (Supplementary B1) which is usually obtained by subtracting the running mean from raw accelerometer time-series (Lush et al., 2018; Sakai et al., 2019) or high-pass filtering (cut-off frequency: 0.3 Hz; Smith et al., 2016). For formulae and a more in-depth explanation of these series, we refer to Qasem et al. (2012).

Some additional time-series are also calculated because they give an approximation for the animal's body tilt during the expressed behaviours (e.g. head up, head down, head tilted to the right side), such as the pitch, roll or sway time-series. All of these measures are derived from the *static acceleration* (Supplementary B1) which is often isolated using running means (Lush et al., 2018) or low-pass filtering (cut-off frequency: 0.3 Hz; Riaboff et al., 2020). For formulae and a more in-depth explanation of these series, we refer to Lush et al. (2018) and Walker et al. (2015).

As illustrated in Fig. 4b, the dynamic and static time-series and their derived time-series OBDA, VeDBA and pitch, roll, heading or sway, respectively, are used alone in the further steps in 8% of the 66 surveyed papers and combined together in 6% of them.

3.2.3. Time-series associated to a specific frequency-band

Some behaviours may be related to specific frequencies. If these frequencies are known, the raw time series can be filtered according to these frequencies using a bandpass filter. A peak observed in the selected band is then used to identify the behaviour and discriminate it from others (Supplementary Data B1). This technique has been found in one study (Andriamandroso et al., 2017) where the authors used a band-pass filter between 1 Hz and 2 Hz to isolate the frequencies related to the periodicity of the chewing movement during grazing and ruminating, and have used this filtered time-series to discriminate chewing from the other behaviours thereafter. Recommendations regarding the time-series that may be usefully calculated are suggested in Section 6.

3.3. Segmentation of the time-series into time-windows

The time-series obtained are then split into segments at regular intervals (*fixed segmentation*) (Fig. 2), usually called *epochs* (Robert et al., 2009), *interval* (Giovanetti et al., 2017) or *windows* (Smith et al., 2016) in the literature. In this review, we call these segments *windows* thereafter. The window is the fundamental statistical unit in subsequent analyses.

The number of data samples that are collected together in a window (Supplementary Data B2) is usually called the *time-epoch* (Robert et al., 2009), the *time-interval* (Giovanetti et al., 2017) or the *window size* (Smith et al., 2016). In this review, we call the number of data samples in the window the *window size* (WS) thereafter. Different window sizes (Supplementary Data B2) have been used to split the time-series in the surveyed papers, from 1 sec. as the shortest (Hu et al., 2020; Moreau et al., 2009) to one hour as the largest (Mattachini et al., 2016), with a median window size of 10 sec. In our survey, most studies (46%) used a WS $<$ 10 sec. but WS $>$ 1 min were also used in ~20% of the studies (Fig. 4c). Several studies investigated different window sizes to find the one that yields the best predictive performance with their classification model (Alvarenga et al., 2016; Robert et al., 2009). Considering several different windows for the next steps instead of just one window size has also been found in the literature (Hu et al., 2020).

The percentage of data in common between two consecutive windows is called the *overlap* (Supplementary Data B2). As illustrated in Fig. 4c, over the 66 surveyed papers, 80% used no *overlap* between windows; that is, there is no common data across two successive time-windows. Hence, 20% of studies have also considered overlap between successive time-windows (e.g. le Roux et al., 2017; Riaboff et al., 2020). The effect of overlap has only been investigated in one study, evaluating the performance of the prediction of the models using several percentages of overlap (from 0% to 90%; Riaboff et al., 2019). Recommendations relating to which segmentation techniques increase the predictive accuracy are proposed in the Section 6.

3.4. Features calculation into each time-window

With the exception of studies involving a deep learning algorithm (8% of the studies; Fig. 4d), features are usually extracted from each time-window at the last stage of the *pre-processing* step (Fig. 2) in order to describe each time window with features. Across the 66 papers, features are mostly calculated in the time-domain (TD) alone (~66% of the studies), but the frequency domain (FD) using the Fourier Transform is also commonly used (in 33% of studies; e.g. Rahman et al., 2018). Sometimes, wavelet features are also extracted in the time-frequency domain (TFD; e.g. Hokkanen et al., 2011).

Whatever the domain, features may provide information on (i) the motion intensity (MI), (ii) the orientation of the animal's body, (iii) the shape of the signal and (iv) the physical description of the movement (e.g., periodical, predictable). A non-exhaustive list of features used for each of these information categories along with the domain from which are calculated is provided in Appendix 3. Equations for these features are detailed in Supplementary Data B3. It is worth noting that in the 66 surveyed papers, only 21% of studies used a single type of features to develop the classification model thereafter (Fig. 4d), most often the MI (10%) and orientation features (3%). Hence, 79% of the studies used a combination of several type of features, the most frequent combinations found being the (i) MI and orientation features (33%), (ii) the MI, orientation and physical features (10%) and (iii) all the above feature types together (14%). Recommendations for features to compute to get good predictive performance are suggested in the Section 6.

At the end of the pre-processing step, each time-window is defined by a set of accelerometer features which are associated with the behaviour observed at the same time during data collection.

4. *Development of a behavioural classification model*: Splitting, modelling and validation options

In this step, a classification model to predict ruminant behaviours (outputs) from accelerometer features (inputs) previously extracted is trained and validated. The 66 papers are reviewed by addressing the three sub-steps applied in model development (Fig. 2), that is, (i) *Dataset Splitting*, (ii) *Model Training* and (iii) *Model Validation*. The approach adopted in the studies varies according to three main sets of choices regarding (i) the technique applied to validate the model and the criterion used to split the dataset, (ii) the classifier used and (iii) the way in which the performance of the model is reported. Fig. 5 summarises the options adopted in this literature. Appendix 4 presents detailed profiles of the 30 core papers (see Supplementary Materials A3 for the details of the remaining 36 papers). In the following sub-sections, we elaborate on the specifics of the three sets-of-choices made.

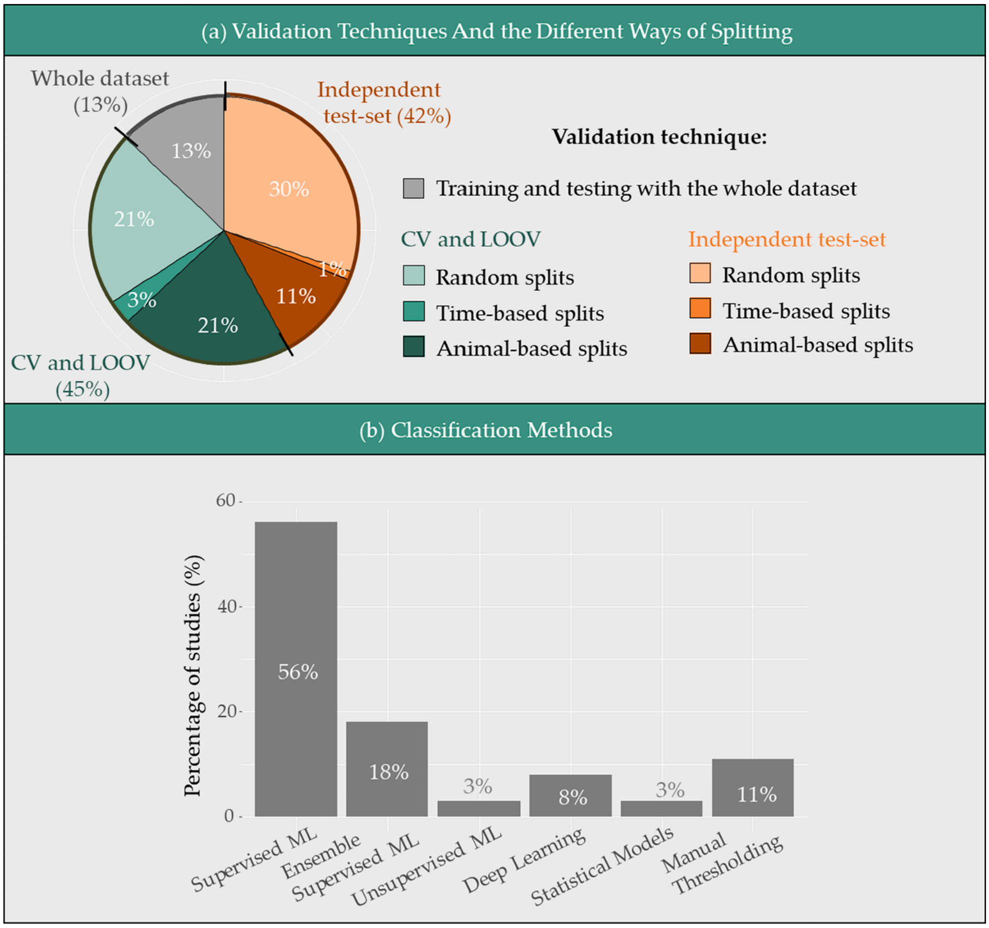

Fig. 5. In *Development of a Behavioural Classification Model*: Distribution of the different techniques used in the surveyed papers (N = 66) to (a) validate the model using train and set-tests and (b) classify the time-windows into behavioural categories from the accelerometer features.

4.1. Dataset splitting: Validation techniques and options to split the dataset

Among the 66 reviewed papers, 42% split the time-windows dataset into a *training-set* and a *test*-set (e.g. Martiskainen et al., 2009; Fig. 5a). This step involves using a percentage of the dataset to train the model (*training-set*) and the remaining percentage to assess the performance of the model independently (*test-set*). Cross-Validation (CV) and Leave-One-Out validation (LOOV) are used in 45% of the studies (e.g. Mansbridge et al., 2018; Fig. 5a). Using CV, the dataset is split into K folds, with K-1 folds used to train the model, and the Kth used to test it. The LOOV is the specific case of the CV where there are as many folds as there are observations in the dataset. The procedure is repeated K times, so that all folds have been used to test the model. Finally, 13% used the whole dataset both for training and validation (e.g. Busch et al., 2017; Fig. 5a). Recommendations on the different techniques used and their impact on the reliability of the model validation are discussed in Section 6.

When the dataset is split into several portions in the surveyed papers, either into a training data-set and test data-set, or into a cross-validation procedure, one of the following three criteria is usually chosen to split the dataset (Supplementary Data B4):

- *Random splits:* The split is done in a completely random way; that is, the partitioning of the time-windows into *training* and *test* sets is done regardless of the animals and periods from which the time-windows came (e.g. Dutta et al., 2015). This option is used in the 63% of studies (Fig. 5a).
- *Time-based splits:* The split is done regardless of the animals from which the time-window came, but the period when the animals are observed is taken into account (e.g. Riaboff et al., 2020). Validation is therefore carried out by considering the time-period associated to the time-window. This option is used in 5% of the studies (Fig. 5a).
- *Animal-based splits:* The split is drawn from a different sample of animals than those used to train the model (e.g. Tamura et al., 2019). This option has been used in 32% of the studies (Fig. 5a).

Recommendations on the criterion to be used for splitting the dataset and its impact on the reliability of the model validation, are also discussed in Section 6.

4.2. Model training: Options for the classifiers

The objective of this step is to train and validate a classification model that best discriminate the different behaviours (Fig. 2). Of the 66 reviewed papers, six main categories of models can be identified:

- *Supervised Machine Learning (SML):* This category of classifiers has been used in 56% of the studies (Fig. 5b). It mainly includes Linear Discriminant Analysis (LDA; Giovanetti et al., 2017) or Quadratic Discriminant Analysis (QDA; Barwick et al., 2018), Support Vector Machines (SVM; Martiskainen et al., 2009), k-Nearest Neighbour (k-NN; Sakai et al., 2019), Naïve Bayes models (NB; Benaissa et al., 2017) and Decision Trees (DT; Robert et al., 2009). The best hyperparameters for each model are usually found using Grid Search and a validation dataset (e.g. Martiskainen et al., 2009).

- *Supervised Ensemble Machine Learning (SEML):* This category of classifiers has been used in 18% of the studies (Fig. 5b). It essentially involves ensemble methods for classification, such as Random Forest (RF; Lush et al., 2018), eXtreme Gradient Boosting (XGB; Riaboff et al., 2020) or Adaboost (ADA; Dutta et al., 2015). The best hyperparameters for each model are usually found using Grid Search and a validation dataset (e.g. Riaboff et al., 2020)
- *Unsupervised Machine Learning (UML):* This category of classifiers has been used in 3.2% of the studies (Fig. 5b). It especially includes the k-means classification (Vázquez Diosdado et al., 2015).
- *Deep Learning (DL):* These methods have been used in 8% of studies (Fig. 5b). This category of classifiers includes the different varieties of Artificial Neural Networks, including Multilayer Perceptions (MLP; Nadimi et al., 2012), Convolutional Neural Networks (CNN; Peng et al., 2019) and Recurrent Neural Networks (RNN), notably Long Short-Term Memory (LSTM; Peng et al., 2020). It should be mentioned that this category is somewhat marginal to the other listed categories as accelerometer features are not necessarily used as model inputs, and accelerometer time-series are usually used instead.
- *Statistical Models (SM):* This category has been used in 3.2% of the studies (Fig. 5b). It includes generalised linear models like logistic regression (LR; le Roux et al., 2019) or models based on a Markov processes like Hidden Markov Models (HMM; Konka et al., 2014).
- *Manual Thresholding (MT):* This classification has been used in 11% of the studies (Fig. 5b). The manual assignment of thresholds for each feature to discriminate behaviours is done using the feature distribution, either with the observational data directly (Arcidiacono et al., 2017), or by first modelling the feature distribution (González et al., 2015).

Many papers test several classification methods comparatively from several categories of classifiers to find the one with the best classification performance in their study (Dutta et al., 2015; Hu et al., 2020; Riaboff et al., 2020; Smith et al., 2016). These comparisons will be used to support the recommendations on the models to choose in Section 6.

4.3. *Model validation: Options to evaluate model performance*

The *validation* step aims to assess the accuracy and robustness of the developed model (Fig. 2). This evaluation includes the three distinct sub-steps that were typically found across the 66 articles:

- *Predictive testing:* Behaviours are predicted for each time-window from the developed model, using the test-data-set (42% of the studies), CV technique (45% of the studies) or the whole dataset (13% of the studies).
- *Confusion matrix calculating*: The predicted behaviours of each time-window are compared to the actual observed behaviours (references obtained from the observations) as a confusion matrix. When more than 2 behaviours are predicted, a confusion matrix per behaviour is usually used to calculate true positives, false positives, true negatives and false negatives associated to each behaviour.
- *Performance analysis:* The performance recorded in the confusion matrix is analysed using a variety of standard metrics for model evaluation, the most common ones identified in the 66 papers being accuracy, F-score measure, Cohen's Kappa, Area Under Curve, sensitivity, specificity and precision. The formulae for these metrics are provided in Supplementary Data B5 and a detailed explanation can be found in Sokolova and Lapalme (2009). Depending on the studies, performance are related for the overall model (e.g. Watanabe et al., 2008), for each behaviour (e.g. Hokkanen et al., 2011), or for both (e.g. Alvarenga et al., 2016). Recommendations on how to report the performance of models are provided in Section 6.

5. Predicting ruminant behaviour: Good overall and by-behaviour performance, but still limitations for field applications

We profile the findings from the 66 surveyed papers, summarising their results in terms of the metrics used. Fig. 6 summarises the overall predictive performance for the surveyed studies using overall accuracy, Kappa, F-score and precision, and also profiles performance for each of the predicted behaviours using the accuracy, sensitivity and specificity per behaviour. Appendix 5 details the overall performance and performance-by-behaviour for the 30 core papers (see Supplementary Materials A4 for the performance of the remaining 36 papers). In the following sub-sections, we summarise the main positive and negative trends that emerge from these results and consider practical aspects regarding the use of these technologies in commercial farms.

5.1. *High predictive performance for the common behaviours*

Across the surveyed papers, overall predictive performance of models tend to be very good on key metrics (Fig. 6a): accuracy is > 80% in > 75% of the 66 papers, with ~50% of them achieving an accuracy of > 90% for correct classifications. The Cohen's Kappa scores for the agreement between the predicted and observed behaviours is also very high, at > 0.8 (on a scale from 0 to 1) in more than half of the surveyed papers. Almost 40% of the studies reached a F-score of > 0.8 (on a scale from 0 to 1), corresponding to excellent sensitivity and precision in the classifications made. Finally, the precision scores were also > 80% in more than the half of the surveyed papers, and > 90% in 17% of them.

The predictive performance by behaviour separately is also good for key metrics, at least for *common predicted behaviours* (i.e., feeding, ruminating, resting, moving, lying, standing; Fig. 6b and 6c). Indeed, for each of these common predicted behaviours, accuracy and sensitivity scores were > 80% in 50% of the surveyed papers. Specificity scores were > 80% in 75% the surveyed studies, and exceeded 90% in > 50% of them.

These results show that the prediction of the dominant ruminant behaviours from raw accelerometer data is quite feasible and successful. Thus, these methods provide a good basis for prediction of ruminant behaviour in the field. Note that these basic behaviours can in turn be used to quantify other issues, such as animal performance, health or welfare.

5.2. *Poor prediction of rarely observed behaviours and large sets of behaviours*

Although common ruminant behaviours can be predicted relatively accurately (i.e., feeding, ruminating, resting, moving, lying, standing; Fig. 6b and 6c), some other behaviours are harder to predict. The latter include those that are only expressed occasionally by animals or ones that tend to be rarely studied. For instance, transitional behaviours such as "lying down" and "standing up" are often poorly predicted (Martiskainen et al., 2009; Vázquez Diosdado et al., 2015). In the surveyed papers, the best reported sensitivities and specificities for "lying down" and "standing up" reached only 86% and 85%, respectively (Wang et al., 2018). Similarly, some maintenance behaviours (e.g. urinating, drinking), self-grooming behaviours and social-interaction behaviours tend to have lower accuracies, at < 80% (Lush et al., 2018; Rodriguez-Baena et al., 2020). Some of these prediction issues clearly arise from a lack of sufficient data with which to train models (Fogarty et al., 2020). Arguably, these predictive shortfalls present obstacles to some assessments as common behaviours are not always sufficient to infer certain events (O'Leary et al., 2020). In particular, transitions between postures can be important indicators to assess animal health (e.g., lameness detection; Yunta et al., 2012), grooming may be a relevant indicator of well-being (Keeling, 2019), and maintenance behaviours, such as urinating, may be useful to address environmental issues (Lush et al.,

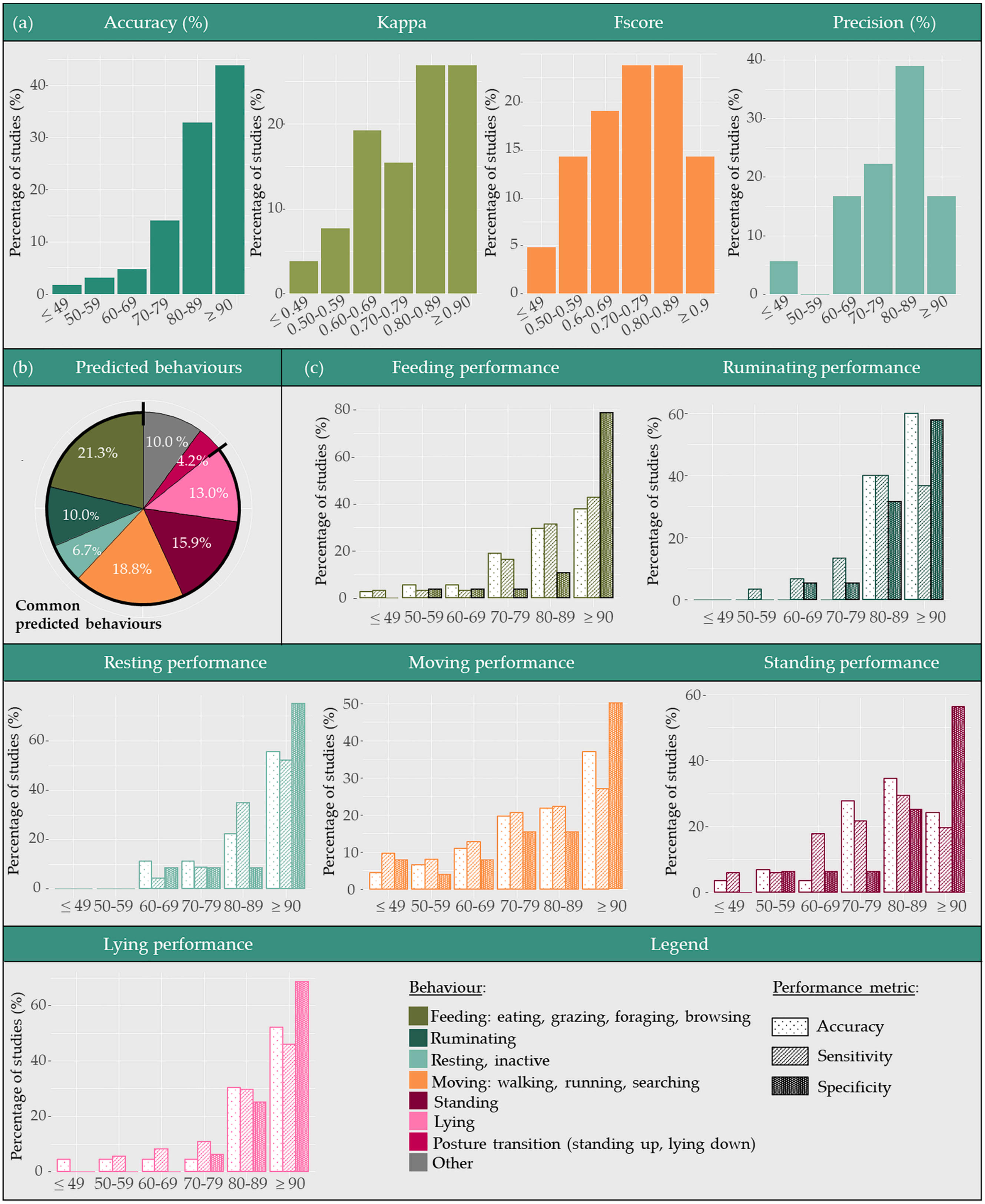

Fig. 6. *Performance* obtained in the surveyed papers (N = 66); overall performance are reviewed based on the accuracy, Kappa, Fscore and precision metrics (a); the performance reached for the main behaviours predicted in the literature (b) are reviewed based on their accuracy, sensitivity and specificity (c).

2018).

Furthermore, prediction issues can arise when the number and diversity of behaviours to predict with the model increase, since prediction performance tends to decrease in such situations (Martiskainen et al., 2009; Peng et al., 2020). For instance, Hänninen (2010) reported a decrease in predictive accuracy from 87% to 63% when the behavioural categories being predicted increased from 3 to 6, respectively. Bishop-Hurley et al. (2014) have argued that the limited range of predicted behaviours available is quite a serious challenge in commercial uses of this technology, as there are many practical problems that require tracking and capturing a diversity of behaviours (e.g., the detection of lameness in grazing cattle; Riaboff et al., 2021).

5.3. Experimental performance may not extend to field applications

Although current models report impressive performance, this performance is very much experimental-based and may not extend to commercial deployments. Recall that an original dataset can be split in different ways for model validation purposes (Sections 4.1). In this respect, the performance obtained with the same model seems to be greatly affected by the split used. Several studies report a substantial decrease in performance when validation is done with animals other than those used to train the model, that is, using an *animal-based* split, compared to a *random* or *time-based* splits. For instance, Rahman et al. (2018) reported a 0.4 decrease in F-score (scale 0–1) using the same original dataset when it was split by animal in a Leave-One-Out-Animal validation technique, compared to the random split using a *k*-fold Cross-Validation. In the same way, Riaboff et al. (2020) also reported a 13% decrease in accuracy when splitting the dataset by animal as opposed to a time-based split, using the independent test-set in both cases. These decrements in model performance come from high inter-animal variability based on differences in physical characteristics (e.g., muscles, tendons, joints) that lead to different expressions of the same behaviour (e.g., motion intensity, speed, posture; Barwick et al., 2020); hence the same behaviour ends up being expressed by different accelerometer patterns. Consequently, splitting an original dataset by animal can lead to markedly different feature-distributions between the training and test-sets and consequential poor predictive performance (Rahman et al., 2018). A decrease in performance can also be expected when validation is applied with a time based-split, compared to a random split (Riaboff et al., 2020). Indeed, behaviour prediction may also be a non-stationary problem since animal behaviour is continuously changing, constantly adapting to varying management and environmental conditions. Thus a model trained at one time for a given behaviour may not generalise to what nominally is the same behaviour at a later time (Vázquez-Diosdado et al., 2019). This lack of model generalization from one animal to another and from one time-period to another is an additional major constraint for applications in the field.

5.4. Technical limitations of the experimental devices for a commercial use

Although beyond the scope of this review, it is worth mentioning that most of the devices reviewed here in an experimental context can be used for research purposes, but are not compatible with commercial use. For example, in most studies, manual data extraction (Robert et al., 2009), calculation time for extracting complex features (Smith et al., 2016) or predicting behaviours with computationally intensive algorithms are unlikely to be suitable for long-term or for scaled-up use when one considers the battery-life of these devices. These technical issues are also a major constraint to the development of marketable systems.

6. Recommendations & lines to explore for future studies

In this section, we propose an overall optimal framework for achieving good predictive performance based on detailed choices at each step in the methodology (Section 6.1). Then we propose some perspectives to address the challenge of increasing the range of well-predicted behaviours (Section 6.2), and to address the challenge of increasing the generality of models (Section 6.3). Finally, we briefly consider the issues surrounding the transitioning of these technologies from a research environment to commercial settings (Section 6.4).

6.1. General recommendations for good and robust predictive performance

General recommendations for further research investigations are divided into proposals for each of the key techniques identified in the 3 steps of the methodology. Key techniques and proposals are made to achieve good and robust predictive performance in a properly designed

Table 2
Overall recommendations for the key techniques identified for each step of the 3-step methodology to achieve high predictive performance in a properly designed scheme for model evaluation, based on the reviewed papers (N = 66).

3-Steps	Key Techniques	General recommendations	
Data Collection	Equipped animals	Include as much variability as possible: > 25 animals from > 2 different farms	
	Observation of animals	Include as much variability as possible: Continuous observations during > 40 h	
	Sampling rate	[10 Hz; 20 Hz]	
	Sensor positioning	Feeding behaviour	Jaw; Neck; Ear
		Motion; Restless	Ear; Neck
		Posture; Transition	Leg
Data Pre-Processing	Cleaning raw data	Clean any missing data.	
	Additional time-series	Calculating Amag, OBDA and/or VeDBA, pitch, roll, sway, etc, and combining several time-series.	
	Segmentation	WS: [3 sec; 30 sec]; A mix of different window-sizes; Adding overlap	
	Features	Combination of MI, orientation, shape and physical features	
Development of a Model	Validation technique; dataset splitting	Split according to the *individual* criterion; Independent dataset to test the model; Balanced dataset between the different behaviours.	
	Classifier	*Supervised Machine Learning*: SVM	
		Supervised Ensemble Machine Learning: RF, XGB	
		Deep Learning: PNN, LSTM, CNN	
	Metrics of performance	Provide metrics of performance for both the overall model (e.g. Fscore, Cohen's Kappa) and for each predicted behaviour (e.g. sensitivity and specificity per behaviour).	

Note: For the abbreviations used in the table, we refer to the list of abbreviations provided at the beginning of the review.

scheme for model evaluation, with respect to the assessment of the 66 papers reviewed. The main recommendations are summarized in Table 2 and detailed in the following sections.

6.1.1. *General recommendations for data collection*

In data collection, a general recommendation can be made about the number of animals to equip during the experiment. Good performance of prediction have already been reached with few equipped animals (<10 animals; e.g. Arcidiacono et al., 2017) but the lack of variability in the dataset can disimprove the generality of the model (see Section 5.3). While this is not a major problem when the model is used thereafter with the same animals for experimental purposes, it is a critical issue for commercial deployments. Therefore, the more animals that are involved, the greater the variability between individuals in the training set, the more robust the model will be. In this respect, equipping 25 animals (Smith et al., 2016) seems to be a reasonable minimum to achieve both robust prediction, that is, good performance using appropriate technique for validation (see Section 6.1.3 to follow). Although rarely seen in the papers surveys, equipping animals from at least 2 different farms is also recommended to improve model robustness. In the same way, the longer the animals are observed, the more inter- and intra-individual variability is taken into account in the dataset, the more robust the model will be. In this regard, about 40 h of observation appears to be the minimum time to achieve both high and robust prediction (González et al., 2015). It is worth mentioning that continuous observation ensures correct association between accelerometer data and observed behaviours over continuous intervals, to both train and validate the model properly. This technique is thus to be preferred to scan sampling (e.g. Tamura et al., 2019) as it is necessary to extend the observed behaviour between the different observation scans, while the animal may have expressed another behaviour over the interval. Video recordings also reduce the degree of uncertainty related to direct observation since it is possible to go back to the observations if necessary, but behaviours should be encoded after filming the animals, which can be more constraining.

Recommendations can also be made on the sampling rate, as studies show that reducing the sampling rate (using down sampling) reveal performance decrements below 10 Hz (Benaissa et al., 2018; Walton et al., 2018). Furthermore, the benefit of sampling rates higher than 20 Hz is not clear since high predictive performance tends to be achieved for many behaviours with 20 Hz (Peng et al., 2019). As increasing the sampling rate decreases the battery life, a sampling rate between [10; 20 Hz] is probably an appropriate trade-off (Benaissa et al., 2018). However, it should be noted that a thorough analysis of the maximum frequencies that can be generated by animal movements according to their species and breed is necessary to make strong recommendations on optimal sampling rates. To the authors' knowledge, such a study has not yet been carried out for livestock.

Sensor positioning must be chosen according to the behaviours being predicted. Indeed, neck-, jaw- and ear-mounted sensors seem adapted to predict feeding behaviours, such as grazing, feeding, ruminating, both in cattle (Smith et al., 2016; Weizheng et al., 2019) and sheep (Barwick et al., 2018; Giovanetti et al., 2017). Ear- and neck- mounted sensors can be recommended to predict behaviours related to motion and restlessness, both in cattle (Peng et al., 2019) and in sheep (Fogarty et al., 2020). Leg-mounted sensors seem the most suitable to discriminate between postures and detect the lying down and standing up transitions, both in cattle (Wang et al., 2018) and goats (Zobel et al., 2015).

6.1.2. *General recommendations for data pre-processing*

Although removing noise from the raw accelerometer signal does not necessarily appear to improve performance based on studies assessing its impact on performance (Riaboff et al., 2019), it may be advisable to use such filters when the accelerometer signal is likely to be very noisy (e.g. when using ear-mounted sensors; Chapa et al., 2020). Removing accelerometer data indicating that the data could not be generated for a time stamp (e.g. artificial zeros, outliers, symbols, etc) is recommended to avoid biasing the analysis later on.

Calculating new time-series appear really relevant to reach a high level of performance prediction. Most especially, OBDA and/or VeDBA extracted from dynamic acceleration are useful to predict behaviours with different levels of activity/energy expenditure (Lush et al., 2018; Vázquez-Diosdado et al., 2019) while pitch, roll, sway, etc, extracted from the static acceleration are useful to predict behaviours involving different posture and body/head tilt (Lush et al., 2018; Zobel et al., 2015). The magnitude of the acceleration, called Amag in this review, also appears relevant when the axis direction is not helpful (e.g. categories of behaviour related to the activity level) or not interpretable (e.g. different sensor positioning to train a same classifier, collar subjected to rotation, etc) (Kamminga et al., 2018; Mansbridge et al., 2018; Tamura et al., 2019). Finally, combining different time-series is promising to predict a large set of behaviours, involving different level of motion intensity and postures, as it is a way to provide complementary inputs into the classifier (Lush et al., 2018; Riaboff et al., 2020; Vázquez Diosdado et al., 2015).

In the 66 surveyed articles, the main differences concerning the segmentation step are related to the chosen window size and whether or not an overlap is added. In that respect, it seems that WS < 10 sec result in better performance than larger windows (Lush et al., 2018; Peng et al., 2019; Robert et al., 2009), possibly because short WS have less variability or transition between behaviours (Banos et al., 2014). However, WS > 10 sec capture the high motion variability found in activities with a complex description (Banos et al., 2014) and therefore may be relevant to predict behaviours with different sub-units, such as grazing which involves alternation between chewing and biting (Gibb, 1996). Large WS > 30 sec may reduce the number of time-windows available as each window spanning several behaviours are usually removed from the data set (Andriamandroso et al., 2017): The longer the windows, the more windows spanning several behaviours, the less time-windows are available for training and validation. Such large window size may also lead to optimal performance in experimental context as validation is usually done using single-behaviour windows (Tamura et al., 2019), while on the field there will surely be a mix of behaviours within WS > 30. Therefore, adopting a window size between [3 sec; 30 sec] appears to be a good trade-off. Another interesting trade-off is to use several different window sizes to calculate features instead of using a single one (Hu et al., 2020), as features extracted from time windows of mixed sizes, ranging from 2 s to 15 s, significantly improved the classification of behaviour. It is worth noting that the best WS is also highly dependent on the sampling rate and behaviours to predict. Thus, we encourage researchers to identify the best WS for their study in a preliminary analysis. Adding overlap is also a way to increase the predictive performance, especially for large WS when there are not enough windows available for training and validation (Riaboff et al., 2019).

Regarding the features to extract from each time window, using a combination of several types of features, including motion intensity, orientation, shape and physical features (Appendix 3) is highly recommended, especially when a broad range of behaviours is predicted. Indeed, high predictive performance are usually obtained combining different types of features, as it provides an exhaustive and complementary description of the time-series into the classifier (Riaboff et al., 2020; Walton et al., 2018).

6.1.3. *General recommendations for model development*

Regarding the techniques used to train and test the model, the most important decision seems to involve the criterion used to split the data set. As explained in Section 5.3, several studies report a substantial decrease in performance when validation is done with other animals than those used to train the model compared to random or time-based splits due to the variability between animals and time periods. Therefore, splitting the dataset randomly does not seem appropriate for testing the robustness of the model and may lead to overly-optimistic performances, not representative of those obtained in a commercial context. In this respect, we recommend to use an animal-based split to evaluate the robustness of the model properly (Rahman et al., 2018). Testing the model with an independent dataset (Riaboff et al., 2020) is also a good way to ensure that the parameters used to train the model (window size, overlap, classifier, hyperparameters, etc.) are suitable for another dataset than the one used to develop the model and are not too specific, which could lead to overly-optimistic performance. Consistently, testing the model with the same dataset than the one used to train it, as done in 13% of the studies (Fig. 5a), conflates training and testing, and is thus a validation technique to be proscribed. Another reason leading to unreliable and sub-optimal performance is the unbalanced distribution of time-windows within the different behaviours to predict; that is, with some behaviours over-represented, and other under-represented. In this way, using stratified CV (Riaboff et al., 2019) or using under sampling (Fogarty et al., 2020; Sakai et al., 2019) to improve the balance of the dataset between the different behaviours to predict is also recommended.

In studies comparing several classifiers, best predictive performance have often been obtained with SVM, RF and XGB (Mansbridge et al., 2018; Riaboff et al., 2020; Vázquez Diosdado et al., 2015). Although Deep Learning is seldom used in the community, contrary to Human Activity Recognition (e.g. Song-Mi Lee et al., 2017), excellent performance has been achieved with PNN and LSTM models (Peng et al., 2020; Weizheng et al., 2019) and more recently with a Convolutional Neural Network (Pavlovic et al., 2021). These Machine Learning and Deep Learning classifiers are thus recommended for the classification of ruminant behaviours from accelerometer data.

Regarding the way in which performance metrics are reported, it seems that providing the overall performance of the model (e.g. overall accuracy and Cohen's Kappa; González et al., 2015), but also the performance for each behaviour (e.g. accuracy, sensitivity and specificity of each predicted behaviour; Alvarenga et al., 2016), is the best way to ensure a complete and unbiased evaluation of the model (Decandia et al., 2018).

6.2. How to increase the range of well-predicted behaviours

As explained in Section 5.2, current studies encounter difficulties in predicting (i) rarely observed, short and transitional behaviours (e.g. grooming, drinking, urinating, transition between postures) and (ii) a broad spectrum of behaviours using a single accelerometer sensor. To overcome these difficulties, we propose several lines to investigate in the 3-step methodology, detailed in the following sub-sections.

6.2.1. Data Collection: Use additional devices

Adding several accelerometer devices at different positions on an animal is a good way to extend the spectrum of well-predicted behaviours (Benaissa et al., 2017) even though this solution may not be practicable for use in the field. Alternatively, additional sensors alongside accelerometers could be used, such a magnetometer or a gyroscope. These additional devices will provide more information that should improve predictive performance. For instance, Walton et al. (2018) used a gyroscope in addition to a collar-mounted accelerometer and achieved good performance for the lying behaviour, a behaviour that is usually quite hard to predict with collar-mounted sensors (Hamalainen et al., 2011). Furthermore, Guo et al., (2018) found that gyroscope features were included in the top-5 most important features for discriminating grazing from non-grazing behaviours, highlighting the potential utility of additional sensing capabilities. Also, Brugarolas et al. (2013) have argued that adding a gyroscope should improve the classification of cyclic behaviours. Nevertheless, further investigations are needed to assess the utility of gyroscopes as sometimes they do not deliver major improvements (Kleanthous et al., 2019; Sakai et al., 2019). Similarly, adding a magnetometer also seems to improve predictive performance (Sakai et al., 2019). Additional data can also be used to reassess the outputs at the end of the prediction. For example, Wang et al. (2019) observed a significant improvement in the prediction of standing and feeding behaviours that were poorly predicted with the Adaboost model by re-evaluating the prediction with location data, using D-S evidence theory. Indeed, location sensors are also relevant, as knowing where the animals are can provide additional about what they are doing (Fogarty et al., 2021). The addition of sensors other than the accelerometer in the same device is therefore probably a good way to improve the prediction of behaviours that are difficult to predict using an accelerometer on its own.

6.2.2. Data pre-processing: Investigate new segmentation techniques

Most studies use a fixed-window size when segmenting a time-series; that is, the time interval used to split the signal is kept constant across the whole time-series (see Section 3.3). However, in related areas, such as Human Activity Recognition from accelerometer sensors, different segmentation techniques are often used and should be investigated to predict ruminant behaviours. For example, the signal can be split into different window-sizes determined by partitioning (top-down approach) or aggregating (bottom-up approach) on a chosen criterion (Keogh et al., 2001). In this way, the segmentation can be carried out at a specific location within the time-series, in a way that minimizes the intra-segment variability and maximizes the variability between the different segments. This approach looks promising as a way to avoid splitting the signal in the middle of the expression of the same behaviour, an issue that can arise in predicting short and transitional behaviours (such as chewing or biting in the grazing process or standing up/lying down transitions).

6.2.3. Model development: Use temporal structure in time-series or sequence of behaviours

In general, good predictive performance can be achieved using Machine Learning (ML) and Ensemble methods but often these classifiers do not take the temporal structure within the time-series into account. Although they are not yet widely used in this literature, classifiers that make use of the dependence between samples in the accelerometer time-series, such as LSTM, seem promising as predictive models for varied and rarely-observed behaviours (Peng et al., 2020, 2019). In that regard, the range of Machine Learning and Deep Learning classifiers designed for time-series classification specifically (Bagnall et al., 2017; Ismail Fawaz et al., 2019) deserve to be tested in ruminant behaviour prediction. It should be noted that accelerometer features could also be used with such classifiers instead of using the raw time-series, calculating features in each time-window and then retaining the time stamp of the corresponding time-window.

Another option is to consider the temporal structure within the continuous sequence of the expressed behaviours. Assuming that the sequence of behaviours expressed by animals is not random, and that there is, on the contrary, a direct relationship between the behaviour expressed at window t and that expressed at window $t + 1$, then one could use this temporal structure to improve the prediction of behaviours. For instance, Riaboff et al. (2020) used a Viterbi HMM algorithm to reassess the predicted behaviours by a ML model for each time-window considering the temporal structure within a behaviour sequence. Such an approach leads to an improvement in predictive performance, especially for behaviours with the lowest sensitivities. Considering temporal structure, either in the accelerometer data or a continuous sequence of behaviours, can thus improve prediction for certain behaviours.

6.3. *Methods to improve of model generality*

As explained in Section 5.3, current studies encounter difficulties in developing robust models that generalise well; that is, models that maintain good predictive performance when tested on different animals and/or over different time-periods. These difficulties present a major obstacle to the commercial use of these technologies in farming. To overcome these issues, we propose several methods and techniques to investigate in a 3-step methodology, detailed in the following subsections.

6.3.1. *Data collection: Include as much variability as possible*

As explained in Section 6.1.1, equipping a large number of animals from different farms, each observed continuously during a long time is essential to acquire a training dataset representative of the observed variability between animals (Hu et al., 2020). To further develop robust models, including different breeds (Andriamandroso et al., 2017), different parities and lactation periods, observing animals at different times of the year (Riaboff et al., 2020) and possibly under different breeding conditions (e.g. different types of housing, such as straw yards vs. cubicles) or different management and feeding practices (e.g. different grass heights for pasture-based systems; Guo et al., 2018) seems to be a promising way to include as much variability as possible in a training dataset, and consequently to develop models that have greater generality.

6.3.2. *Model development: Use classifiers that prevent overfitting and/or adaptive over time*

As mentioned in Section 6.1.3, SVM, RF or XGB classifiers lead to good predictive performance. Moreover, beyond the good performance obtained in an experimental framework, such classifiers include hyperparameters specifically designed to avoid *overfitting* (i.e., where a model becomes too adapted to its training dataset, and therefore does not generalise well). For example, the margin width relative to the hyperplane in the SVM model (Burges, 1998) can be adapted (*hyperparameter C: cost of constraints violation*) so that it allows some misclassification in the training dataset for observations that are too far from the mean distribution (i.e. *outliers*). Ensemble machine learning methods such as RF, Adaboost or XGB use several weak classifiers to avoid overfitting and also include hyperparameters to optimize (Breiman, 2001; Friedman, 2001). In that regard, some models like manual thresholding or DT are more prone to overfitting and thus the excellent performance obtained with such algorithms in an experimental context need to be qualified (Tamura et al., 2019).

Another way to improve the generality of a model is to use an evolving model, that adapts the prediction as it goes along according to the data it receives. For example, Vázquez-Diosdado et al. (2019) first predicted a new instance with a trained and non-evolving off-line model (*k*-NN classifier) and then use an evolving on-line model (k-means classification) which was updated as the data were received to reassess the predictions from the off-line model. When this combination is tested on new animals over new time periods, better performance was obtained than when the offline model was used on its own.

6.4. *Moving from experimental to commercially-usable systems*

As explained in Section 5.4, it seems important to point out that most of the methods mentioned in the review were developed in experimental settings and do not yet transfer to on-farm commercial use at scale. Making a given tool compatible with commercial use in farming (positioning of sensors, automatic transfer of comprehensible data, battery life, etc.) is already the subject of many studies (Kuźnicka and Gburzyński, 2017; Nadimi et al., 2008). Several lines have been investigated, including the automatic transfer of data using edge computing (Hendriks et al., 2020) or limiting the volume to transfer (Benaissa et al., 2018; Kuźnicka and Gburzyński, 2017). Decreasing computation time by (i) adapting the pre-processing (increasing the window size; Walton et al., 2018), (ii) extracting variables with low computational requirements (Vázquez-Diosdado et al., 2019) or (iii) selecting suitable classifiers (Vimalajeewa et al., 2021) is also an issue addressed in several studies.

7. Conclusions

This paper has reviewed the key literature predicting ruminant behaviours from raw accelerometer data. In the paper, we have (i) profiled the main characteristics of 66 key articles on this topic area within the 3-step methodology used in the area, (ii) reported on main results that have arisen from these studies, (iii) detailed the challenges and issues that arise in this area and (iv) made recommendations and suggested lines to investigate for resolving these challenges in future studies. Overall, the good predictive performance obtained in most of the studies supports the potential of this methodology to predict ruminant behaviour from accelerometer sensor data, across a variety of farming applications. The main limitations that arise include the difficulties surrounding the prediction of certain behaviours and the lack of generality and robustness in developed models. For that purpose, we encourage researchers to include as much variability as possible in their datasets and to use classifiers that prevent overfitting. Some lines to investigate are also proposed to address these issues, but, in general, we would argue that the techniques used need to be objective-driven (e.g. adapted to the behaviours to predict, to the desired application on farms, etc). In this regard, we recommend that a series of pilot studies be carried out to identify the most appropriate methods and techniques for the objectives of the study. Finally, while the large range of relevant techniques and methods reported in ruminants and related communities should help to solve the mentioned limitations, further work is still needed to combine performance and operability of the systems developed.

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 (i) Science Foundation Ireland (SFI) to the *Insight SFI Centre for Data Analytics* under Grant Number 12/RC/2289_P2 and (ii) SFI and the Department of Agriculture, Food and Marine on behalf of the Government of Ireland to the *VistaMilk SFI Research Centre* under Grant Number 16/RC/3835.

Appendix 1. ***In Data Collection*****: Detail of the key choices made when predicting ruminant behaviour from accelerometers in each surveyed paper (note - based on the 30 core articles*; see Supplementary data A1 for the other 36 studies).**

References	Breed[1]	No animals	Accelerometer	Position	Sampling frequency	Additional sensor	Encoded behaviours[2,3]	Duration of observation[4]
Martiskainen et al., 2009	Ayrshire, Holstein	30	ADXL330, 3D, ± 3 g	Neck	10 Hz		S, L, Rum, F, W, lame W	95.5 hours
Dutta et al., 2015	Holstein	24	HMC6343, 3D	Neck	10 Hz	3D-Magnetometer	G, Sea, Rum, R, Scr, U	6 hours
Gonzalez et al., 2015	Brahman and cross, cross Belmont Red	58	HMC6343, 3D, ± 4 g	Neck	10 Hz	GPS	F, Rum, W, R, O	43 hours
Vázquez Diosdado et al., 2015	Holstein	6	MMA8451Q, 3D, ± 8 g	Neck	50 Hz		F, S, L, LD, SU	33.3 hours
Zobel et al., 2015	Cross Saanen	10	HOBO® Pendant G, 3D	Leg	0.02 Hz		L, S, perching, U, lifting rear legs, O	144 hours
Alvarenga et al., 2016	Merino	4; 6	AML, 3D, ± 8 g	Under-jaw	5, 10, 25 Hz		G, L, Run, S, W	3.36 hours
Smith et al., 2016	Holstein	24	HMC6343, 3D	Neck	10 Hz		G, W, Rum, R, O	6 hours
Abell et al., 2017	Angus and cross, Simmental, Gelbvieh	2	Smartbow	Ear, wither, Neck	1 Hz		L, S, W, Moun, O	Unspecified
Andriamandroso et al., 2017	Holstein, cross Blonde d'Aquitaine and Belgian	19	STMicro STM33DH (iPhone 4S), 3D	Neck	100 Hz	3D-Gyroscope 3D-Magnetometer	G, Rum, O	44 hours
Arcidiacono et al., 2017	Unspecified	5	KXTJ9, 3D, ± 8g	Neck	4 Hz		F, S	5 hours
Benaissa et al., 2017	Holstein	16	HOBO® Pendant G, 3D, ± 3 g	Neck, leg	1 Hz		L, S, F	96 hours
Giovanetti et al., 2017	Sarda	2	ADXL335, 3D	Under-jaw	62.5 Hz		G, Rum, R	42 hours
Barwick et al., 2018	Cross Merino and Poll Dorset	5	GCDC X16 mini, 3D	Neck, leg, ear	12 Hz		G, S, L, W	12.5 hours
Benaissa et al., 2018	Holstein	10	Axivity AX3, 3D	Neck	10 Hz		F, Rum, O	60 hours
Decandia et al., 2018	Sarda	8	BEHARUM, 3D	Under-jaw	62.5 Hz	Force sensor	G, Rum, O	3 hours
Lush et al., 2018	Welsh moutain	30	Daily Diary'tag, 3D	At the rear of the back	40 Hz		Fo, W, Run, S, L, U, Scr, Gro, Int	5.6 hours
Mansbridge et al., 2018	Cross Texel, cross, Suffolk, Mule	6	BMI160, 3D, ± 8 g	Neck, ear	16 Hz	3D-Gyroscope	G, Rum, O	16 hours
Walton et al., 2018	Cross Texel, cross Suffolk, Mule	6	BMI160, 3D, ± 8 g	Ear, neck	8, 16, 32 Hz	3D-Gyroscope	W, S, L	16 hours
Wang et al., 2018	Holstein	5	ADXL345, 3D, ± 8 g	Leg	1 Hz	Location sensor (RSS)	F, L, S, LD, SU, W, active walking	200 hours
Peng et al., 2019	Japanese Black	8	RT-BT-9axis IMU, 3D	Neck	20 Hz	3D-Gyroscope, 3D-Magnetometer	F, L, RumL, RumS, licking (social and salt), W, head butt	68 hours
Riaboff et al., 2019	Holstein	10	LSM9DS1, 3D; ± 2 g	Neck	59.5 Hz		G, W, L, S	5 hours
Sakai et al., 2019	Saanen	3	NinjaScan-light, 3D	Back	100 Hz	3D-Gyroscope, 3D-Magnetometer	L, S, G	12 hours
Tamura et al., 2019	Holstein	38	H30CD, 3D, ± 2 g;	Neck	20 Hz		F, Rum, S, L, W, D, O	6 hours (scan sampling)
Vázquez-Diosdado et al., 2019	Cross Lleyn, Aberfield, Explana and Berichon du Cher	17	InterQuark SE C1000, 3D	Ear	16 Hz	3D-Gyroscope	W, S, L	12 hours
Wang et al., 2019	Holstein	12	ADXL345, 3D, ± 2 g	Leg	1 Hz		S, L, W, active walking, SU, LD	4 hours
Weizheng et al., 2019	Holstein	5	MSR145, 3D, ± 10 g	Under-jaw	5 Hz		F, Rum, O	Unspecified
Fogarty et al., 2020	Cross Merino	12	Axivity AX3, 3D	Ear	12.5 Hz		G, W, S, L	6 hours
Hu et al., 2020	Merino	17	wGT3X-BT, 3D	Neck	30 Hz		G, Rum, W, S	8.5 hours
Peng et al., 2020	Japanese Black	3	RT-BT-9axisIMU, 3D	Neck	20 Hz	3D-Gyroscope, 3D-Magnetometer	F, RumL, RumS, L, S, L before calving, S before calving	150 hours
Riaboff et al., 2020	Holstein	86	LSM9DS1, 3D, ± 2 g	Neck	59.5 Hz		G, W, RumL, RumS, RL, RS, Gro, U, Int, Run, SU, LD, G while L	57.35 hours

Note: * The 30 papers presented in Appendix 1 out of the 66 reviewed were selected based on their complete methodological practices; that is, there is no main missing elements in the description of the material and methods used and the results are reported in a complete and understandable way. Furthermore, we selected papers predicting at least two distinct behaviours in order to display the predictive performance obtained in a challenging framework. The details of the remaining 36 articles are reported in Supplementary Materials.

[1] Cell background is displayed in dark green, orange or light green depending on the study was on cattle, sheep or goat, respectively.

[2] Behaviours observed at pasture are written on green background; behaviours observed in the barn are written on orange background.

[3] Notations used for behaviours: G: grazing, F: feeding, Rum: ruminating, R: resting, W: walking, S: standing, L: lying, RumL: ruminating while lying; RumS: ruminating while standing; RL: resting while lying; RS: resting while standing; SU: standing up, LD: lying down, Run: running, Scr: scratching, Sea: searching; Fo: foraging; Gro: grooming, Int: interaction, Moun: mounting; U: urinating, D: drinking, O: other.

[4] Depending on the studies, the time displayed is the exact or approximate duration of observation. Observations using time-sampling are notified in brackets.

Appendix 2. ***In Pre-processing*****: Key choices made when predicting ruminant behaviour from accelerometers in each surveyed (note - based on the 30 core articles*; see Supplementary Data A2 for the other 36 studies).**

	Key Decision Sets in Data Pre-Processing															
	Cleaning	Time series					Segmentation						Features categories			
Reference	Cleaning/Removing noise	Raw x, y, z time-series	Amag	AccDy, OBDA and/or VeDBA	AccSt, pitch, roll and/or sway	TS from other sensors	Strictly less than 10 sec.	10 ≤ WS ≤ 20 sec.	20 < WS ≤ 30 sec.	30 < WS ≤ 60 sec.	Strictly more than 60 sec.	Overlap	MI	Orientation	Shape	Physical
Martiskainen et al., 2009	🟩	🟩						🟩					🟩	🟩	🟩	
Dutta et al., 2015		🟩			🟩		🟩						🟩	🟩	🟩	
Gonzalez et al., 2015		🟩				🟩		🟩					🟩	🟩		
Vázquez Diosdado et al., 2015				🟩	🟩						🟩		🟩	🟩		
Zobel et al., 2015	🟩				🟩					🟩				🟩		
Alvarenga et al., 2016		🟩					🟩						🟩	🟩		
Smith et al., 2016	🟩		🟩	🟩	🟩		🟩	🟩	🟩				🟩	🟩	🟩	🟩
Abell et al., 2017	🟩	🟩					Unspecified						🟩	🟩		
Andriamandroso et al., 2017				🟩	🟩	🟩	🟩	🟩	🟩	🟩			🟩	🟩		
Arcidiacono et al., 2017		🟩					🟩						🟩	🟩		
Benaissa et al., 2017			🟩	🟩						🟩			🟩			🟩
Giovanetti et al., 2017	🟩	🟩	🟩							🟩			🟩	🟩		
Barwick et al., 2018		🟩	🟩					🟩					🟩	🟩		
Benaissa et al., 2018	🟩			🟩						🟩			🟩			
Decandia et al., 2018		🟩	🟩			🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩		
Lush et al., 2018		🟩		🟩	🟩		🟩	🟩					🟩	🟩		🟩
Mansbridge et al., 2018			🟩			🟩	🟩					🟩	🟩		🟩	🟩
Walton et al., 2018	🟩		🟩			🟩	🟩					🟩	🟩		🟩	🟩
Wang et al., 2018		🟩				🟩	🟩						Time-series			
Peng et al., 2019	🟩	🟩				🟩	🟩	🟩				🟩	Time-series			
Riaboff et al., 2019	🟩	🟩	🟩	🟩			🟩	🟩	🟩			🟩	🟩	🟩	🟩	🟩
Sakai et al., 2019		🟩		🟩	🟩		🟩						🟩	🟩	🟩	
Tamura et al., 2019	🟩		🟩							🟩			🟩			
Vázquez-Diosdado et al., 2019				🟩		🟩	🟩						🟩		🟩	
Wang et al., 2019	🟩	🟩					🟩						Time-series			
Weizheng et al., 2019			🟩	🟩			🟩	🟩	🟩	🟩			🟩			🟩
Fogarty et al., 2020		🟩	🟩				🟩	🟩	🟩				🟩	🟩		
Hu et al., 2020		🟩	🟩				🟩	🟩					🟩		🟩	
Peng et al., 2020	🟩	🟩		🟩		🟩	🟩	🟩				🟩	Time-series			
Riaboff et al., 2020		🟩	🟩	🟩	🟩			🟩				🟩	🟩	🟩	🟩	🟩

Note: * The 30 papers presented in Appendix 2 out of the 66 reviewed were selected based on their complete methodological practices; that is, there is no main missing elements in the description of the material and methods used and the results are reported in a complete and understandable way. Furthermore, we selected papers predicting at least two distinct behaviours in order to display the predictive performance obtained in a challenging framework. The details of the remaining 36 articles are reported in Supplementary Materials. The background of the cells is green when the technique has been applied, grey otherwise. Abbreviations used in the Table: AccSt: Static Acceleration; AccDy: Dynamic Acceleration; Amag: Magnitude of the acceleration; OBDA: Overall Body Dynamic Acceration; VeDBA: Vector Dynamic Body Acceleration; WS: Window Size; MI: Motion Intensity.

Appendix 3. Quantitative features used to describe the time-series into each time-window, organized by information-category and signal domain.

Category	Domain	Time-series	Example of Features	References
Motion Intensity	TD	x-axis, y-axis, z-axis	Standard-deviation, movement variation, median, first and third quartile, interquartile, minimum, maximum, range, root mean square	Williams et al., 2017 Barwick et al., 2018 Riaboff et al., 2020 Vázquez Diosdado et al., 2015 Robert et al., 2009 Benaissa et al., 2018 Riaboff et al., 2020
		Dynamic acceleration, *Amag*, OBDA, VeDBA	Mean, standard-deviation, movement variation, median, first and third quartile, interquartile, minimum, maximum, range, root mean square	
Orientation of the body	TD	x-axis, y-axis, z-axis	Mean, median	

(*continued on next page*)

(*continued*)

Category	Domain	Time-series	Example of Features	References
				Kleanthous et al., 2018 Alvarenga et al., 2016
		Static acceleration, pitch, roll, sway	Mean, median, standard-deviation, minimum, maximum	Lush et al., 2018
Characterization of the shape	TD	x-axis, y-axis, z-axis, pitch, roll	Skewness, kurtosis	Dutta et al., 2015
	FD	x-axis, y-axis, z-axis, pitch, roll	Spectral flatness, spectral centroid, spectral spread, spectral kurtosis	Smith et al., 2016
Physical description of the movement	FD	x-axis, y-axis, z-axis, *Amag*, pitch, roll, dynamic acceleration	Spectral entropy, fundamental frequency, maximum and second maximum power spectral density, wavelet features	Smith et al., 2016
	TFD			Lush et al., 2018 Riaboff et al., 2020 Hokkanen et al., 2011

Note: Equations of features listed in the table are provided in Supplementary Data B3.

Abbreviations: TD: time-domain; FD: frequency-domain. TFD: time-frequency domain

Appendix 4. In the *Development of a Behavioural Classification Model*: Key choices made when predicting ruminant behaviour from accelerometers in each surveyed paper (note - based on the 30 core articles*; see Supplementary Data A3 for the other 36 studies).

	Data splitting			Model training						Model Validation		
	Split criterion			Type of classifiers						Technique		
Reference	Random	Time-sequence	Individual	SML	SEM	UML	DL	SM	MT	Independent dataset	Cross-Validation	Bootstrap
Martiskainen et al., 2009	🟩			SVM						🟩		
Dutta et al., 2015	🟩			DT, LDA, NB, kNN	Bag, RS, Ada		ANFIS			🟩		
Gonzalez et al., 2015			🟩						🟩	🟩		
Vázquez Diosdado et al., 2015	Unspecified			DT, SVM		k-means		HMM		Unspecified		
Zobel et al., 2015			🟩						🟩	🟩		
Alvarenga et al., 2016	🟩			DT						🟩		
Smith et al., 2016			🟩	SVM, NB, kNN	RF			LR			🟩	
Abell et al., 2017	🟩			DT	RF					🟩		
Andriamandroso et al., 2017	🟩								🟩	🟩		
Arcidiacono et al., 2017	🟩								🟩	🟩		
Benaissa et al., 2017			🟩	kNN, NB, SVM							🟩	
Giovanetti et al., 2017			🟩	LDA								🟩
Barwick et al., 2018	🟩			QDA							🟩	
Benaissa et al., 2018			🟩	DT, SVM							🟩	
Decandia et al., 2018	🟩			LDA								🟩
Lush et al., 2018	🟩				RF					🟩		
Mansbridge et al., 2018	🟩			SVM, k-NN	RF, Ada						🟩	
Walton et al., 2018	🟩				RF					🟩		
Wang et al., 2018	🟩				Ada (MLP)					🟩		
Peng et al., 2019	🟩						LSTM,CNN			🟩		
Riaboff et al., 2019	🟩			DT							🟩	
Sakai et al., 2019	🟩			kNN, DT							🟩	
Tamura et al., 2019			🟩	DT						🟩		
Vázquez-Diosdado et al., 2019			🟩	kNN		k-means			🟩	🟩		
Wang et al., 2019	🟩				Ada (MLP)					🟩		
Weizheng et al., 2019			🟩	k-NN, SVM			PNN				🟩	
Fogarty et al., 2020			🟩	DT, SVM, LDA, QDA							🟩	
Hu et al., 2020	🟩			SVM, LDA	RF						🟩	
Peng et al., 2020	🟩						LSTM			🟩		
Riaboff et al., 2020		🟩		SVM	XGB, Ada, RF			HMM		🟩		

Note: * The 30 papers presented in Appendix 4 out of the 66 reviewed were selected based on their complete methodological practices; that is, there is no main missing elements in the description of the material and methods used and the results are reported in a complete and understandable way. Furthermore, we selected papers predicting at least two distinct behaviours in order to display the predictive performance obtained in a challenging framework. The details of the remaining 36 articles are reported in Supplementary Materials. The background of the cells is green when the technique has been applied, grey otherwise (n.b., for the abbreviations see list at the beginning of the review).

Appendix 5. *Model Performance* profiles showing overall performance (in terms of Accuracy, Kappa, F-score, Sensitivity, Specificity and Precision) and performance for each behaviour examined (in terms of Accuracy, F-score, Sensitivity, Specificity and Precision) (note - based on the 30 core articles*; see Supplementary Data A4 for the other 36 studies).

References[1]	Overall performance[2]						Performance per behaviour[3]			
	Acc. (%)	Kappa	Fscore	Sens. (%)	Spe. (%)	Prec. (%)	Acc. (%)	Sens. (%)	Spe. (%)	Prec. (%)
🟧 Martiskainen et al., 2009		0.69				78	F: 96; L: 84; Rum: 92; S: 87; W: 99	F: 75; L: 80; LD: 0; S: 80; SU: 71; Rum: 75; W: 79	F: 81; L: 83; LD: 0; S: 65; SU: 29; Rum: 86; W: 79	
🟩 Dutta et al., 2015	82-96		0.51-0.89	90-97	49-89		G: 86-98; R: 81-97; Rum: 82-97; Sea: 77-93; O: 84-92	G: 93-98; R: 85-98; Rum: 87-99; Sea: 89-97; O: 35-85	G: 53-93; R: 65-92; Rum: 60-90; Sea: 33-83; O: 65-92	
🟩 Gonzalez et al., 2015	90	0.87						F: 98; R: 86; Rum: 87; W: 25; O: 0	F: 99; R: 95; Rum: 95; W: 100; O: 98	
🟧 Vázquez Diosdado et al., 2015				68-88		69-88		F: 60-100; L: 55-93; LD: 67; S: 38-92; SU: 64		F: 86-93; L: 85-100; LD: 83; S: 29-79; SU: 70
🟧 Zobel et al., 2015				100	99					
🟩 Alvarenga et al., 2016	83-86	0.75-0.79					G: 85-93; L: 82-91; Run: 50-100; S: 56-86; W: 78-100	G: 84-92; L: 81-91; Run: 83-100; S: 74-84; W: 55-92	G: 95-97; L: 88-93; Run: 99-100; S: 88-96; W: 78-100	
🟩 Smith et al., 2016	94							G: 99; W: 69; Rum: 87; R: 82		G: 97; W: 77; Rum: 88; R: 88
🟩 Abell et al., 2017							L: 86-99; Moun: 68-80; S: 76-91; W: 77-87	L: 87-98; Moun: 70-81; S: 76-93; W: 68-81	L: 86-100; Moun: 68-80; S: 77-87; L: : 66-77	
🟩 Andriamandroso et al., 2017							G: 91; Rum: 97; O: 88	G: 91; Rum: 53; O: 88	G: 91; Rum: 99; O: 88	G: 94; Rum: 85; O: 79
🟧 Arcidiacono et al., 2017	94			93	96	95				
🟧 Benaissa et al., 2017	84-98							F: 72-99; L: 72-93; S: 43-96		F: 73-99; L: 81-99; S: 49-94
🟩 Giovanetti et al., 2017	93	0.89					G: 96; Rum: 95; R: 94	G: 96; Rum: 89; R: 93	G: 97; Rum: 97; R: 95	G: 95; Rum: 89; R: 94
🟩 Barwick et al., 2018							G: 86-97; L: 95-100; S: 87-97; W: 98-100	G: 85-92; L: 61-100; S: 69-98; W: 95-98	G: 88-98; L: 99-100; S: 89-95; W: 98-100	G: 91-96; L: 93-100; S: 54-96; W: 93-100
🟧 Benaissa et al., 2018	82-93							F: 85-96; Rum: 74-94; O: 78-89	F: 92-96; Rum: 87-96; O: 88-99	F: 83-92; Rum: 70-92; O: 87-98
🟩 Decandia et al., 2018	79-90	0.60-0.8					G: 88-94; Rum: 81-90; O: 85-95	G: : 88-95; Rum: 68-82; O: 57-88	G: 87-94; Rum: 89-95; O: 90-96	G: 87-95; Rum: 74-88; O: 49-78
🟩 Lush et al., 2018	94-96	0.93-0.95		80-87		92-94	Fo: 94-96; L: 100; Run: 72-83; S: 77-79; U: 33-69; W: 81-90	Fo: 97; L: 100; Run: 72-90; S: 76-81; U: 50-64; W: 85-91		Fo: 92-94; L: 100; Run: 90-97; S: 94-99; U: 81-90; W: 87-90
🟩 Mansbridge et al., 2018	67-92							G: 90-93; non-eating behaviour: 93-95; Rum: 86-87	G: 98; non-eating behaviour: 89-91; Rum: 97	G: 95-96; non-eating behaviour: 89; Rum: 89-92
🟩 Walton et al., 2018	88-95	0.80-0.91						L: 93-98; S: 82-93; W: 80-92	L: 92-97; S: 80-93; W: 81-93	L: 92-97; S: 81-95; W:81-93
🟧 Wang et al., 2019							F: 75; L: 92; LD: 99; S: 75; SU: 99; W: 97-99	F: 73; L: 93; LD: 82; S: 78; SU: 74; W: 92-94		F: 75; L: 82; LD: 86; S: 72; SU: 85; W: 86-89
🟧 Peng et al., 2019	80-89		0.79-0.89	80-89		79-89	F: 96-99; Head butt: 70-82; L: 74-90; Lick salt: 85-97; Moving: 70-85; RumL: 86-95; RumS: 90-93; Social lick: 46-80			
🟩 Riaboff et al., 2019	76-95		0.65-0.96				G: 95-100; L: 81-91; S: 75-95; W: 90-98	G: 91-100; L: 90-94; S: 60-81; W: 70-92	G: 100; L: 75-90; S: 91-95; W: 100	
🟩 Sakai et al., 2019	75-87							G: 74-95: L: 86-93; S: 45-67		G: 86-91; L: 87-95; S: 39-70
🟧 Tamura et al., 2019								E: 100; L: 100; Rum: 100	E: 100; L: 100; Rum: 100	
🟩 Vázquez-Diosdado et al., 2019	85		0.60	58	83	70	L: 84; S: 78; W: 93	L: 66; S: 90; W: 17	L: 91; S: 58; W: 99	L: 76; S: 79; W: 55
🟧 Wang et al., 2019	93			76	96	77	L: 93; LD: 94; S: 93; SU: 94; W: 93	L: 86; LD: 66; S: 87; SU: 69; W: 74-75	L: 96; LD: 97; S: 94; SU: 96; W: 95-97	L: 86; LD: 66; S: 83; SU: 68; W: 73-83
🟧 Weizheng et al., 2019	87-95	0.80-0.92						F: 86-96; Rum: 87-94; O: 84-94	F: 93-96; Rum: 90-98; O: 94-98	F: 88-93; Rum: 79-94, O: 90-97
🟩 Fogarty et al., 2020	53-98	0.3-0.8						active: 97; G: 90; inactive: 99; L: 56; prostrate: 100; S: 63; upright: 81; W: 66	active: 99; G: 98; inactive: 97; L: 93; prostrate: 81; S: 84; upright: 100; W: 45	active: 97; G: 97; inactive: 99; L: 70; prostrate: 79; S: 45; upright: 100; W: 25
🟩 Hu et al., 2020	72-100							G: 90-100; Rum: 0-100; S: 77-100; W: 0-100		G: 82-100; Rum: 0-100; S: 65-100; W: 0-100
🟧 Peng et al., 2020	77-80		0.77-0.80	77-80		77-81		G: 74-84; L: 55-63; RumL: 87-93; RumS: 87-91; S: 78-81		
🟩 Riaboff et al., 2020	95-98	0.91-0.96						G: 100; RL: 83-95; RS: 68-82; RumL: 95-99; RumS: 91-95; W: 70-84	G: 99; RL: 98-99; RS: 98-100; RumL: 99; RumS: 99; W: 100	

Note:

* The 30 papers presented in Appendix 5 out of the 66 reviewed were selected based on their complete methodological practices; that is, there is no main missing elements in the description of the material and methods used and the results are reported in a complete and understandable way. Furthermore, we selected papers predicting at least two distinct behaviours in order to display the predictive performance obtained in a challenging framework. The details of the remaining 36 articles are reported in Supplementary Materials.

[1] Behaviours observed and then predicted at pasture are written on green background; behaviours observed and then predicted in the barn are written on orange background.

[2] Overall performance reported in the studies. Note that when several models have been explored, minimum and maximum performance are provided in the Table (notation: min-max). Notations used: Acc:: accuracy; Sens.: sensitivity, Spe.: specificity; Prec: precision.

[3] Performance per behaviour reported in the studies. Note that when several models have been explored, minimum and maximum performance are provided in the Table (notation: min-max). Results on event- or disease-related behaviour are not reported in the Table.

Notations used for behaviours (alphabetical order): D: drinking; F: feeding; Fo: foraging; G: grazing; Gro: grooming; Int: interaction; L: lying; LD: lying down; Moun: mounting; O: other; R: resting; RL: resting while lying; RS: resting while standing; Rum: ruminating; RumL; ruminating while lying; RumS: ruminating while standing; Run: running; S: standing; Scr: scratching; Sea: searching; SU: standing up; U: urinating; W: walkin

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2021.106610.

References

Abell, K.M., Theurer, M.E., Larson, R.L., White, B.J., Hardin, D.K., Randle, R.F., 2017. Predicting bull behavior events in a multiple-sire pasture with video analysis, accelerometers, and classification algorithms. Comput. Electron. Agric. 136, 221–227. https://doi.org/10.1016/j.compag.2017.01.030.

Alvarenga, F.A.P., Borges, I., Palkovič, L., Rodina, J., Oddy, V.H., Dobos, R.C., 2016. Using a three-axis accelerometer to identify and classify sheep behaviour at pasture. Appl. Animal Behaviour Sci. 181, 91–99. https://doi.org/10.1016/j.applanim.2016.05.026.

Andriamandroso, A.L.H., Lebeau, F., Beckers, Y., Froidmont, E., Dufrasne, I., Heinesch, B., Dumortier, P., Blanchy, G., Blaise, Y., Bindelle, J., 2017. Development of an open-source algorithm based on inertial measurement units (IMU) of a smartphone to detect cattle grass intake and ruminating behaviors. Comput. Electron. Agric. 139, 126–137. https://doi.org/10.1016/j.compag.2017.05.020.

Arcidiacono, C., Porto, S.M.C., Mancino, M., Cascone, G., 2017. Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data. Comput. Electron. Agric. 134, 124–134. https://doi.org/10.1016/j.compag.2017.01.021.

Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E., 2017. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Min. Knowl. Disc. 31 (3), 606–660. https://doi.org/10.1007/s10618-016-0483-9.

Bailey, D.W., Trotter, M.G., Knight, C.W., Thomas, M.G., 2018. Use of GPS tracking collars and accelerometers for rangeland livestock production research1. Trans. Animal Sci. 2, 81–88. https://doi.org/10.1093/tas/txx006.

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.

Barwick, J., Lamb, D.W., Dobos, R., Welch, M., Schneider, D., Trotter, M., 2020. Identifying Sheep Activity from Tri-Axial Acceleration Signals Using a Moving Window Classification Model. Remote Sensing 12, 646. https://doi.org/10.3390/rs12040646.

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. https://doi.org/10.1016/j.compag.2018.01.007.

Benaissa, S., Tuyttens, F.A.M., Plets, D., Cattrysse, H., Martens, L., Vandaele, L., Joseph, W., Sonck, B., 2018. Classification of ingestive-related cow behaviours using RumiWatch halter and neck-mounted accelerometers. Appl. Animal Behaviour Sci. 211, 9–16. https://doi.org/10.1016/j.applanim.2018.12.003.

Benaissa, S., Tuyttens, F.A.M., Plets, D., de Pessemier, T., Trogh, J., Tanghe, E., Martens, L., Vandaele, L., Van Nuffel, A., Joseph, W., Sonck, B., 2017. On the use of on-cow accelerometers for the classification of behaviours in dairy barns. Res. Vet. Sci. 125, 425–433. https://doi.org/10.1016/j.rvsc.2017.10.005.

Benaissa, S., Tuyttens, F.A.M., Plets, D., Trogh, J., Martens, L., Vandaele, L., Joseph, W., Sonck, B., 2020. Calving and estrus detection in dairy cattle using a combination of indoor localization and accelerometer sensors. Comput. Electron. Agricult. 168 (105153) https://doi.org/10.1016/j.compag.2019.105153.

Bishop-Hurley, G., Henry, D., Smith, D., Dutta, R., Hills, J., Rawnsley, R., Hellicar, A., Timms, G., Morshed, A., Rahman, A., D'Este, C., Shu, Y., 2014. An investigation of cow feeding behavior using motion sensors. In: In: 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings. Presented at the 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), IEEE, Montevideo, Uruguay, pp. 1285–1290. https://doi.org/10.1109/I2MTC.2014.6860952.

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. J. Dairy Sci. 99 (9), 7458–7466. https://doi.org/10.3168/jds.2015-10843.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32.

Brugarolas, R., Loftin, R.T., Yang, P., Roberts, D.L., Sherman, B., Bozkurt, A., 2013. Behavior recognition based on machine learning algorithms for a wireless canine machine interface. In: In: 2013 IEEE International Conference on Body Sensor Networks. Presented at the 2013 IEEE International Conference on Body Sensor Networks (BSN), IEEE, Cambridge, MA, USA, pp. 1–5. https://doi.org/10.1109/BSN.2013.6575505.

Burges, C.J.C., 1998. A Tutorial on Support Vector Machines for Pattern Recognition. Data Min. Knowl. Disc. 2, 121–167.

Busch, P., Ewald, H., Stupmann, F., 2017. Determination of standing-time of dairy cows using 3D-accelerometer data from collars. In: in: 2017 Eleventh International Conference on Sensing Technology (ICST). Presented at the 2017 Eleventh International Conference on Sensing Technology (ICST), IEEE, Sydney, NSW, pp. 1–4. https://doi.org/10.1109/ICSensT.2017.8304492.

Carvalho, P.C.D.F., 2013. Harry Stobbs Memorial Lecture: Can grazing behavior support innovations in grassland management? Tropical Grasslands 1 (2), 137. https://doi.org/10.17138/TGFT(1)137-155.

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. Behav. Process. 181, 104262. https://doi.org/10.1016/j.beproc.2020.104262.

Decandia, M., Giovanetti, V., Molle, G., Acciaro, M., Mameli, M., Cabiddu, A., Cossu, R., Serra, M.G., Manca, C., Rassu, S.P.G., Dimauro, C., 2018. The effect of different time epoch settings on the classification of sheep behaviour using tri-axial accelerometry. Comput. Electron. Agric. 154, 112–119. https://doi.org/10.1016/j.compag.2018.09.002.

Delagarde, R., Caudal, J.-P., Peyraud, J.-L., 1999. Development of an automatic bitemeter for grazing cattle. Annales de Zootechnie 48 (5), 329–339. https://doi.org/10.1051/animres:19990501.

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. https://doi.org/10.1016/j.compag.2014.12.002.

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.

Fogarty, E.S., Swain, D.L., Cronin, G.M., Moraes, L.E., Bailey, D.W., Trotter, M., 2021. Developing a Simulated Online Model That Integrates GNSS, Accelerometer and Weather Data to Detect Parturition Events in Grazing Sheep: A Machine Learning Approach. Animals 11, 303. https://doi.org/10.3390/ani11020303.

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. https://doi.org/10.1016/j.compag.2019.105175.

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Statist. 29, 1189–1232.

Galindo, F., Broom, D.M., 2002. The Effects of Lameness on Social and Individual Behavior of Dairy Cows. J. Appl. Anim. Welfare Sci. 5 (3), 193–201. https://doi.org/10.1207/S15327604JAWS0503_03.

Gibb, M.J., 1996. Animal grazing/intake terminology and definitions. In: Pasture Ecology and Animal Intake. Presented at the Concerted Action AIR3-CT93-0947, Dublin.

Giovanetti, V., Decandia, M., Molle, G., Acciaro, M., Mameli, M., Cabiddu, A., Cossu, R., Serra, M.G., Manca, C., Rassu, S.P.G., Dimauro, C., 2017. Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer. Livestock Sci. 196, 42–48. https://doi.org/10.1016/j.livsci.2016.12.011.

González, L.A., Bishop-Hurley, G.J., Handcock, R.N., Crossman, C., 2015. Behavioral classification of data from collars containing motion sensors in grazing cattle. Comput. Electron. Agric. 110, 91–102. https://doi.org/10.1016/j.compag.2014.10.018.

Guo, L., Welch, M., Dobos, R., Kwan, P., Wang, W., 2018. Comparison of grazing behaviour of sheep on pasture with different sward surface heights using an inertial measurement unit sensor. Comput. Electron. Agric. 150, 394–401. https://doi.org/10.1016/j.compag.2018.05.004.

Hamalainen, W., Jarvinen, M., Martiskainen, P., Mononen, J., 2011. Jerk-based feature extraction for robust activity recognition from acceleration data. In: 2011 11th International Conference on Intelligent Systems Design and Applications. Presented at the 2011 11th International Conference on Intelligent Systems Design and Applications (ISDA), IEEE, Cordoba, Spain, pp. 831–836. https://doi.org/10.1109/ISDA.2011.6121760.

Hamilton, A., Davison, C., Tachtatzis, C., Andonovic, I., Michie, C., Ferguson, H., Somerville, L., Jonsson, N., 2019. Identification of the Rumination in Cattle Using Support Vector Machines with Motion-Sensitive Bolus Sensors. Sensors 19, 1165. https://doi.org/10.3390/s19051165.

Hänninen, S., 2010. Finnish Society for Applied Ethology and University of Eastern Finland. In: In: Proceedings of the 22nd Nordic Symposium of the International Society. Presented at the International Society for Applied Ethology, Finnish Society for Applied Ethology, Siilinjärvi, Finland, p. 3.

Heinicke, J., Ibscher, S., Belik, V., Amon, T., 2019. Cow individual activity response to the accumulation of heat load duration. J. Therm. Biol 82, 23–32. https://doi.org/10.1016/j.jtherbio.2019.03.011.

Hendriks, S.J., Phyn, C.V.C., Huzzey, J.M., Mueller, K.R., Turner, S.-A., Donaghy, D.J., Roche, J.R., 2020. Graduate Student Literature Review: Evaluating the appropriate use of wearable accelerometers in research to monitor lying behaviors of dairy cows. J. Dairy Sci. 103 (12), 12140–12157. https://doi.org/10.3168/jds.2019-17887.

Hokkanen, A.-H., Hänninen, L., Tiusanen, J., Pastell, M., 2011. Predicting sleep and lying time of calves with a support vector machine classifier using accelerometer data. Appl. Animal Behav. Sci. 134 (1-2), 10–15. https://doi.org/10.1016/j.applanim.2011.06.016.

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. https://doi.org/10.1016/j.compag.2020.105857.

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.-A., 2019. Deep learning for time series classification: a review. Data Min. Knowl. Disc. 33 (4), 917–963. https://doi.org/10.1007/s10618-019-00619-1.

Jensen, M.B., 2012. Behaviour around the time of calving in dairy cows. Appl. Animal Behav. Sci. 139 (3-4), 195–202. https://doi.org/10.1016/j.applanim.2012.04.002.

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 Interactive, Mobile, Wearable Ubiquitous Technologies 2 (1), 1–27. https://doi.org/10.1145/3191747.

Keeling, L.J., 2019. Indicators of Good Welfare. In: Encyclopedia of Animal Behavior. Elsevier, pp. 134–140. https://doi.org/10.1016/B978-0-12-809633-8.90715-5.

Keogh, E., Chu, S., Hart, D., Pazzani, M., 2001. An online algorithm for segmenting time series. In: Proceedings 2001 IEEE International Conference on Data Mining. Presented at the 2001 IEEE International Conference on Data Mining, IEEE Comput. Soc, San Jose, CA, USA, pp. 289–296. https://doi.org/10.1109/ICDM.2001.989531.

Khanh, P.C.P., Dinh Chinh, N., Cham, T.T., Vui, P.T., Tan, T.D., 2016. Classification of cow behavior using 3-DOF accelerometer and decision tree algorithm. In: In: 2016 International Conference on Biomedical Engineering (BME-HUST). Presented at the 2016 International Conference on Biomedical Engineering (BME-HUST), IEEE, Hanoi, Vietnam, pp. 45–50. https://doi.org/10.1109/BME-HUST.2016.7782100.

Kilgour, R.J., 2012. In pursuit of "normal": A review of the behaviour of cattle at pasture. Appl. Animal Behav. Sci. 138 (1-2), 1–11. https://doi.org/10.1016/j.applanim.2011.12.002.

Kleanthous, N., Hussain, A., Mason, A., Sneddon, J., 2019. Data Science Approaches for the Analysis of Animal Behaviours. In: Huang, D.-S., Huang, Z.-K., Hussain, A. (Eds.), Intelligent Computing Methodologies, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 411–422. https://doi.org/10.1007/978-3-030-26766-7_38.

Kleanthous, N., Hussain, A., Mason, A., Sneddon, J., Shaw, A., Fergus, P., Chalmers, C., Al-Jumeily, D., 2018. Machine Learning Techniques for Classification of Livestock Behavior. In: Cheng, L., Leung, A.C.S., Ozawa, S. (Eds.), Neural Information Processing, Lecture Notes in Computer Science. pp. 304–315. https://doi.org/10.1007/978-3-030-04212-7_26.

Konka, J., Michie, C., Andonovic, I., 2014. Automatic Classification of Eating and Ruminating in Cattle Using a Collar Mounted Accelerometer. IEE Sensors 5.

Kour, H., Patison, K.P., Corbet, N.J., Swain, D.L., 2018. Validation of accelerometer use to measure suckling behaviour in Northern Australian beef calves. Appl. Animal Behav. Sci. 202, 1–6. https://doi.org/10.1016/j.applanim.2018.01.012.

Kuźnicka, E., Gburzyński, P., 2017. Automatic detection of suckling events in lamb through accelerometer data classification. Comput. Electron. Agric. 138, 137–147. https://doi.org/10.1016/j.compag.2017.04.009.

le Roux, S.P., Marias, J., Wolhuter, R., Niesler, T., 2017. Animal-borne behaviour classification for sheep (Dohne Merino) and Rhinoceros (Ceratotherium simum and Diceros bicornis). Anim Biotelemetry 5, 25. https://doi.org/10.1186/s40317-017-0140-0.

le Roux, S.P., Wolhuter, R., Niesler, T., 2019. Energy-Aware Feature and Model Selection for Onboard Behavior Classification in Low-Power Animal Borne Sensor Applications. IEEE Sensors J. 19 (7), 2722–2734. https://doi.org/10.1109/JSEN.2018.2886890.

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. https://doi.org/10.1016/j.compag.2018.04.018.

Mansbridge, N., Mitsch, J., Bollard, N., Ellis, K., Miguel-Pacheco, 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.

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. Animal Behav. Sci. 119 (1-2), 32–38.

Mattachini, G., Riva, E., Perazzolo, F., Naldi, E., Provolo, G., 2016. Monitoring feeding behaviour of dairy cows using accelerometers. J. Agric. Eng. 47 (1), 54. https://doi.org/10.4081/jae.2016.498.

Medria Solutions [WWW Document], 2020. Farmlife Bouquet: Enter in the new era of monitoring! URL https://www.medria.fr/en/solutions/herd-monitoring.html (accessed 11.9.20).

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. Animal Behav. Sci. 119 (3-4), 158–170. https://doi.org/10.1016/j.applanim.2009.04.008.

Nadimi, E.S., Jørgensen, R.N., Blanes-Vidal, V., Christensen, S., 2012. Monitoring and classifying animal behavior using ZigBee-based mobile ad hoc wireless sensor networks and artificial neural networks. Comput. Electron. Agric. 82, 44–54. https://doi.org/10.1016/j.compag.2011.12.008.

Nadimi, E.S., Søgaard, H.T., Bak, T., 2008. ZigBee-based wireless sensor networks for classifying the behaviour of a herd of animals using classification trees. Biosyst. Eng. 100 (2), 167–176. https://doi.org/10.1016/j.biosystemseng.2008.03.003.

O'Leary, N.W., Byrne, D.T., Garcia, P., Werner, J., Cabedoche, M., Shalloo, L., 2020. Grazing Cow Behavior and Lameness (preprint). Engineering. https://doi.org/10.20944/preprints202002.0089.v1.

Pavlovic, D., Davison, C., Hamilton, A., Marko, O., Atkinson, R., Michie, C., Crnojević, V., Andonovic, I., Bellekens, X., Tachtatzis, C., 2021. Classification of Cattle Behaviours Using Neck-Mounted Accelerometer-Equipped Collars and Convolutional Neural Networks. Sensors 21, 4050. https://doi.org/10.3390/s21124050.

Peng, Y., Kondo, N., Fujiura, T., Suzuki, T., Ouma, S., Wulandari, Yoshioka, H., Itoyama, E., 2020. Dam behavior patterns in Japanese black beef cattle prior to calving: Automated detection using LSTM-RNN. Comput. Electron. Agric. 169, 105178. https://doi.org/10.1016/j.compag.2019.105178.

Peng, Y., Kondo, N., Fujiura, T., Suzuki, T., Wulandari, Yoshioka, H., Itoyama, E., 2019. Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units. Comput. Electron. Agric. 157, 247–253.

Qasem, L., Cardew, A., Wilson, A., Griffiths, I., Halsey, L.G., Shepard, E.L.C., Gleiss, A.C., Wilson, R., Ropert-Coudert, Y., 2012. Tri-Axial Dynamic Acceleration as a Proxy for Animal Energy Expenditure; Should We Be Summing Values or Calculating the Vector? PLoS ONE 7 (2), e31187. https://doi.org/10.1371/journal.pone.0031187.

Radeski, M., Ilieski, V., 2017. Gait and posture discrimination in sheep using a tri-axial accelerometer. Animal 11 (7), 1249–1257. https://doi.org/10.1017/S175173111600255X.

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. Inform. Process. Agric. 5 (1), 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. Comput. Electron. Agric. 165, 104961. https://doi.org/10.1016/j.compag.2019.104961.

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. https://doi.org/10.1016/j.compag.2019.105179.

Riaboff, L., Relun, A., Petiot, C.-E., Feuilloy, M., Couvreur, S., Madouasse, A., 2021. Identification of discriminating behavioural and movement variables in lameness scores of dairy cows at pasture from accelerometer and GPS sensors using a Partial Least Squares Discriminant Analysis. Preventive Veterinary Med. 193, 105383. https://doi.org/10.1016/j.prevetmed.2021.105383.

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. Comput. Electron. Agric. 67 (1-2), 80–84.

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. https://doi.org/10.1016/j.jocs.2020.101076.

Rutten, C.J., Steeneveld, W., Vernooij, J.C.M., Huijps, K., Nielen, M., Hogeveen, H., 2016. A prognostic model to predict the success of artificial insemination in dairy cows based on readily available data. J. Dairy Sci. 99 (8), 6764–6779. https://doi.org/10.3168/jds.2016-10935.

Rutter, S.M., Champion, R.A., Penning, P.D., 1997. An automatic system to record foraging behaviour in free-ranging ruminants. Appl. Animal Behav. Sci. 54 (2-3), 185–195. https://doi.org/10.1016/S0168-1591(96)01191-4.

Ruuska, S., Kajava, S., Mughal, M., Zehner, N., Mononen, J., 2016. Validation of a pressure sensor-based system for measuring eating, rumination and drinking behaviour of dairy cattle. Appl. Animal Behav. Sci. 174, 19–23. https://doi.org/10.1016/j.applanim.2015.11.005.

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. Comput. Electron. Agric. 166, 105027. https://doi.org/10.1016/j.compag.2019.105027.

Shalloo, L., Byrne, T., Leso, L., Ruelle, E., Starsmore, K., Geoghegan, A., Werner, J., O'Leary, N., 2021. A review of precision technologies in pasture-based dairying systems. Irish J. Agric. Food Res. https://doi.org/10.15212/ijafr-2020-0119.

Silberberg, M., Meunier, B., Veissier, I., Mialon, M.-M., 2017. Continuous monitoring of cow activity to detect sub-acute ruminal acidosis (SARA). Presented at the EC-PLF, Nantes.

Smith, D., Dutta, R., Hellicar, A., Bishop-Hurley, G., Rawnsley, R., Henry, D., Hills, J., Timms, G., 2015. Bag of Class Posteriors, a new multivariate time series classifier applied to animal behaviour identification. Expert Syst. Appl. 42 (7), 3774–3784. https://doi.org/10.1016/j.eswa.2014.11.033.

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. https://doi.org/10.1016/j.compag.2016.10.006.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manage. 45 (4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.

Song-Mi Lee, Sang Min Yoon, Heeryon Cho, 2017. Human activity recognition from accelerometer data using Convolutional Neural Network. In: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp). Presented at the 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, Jeju Island, South Korea, pp. 131–134. https://doi.org/10.1109/BIGCOMP.2017.7881728.

Tamura, T., Okubo, Y., Deguchi, Y., Koshikawa, S., Takahashi, M., Chida, Y., Okada, K., 2019. Dairy cattle behavior classifications based on decision tree learning using 3-axis neck-mounted accelerometers. Anim. Sci. J. 90 (4), 589–596. https://doi.org/10.1111/asj.13184.

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. Biotelem. 3 (1) https://doi.org/10.1186/s40317-015-0045-8.

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.

Vimalajeewa, D., Kulatunga, C., Berry, D., Balasubramaniam, S., 2021. A Service-based Joint Model Used for Distributed Learning: Application for Smart Agriculture. IEEE Trans. Emerg. Topics Comput. 1–1 https://doi.org/10.1109/TETC.2020.3048671.

Walker, J.S., Jones, M.W., Laramee, R.S., Holton, M.D., Shepard, E.LC., Williams, H.J., Scantlebury, D.M., Marks, N.J., Magowan, E.A., Maguire, I.E., Bidder, O.R., Di Virgilio, A., Wilson, R.P., 2015. Prying into the intimate secrets of animal lives; software beyond hardware for comprehensive annotation in 'Daily Diary' tags. Mov. Ecol. 3 (1) https://doi.org/10.1186/s40462-015-0056-3.

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. R. Soc. Open Sci. 5 (2), 171442. https://doi.org/10.1098/rsos.171442.

Wang, J., He, Z., Ji, J., Zhao, K., Zhang, H., 2019. IoT-based measurement system for classifying cow behavior from tri-axial accelerometer. Cienc. Rural 49, e20180627. https://doi.org/10.1590/0103-8478cr20180627.

Wang, J., He, Z., Zheng, G., Gao, S., Zhao, K., Loor, J.J., 2018. Development and validation of an ensemble classifier for real-time recognition of cow behavior patterns from accelerometer data and location data. PLoS ONE 13 (9), e0203546. https://doi.org/10.1371/journal.pone.0203546.

Watanabe, N., Sakanoue, S., Kawamura, K., Kozakai, T., 2008. Development of an automatic classification system for eating, ruminating and resting behavior of cattle using an accelerometer. Grassland Sci. 54 (4), 231–237.

Watanabe, S., Izawa, M., Kato, A., Ropert-Coudert, Y., Naito, Y., 2005. A new technique for monitoring the detailed behaviour of terrestrial animals: A case study with the domestic cat. Appl. Animal Behav. Sci. 94 (1-2), 117–131. https://doi.org/10.1016/j.applanim.2005.01.010.

Weizheng, S., Fei, C., Yu, Z., Xiaoli, W., Qiang, F., Yonggen, Z., 2019. Automatic recognition of ingestive-related behaviors of dairy cows based on triaxial acceleration. S2214317319301921 Inform. Process. Agric.. https://doi.org/10.1016/j.inpa.2019.10.004.

Williams, L.R., Moore, S.T., Bishop-Hurley, G.J., Swain, D.L., 2020. A sensor-based solution to monitor grazing cattle drinking behaviour and water intake. Comput. Electron. Agric. 168, 105141. https://doi.org/10.1016/j.compag.2019.105141.

Yunta, C., Guasch, I., Bach, A., 2012. Short communication: Lying behavior of lactating dairy cows is influenced by lameness especially around feeding time. J. Dairy Sci. 95 (11), 6546–6549. https://doi.org/10.3168/jds.2012-5670.

Zambelis, A., 2019. Technical note: Validation of an ear-tag accelerometer to identify feeding and activity behaviors of tiestall-housed dairy cattle. J. Dairy Sci 102 (5), 4536–4540. https://doi.org/10.3168/jds.2018-15766.

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. J. Dairy Sci. 98 (2), 1082–1089. https://doi.org/10.3168/jds.2014-8635.

Williams, L.R., Bishop-Hurley, G.J., Anderson, A.E., Swain, D.L., 2017. Application of accelerometers to record drinking behaviour of beef cattle. Anim. Prod. Sci. 59 (1), 122. https://doi.org/10.1071/AN17052.