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

L. Riaboff, Sylvain Poggi, Aurélien Madouasse, S. Couvreur, S. Aubin, Nicolas Bédère, E. Goumand, Alain Chauvin, Guy Plantier

Computers and Electronics in Agriculture 169 (2020) 105179

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

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

L. Riaboff[a,d,*], S. Poggi[b], A. Madouasse[e], S. Couvreur[c], S. Aubin[a], N. Bédère[c,f], E. Goumand[d], A. Chauvin[e], G. Plantier[a]

[a] ESEOTech – LAUM, 49000 Angers, France
[b] INRAE, Agrocampus Ouest, Université de Rennes, IGEPP, F-35650 Le Rheu, France
[c] URSE, Ecole Supérieure d'Agricultures, Univ. Bretagne Loire, 49000 Angers, France
[d] Terrena Innovation, 44150 Ancenis, France
[e] BIOEPAR, INRAE, Oniris, 44307 Nantes, France
[f] INRAE, AGROCAMPUS OUEST, PEGASE, F-35590 Saint-Gilles, France

1. Introduction

Monitoring livestock behaviours is a useful way to detect breeding events such as oestrus (Kamphuis et al., 2012) and health events such as lameness (Chapinal et al., 2009). In pasture-based systems, monitoring the behaviour of grazing ruminants is also important to optimise animal intake and performance (Carvalho, 2013). However, these observations are time consuming, labour-intensive (Penning, 1983) and sometimes access to the animal is not easy, like in mountain environment (Schlecht et al., 2009). For these reasons, applications that involve sensors to automatically monitor the behaviour of livestock have grown rapidly over the last years (Rushen et al., 2012). Among available technologies, promising results have been obtained with accelerometers, already widely used for human activity recognition (Mathie et al., 2004). Accelerometers have been effective in discriminating behaviours of ruminants like grazing, ruminating or lying (Benaissa et al., 2018; Dutta et al., 2015; Robert et al., 2009), depending on where the sensor is attached to the animal. Accelerometers have also been used to predict specific events such as oestrus (Shahriar et al., 2016) or urination (Lush et al., 2018) of livestock for instance.

For their effective application in farming, sensors should provide information about a wide range of behaviours (Rutten et al., 2013). Animal's posture like standing or lying, and feeding behaviours like grazing or ruminating, are both crucial to detect health events and assist in pasture management. For example, Yunta et al. (2012) showed that lying bout durations are longer for moderately lame cows. Norring et al. (2014) highlighted that lameness is associated with changes in feeding behaviour, thereby justifying that both posture and feeding behaviour should be taken into account to detect lameness. Concerning pasture management, most of the studies focused on grazing, ruminating, resting and active behaviours (Andriamandroso, 2017; Giovanetti et al., 2017; Gonzalez et al., 2015), as these are the main behaviours affecting livestock performance. Collecting lying and standing time and identifying the main areas where animals carried out these behaviours is also a relevant information for pasture management and welfare monitoring (Martiskainen et al., 2009; Riaboff et al., 2018). In the latter studies, a satisfactory discrimination of two or three behaviours was obtained (Benaissa et al., 2018; Giovanetti et al., 2017) but the performance dramatically decreased when posture and feeding behaviours were investigated (Alvarenga et al., 2016; Martiskainen et al., 2009). Thus, the limited set of discriminated behaviours is currently a significant barrier to the practical use of sensors in farming (Dutta et al., 2015).

Behaviour prediction generally consists of the three-following steps (i) signal processing, (ii) feature extraction, and (iii) supervised learning for classification. Few studies have considered step (i) to improve the prediction although its importance has been demonstrated in human activity recognition (Bersch et al., 2014). Most authors have focused on step (ii) in selecting the best set of features (Smith et al., 2016). Concerning step (iii), different classification algorithms have already been compared to identify the most suitable classifier (Smith et al., 2016; Vázquez Diosdado et al., 2015). Regarding step (iii), Machine Learning methods have been widely considered, including Support Vector Machine (SVM) (Martiskainen et al., 2009) and Random Forest (RF) (Lush et al., 2018) but less complex methods have sometimes been applied, such as Discriminant Analysis (Barwick et al., 2018) or Decision Tree (Robert et al., 2009). Stochastic models such as Hidden Markov Models (HMM) have sometimes been used to incorporate

* Corresponding author at: ESEOTech – LAUM, 49000 Angers, France.
E-mail addresses: lucile.riaboff@eseo.fr (L. Riaboff), sylvain.poggi@inra.fr (S. Poggi), aurelien.madouasse@oniris-nantes.fr (A. Madouasse), s.couvreur@groupe-esa.com (S. Couvreur), sebastien.aubin@eseo.fr (S. Aubin), Nicolas.Bedere@inra.fr (N. Bédère), Etienne.Goumand@idele.fr (E. Goumand), alain.chauvin@oniris-nantes.fr (A. Chauvin), guy.plantier@eseo.fr (G. Plantier).

https://doi.org/10.1016/j.compag.2019.105179
Received 21 October 2019; Received in revised form 20 December 2019; Accepted 24 December 2019
Available online 17 January 2020

temporal information within the behaviour sequences but it seems less effective than common supervised algorithm (Vázquez Diosdado et al., 2015). Yang (2009) proposed a strategy which consists in first predicting human activities with a decision tree and then smoothing the outputs with an HMM-based Viterbi algorithm. As this 2 steps strategy allowed improving the accuracy of the prediction of human behaviour, it should be interesting to evaluate it for the prediction of livestock behaviour. Finally, regardless of the prediction methods used, the predictions obtained with the model are evaluated usually using independent windows of behaviours (Robert et al., 2009).

In this study, our aim was to develop a methodological framework for predicting 6 behaviours expressed by dairy cows at pasture and relevant for on-farm applications, using a three-dimensional accelerometer fixed on a neck-collar. The 6 behaviours of interest were grazing, walking, ruminating while standing, ruminating while lying, resting while standing and resting while lying. Our approach combined a raw signal pre-processing step investigated in a previous study (Riaboff et al., 2019) and the identification of the best machine learning algorithm among 4 under consideration (Random Forest (RF), Support Vector Machine (SVM), Adaboost (ADA) and eXtreme Gradient Boosting (XGB)). A temporal smoothing with an HMM-based Viterbi algorithm was finally applied to reassess the predictions of the behaviours obtained with the Machine Learning algorithms based on the temporal structure within the sequence of behaviours.

2. Material and methods

An overview of the applied methodological framework is provided in Fig. 1. Each step is detailed in the following sections.

2.1. Data collection

Data collection was carried out on four commercial dairy farms with 55–71 Holstein cows, located in the region Pays-de-la-Loire (France), during the summer of 2017, and the spring and summer of 2018. On each farm, cows that were equipped with a collar were selected based on stage of lactation (days in milk) and number of calvings (parity), as these factors are known to affect the time spent in different behaviours. Therefore, we selected cows between 60 and 300 days in milk, and representative of the herd in terms of parity. A total of 86 cows were equipped with a sensor through the study. Further details on farms and animals are presented in Table 1.

An RF-Track datalogger (RF-Track, Rennes, France) with a LSM9DS1 three-axis accelerometer (STMicroelectronics, Geneva, Switzerland) ± 2 g was used. Data were collected at 59.5 Hz. The accelerometer was powered with two 3.7 V lithium batteries (2.6 Ah). Data were stored on a SD card and downloaded after the experiment. The dataloggers were 98.2 × 51.60 × 36.0 mm in size and weighed 250 g. The three-axis accelerometer was fixed on a collar and positioned on the right side of the neck as recommended by Smith et al. (2016). Collars were tightly adjusted and a 500 g counterweight was added to prevent them from turning around (Fig. 2) (Robert et al., 2009). The x-axis detected the up-down direction, the y-axis detected the backward-forward direction and the z-axis detected the left–right direction.

Each selected cow was equipped with the collar during one day of grazing on the first and second farm, 2, 4 or 6 days of grazing on the third farm depending on the cow, and 3 days of grazing on the last farm. Two trained observers recorded behaviours of the cows at pasture at each observation session. While one observer tracked one cow during 15 min, the other observer focused on another cow. All the cows equipped in the 4 farms were observed at least once by each of the two observers. Therefore, we collected a minimum of 30 min of observation per cow. A total of 57 h 21 min and 57 s of observation were performed. The main behaviours recorded are “grazing”, “walking”, “ruminating – lying”, “ruminating – standing”, “resting – lying“ and “resting – standing”. The behaviours “grooming”, “urinating”, “interaction with other cows”, “running”, “raising”, “lying down” and “grazing while lying” were also observed. As they were each poorly represented, they were combined as “Other”. The behaviours are defined in the ethogram detailed in Table 2. It should be noted that the concordance between the two observers was checked (percentage of agreement > 80%) to ensure that observations could be considered independent of the observer in the rest of the study.

The duration of observation and the number of cows for each behaviour are reported in Table 3.

2.2. Raw accelerometer signal processing

The pre-processing step aimed providing suitable datasets for the subsequent application of the classification algorithms. The pre-processing step was performed in Matlab R2018a. The total duration of observations led to 229 continuous signal sequences of 15-minutes. Due to a malfunction of 4 of the 30 sensors available, we collected only 199 continuous sequences. We first combined the 199 continuous sequences of raw accelerometer signal (thereafter referred to as **observation sequences**) with the corresponding field observations. It should be noted that the accelerometer signal and observations were time synchronized to ensure that direct observations were associated to the correct sequences of accelerometer signal.

These 199 observation sequences were then split into segments (time-window) of the same duration (size of the time-window). When more than one behaviour was associated to a time-window, this time-window was removed from its sequence. In this way, the features extracted from the signal windows were representative of each behaviour specifically, which is the most suitable to train the models. Consequently, the 199 initial continuous sequences of observation were split into shorter continuous sequences. Furthermore, time-windows associated to the “Other” class were discarded from the dataset. Indeed, it combined several behaviours so it is not possible to extract representative features for this class in a supervised classification context. Therefore, we focused on the 6 main behaviours. As signal segmentation may drastically alter the results of the prediction (Bersch et al., 2014; Robert et al., 2009; Smith et al., 2016) several configurations identified as affecting the quality of the prediction in a previous study (Riaboff et al., 2019) were evaluated. In particular, window sizes of 5 s and 10 s were evaluated as they are commonly used in similar studies (Robert et al., 2009), while 50% and 90% overlap were compared as recommended by Bersch et al. (2014). This resulted in testing four different pre-processing configurations. The best results were obtained with the configuration (10 s; 90%), in accordance with Riaboff et al. (2019). The number of observation sequences and time-windows obtained for the configuration (10 s; 90%) for every behaviour are presented in the Table 4. The study was carried out thereafter with this configuration exclusively.

A set of features was then extracted from each 10 s-window, both in the time and frequency domains. Features were computed from the raw signal () of X-axis (a_x), Y-axis (a_y) and Z-axis (a_z). The signal magnitude axis (a_{mag}), considered as orientation independent (Fida et al., 2015) was also computed (Appendix A). Each axis (a_x, a_y, a_z, a_{mag}) was high-pass filtered to get the dynamic component of the acceleration due to the body of the cow and noted $a_{dynamic\_axis}$. A 6-th-order high-pass digital Butterworth filter with a cutoff frequency of 0.3 Hz (Smith et al., 2016) was applied to each axis of the raw data. The Overall Body Dynamic Acceleration (*OBDA*) and the Vector of the Dynamic Body Acceleration (*VeDBA*) were computed from the dynamic component. Each axis (a_x, a_y, a_z, a_{mag}) was also low-pass filtered to get the static component of the acceleration due to the gravity and noted a_{static_axis}. The pitch and roll angle were calculated on a_{static}. The sixty-seven calculated features are listed in Table 5. The formulas used to compute each feature are provided in the Appendix A.

It should be mentioned that the features perfectly correlated were

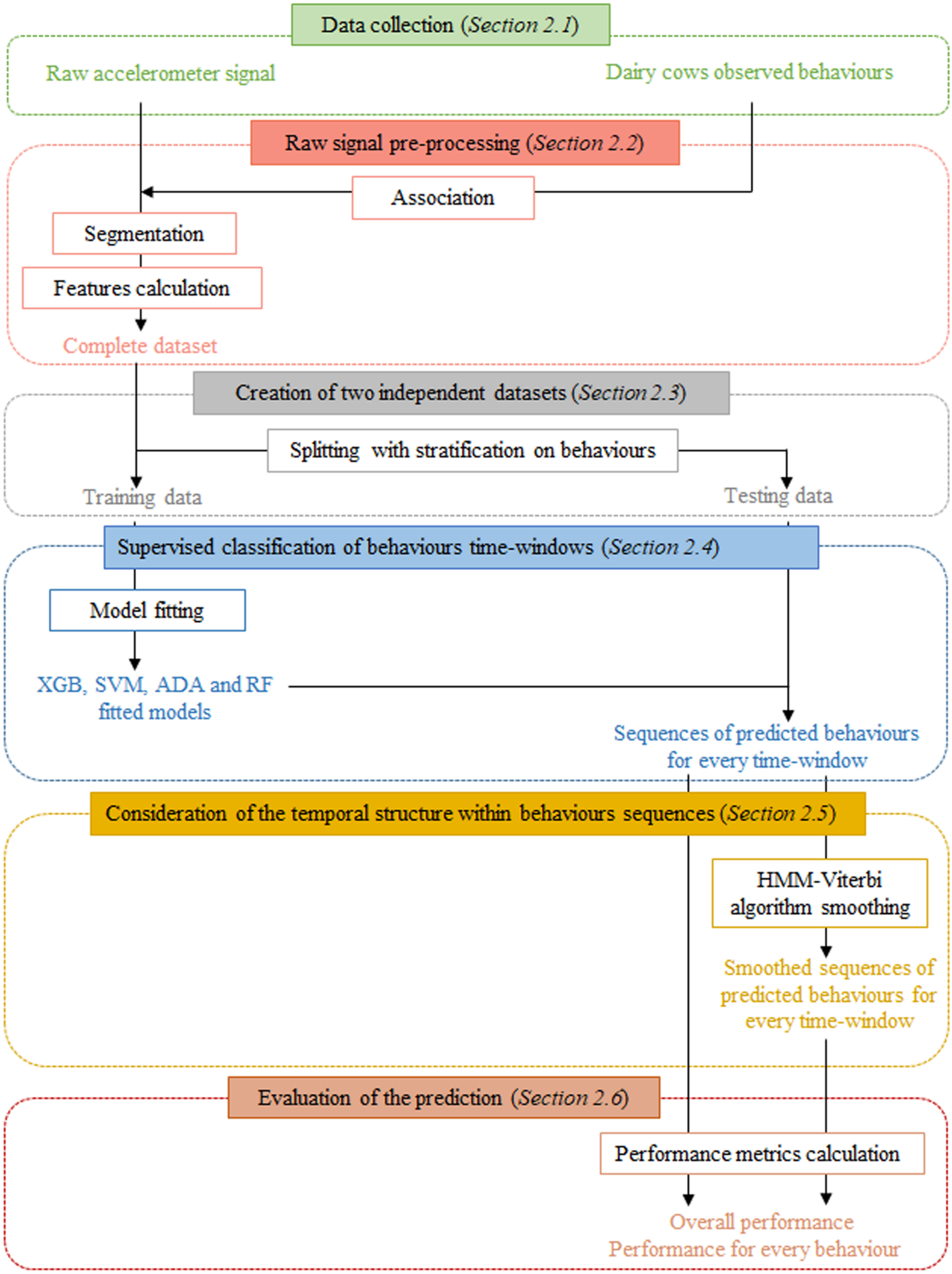

Fig. 1. Overview of the methodological framework used to predict the 6 main dairy cow behaviours at grazing.

excluded. Thus, the RMS on a_x, the mean and median on a_y, the mean on a_z, the RMS on a_{mag}, the pitch and VeDBA were not considered in the rest of the study. The features were finally normalized according to the equation presented in the Appendix A. Therefore, the complete dataset consisted of 61 features associated to each 10 s-window.

2.3. Creation of two independent datasets

As we are in a supervised classification context, the complete dataset was split into two independent datasets. Two-thirds of the 1794 observation sequences (Table 4) were randomly chosen to train the models and the remaining observation sequences were used to evaluate them. Random sampling of observations was stratified by behaviour in order to make sure that each behaviour was equally frequent in both the training and testing datasets. Therefore, the training dataset was composed of the 10 s-windows belonging to the selected two-thirds observation sequences. The testing dataset was composed of the remaining 10 s-windows.

2.4. Supervised classification of behaviour time-windows

The aim of this step was to classify the 10 s-windows from each observation sequence optimally. For this purpose, four supervised Machine Learning algorithms were investigated. Details about the

Table 1
Farms and dairy cows under study.

Farm	GAEC Beloin	GAEC Perrière	GAEC Haute-Roue	GAEC Harmonie
Location	Lion d'Angers, France (N 47°37′43″, O 0°42′42″)	Le Cellier, France (N 47°19′13″, O 1°20′43″)	La Pommeraye, France (N 47°21′21″, O 0°51′33″)	La Rouxière, France (N 47°26′37″, O 1°03′59″)
Period of experiment	July 2017	April 2018	May 2018	July 2018
Number of cows equipped/Herd size	10/71	12/55	40/70	24/61
Paddocks areas	0.3 ha	1.2 ha	From 0.6 to 2 ha	0.3 ha
Pasture access	Day and night	6 h a day	Day and night	6 h a day
Average grass height* (beginning of the observation)	12 cm	12 cm	15 cm	6 cm
Grasslands	Temporary grassland: mixture of *Lolium perenne L. Trifolium repens L.*	Temporary grassland: mixture of *Lolium perenne L.* and *Trifolium repens L.*	Temporary grassland: mixture of *Lolium perenne L.*, *Trifolium repens L.* and *Bromus hordeacus* Permanent grassland: mainly containing *Lolium perenne L*, *Trifolium repens L.*, *Dactylis glomerata* and *Holcus lanatus*	Temporary grassland: mixture of *Lolium perenne L.* and *Trifolium repens L.*
Supplement	Maize silage (10 kg/day), concentrate (2 kg/day) and hay (5 kg/day) after each milking	Maize silage (10 kg/day), concentrate (2 kg/day) and hay (5 kg/day) after each milking	Alfalfa hay (5 kg/day) once a day, a ration of Triticale – Peas – Field Beans (1.83 kg/day) after each AMS frequentation	Maize silage (10 kg/day), concentrate (2 kg/day) and hay (5 kg/day) after each milking
Milk yield (kg/cow/year)	10,000	9000	9000	8000

* Grass height measurements were carried out with the GrassHopper® plate meter (McSweeney et al., 2015) in the pastures before the beginning of the experiment.

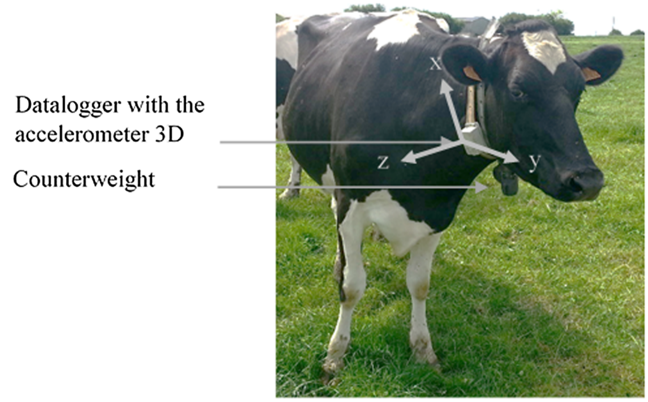

Fig. 2. RF-Track 3D accelerometer and GPS sensor fixed on the collar on the cow. The three coordinate axes of the accelerometer sensor are displayed.

considered algorithms and the methodology used to fit the models and predict behaviours are provided in the following sections.

2.4.1. Description of the considered Machine Learning algorithms

- Extreme Boosting Algorithm (XGB)

XGB is an ensemble classifier derived from the gradient boosting decision tree (Friedman, 2001). XGB combines weak base classifiers into a strong classifier. At each iteration of the training process, the residual of a base classifier is used in the next classifier for optimizing the objective function. In addition, XGB introduces a regularization term to control the complexity of the model and thus prevents overfitting (Shi et al., 2019). Hyperparameters to tune are listed below:

- *nrounds*: number of boosting iterations.
- *max_depth*: maximum depth of a tree. High values will tend to complicate the model with likely overfitting.
- *eta*: controls the learning rate. After each boosting step, it scales the contribution of each tree by a factor range between 0 and 1. In this manner, it prevents overfitting by making the boosting process more conservative.
- *gamma*: minimum loss reduction required to make a further partition on a leaf node of the tree. The smaller, the less conservative the algorithm will be.
- *colsample_bytree*: subsample ratio of columns when constructing each tree.
- *min_child_weight*: minimum sum of instance weight needed in a child. As long as the tree partition step results in leaf nodes with the sum of instance weight above min_child_weight, the building process continues. The larger, the more conservative the algorithm will be.
- *subsample*: subsample ratio of the training instances. For instance, a subsample value of 0.5 means that XGB would randomly sample half of the training data prior to growing trees. Subsampling will occur once in every boosting iteration.

The XGB algorithm was applied using the R package xgboost (Chen et al., 2018) in R 3.6.1 software (R Core Team, 2019).

- Support Vector Machine (SVM)

SVM is a supervised binary classification method (Burges, 1998). The basic idea of the multi-class SVM problem is to construct k classifiers, one for each class. The kth classifier constructs an optimal hyperplane between class n and the $k - 1$ other classes. To construct the optimal separating hyperplane, a kernel function is used to transform the input data into a high-dimensional space. Hyperparameters to tune are:

Table 2
Definition of the different behaviours under study.

Behaviour	Classification description
Grazing	Biting or browsing taking frequent bites, without raising the head
Walking	Movement from one location to another without lowering the head at ground level
Ruminating – Lying	Lying with regurgitating rumen bolus before it is chewing and then re-swallowing
Ruminating – Standing	Standing with regurgitating rumen bolus before it is chewing and then re-swallowing
Resting – Lying	Lying without rumination
Resting – Standing	Standing without movement or rumination
Other	All remaining behaviours

Table 3
Duration of observation and number of cows for each behaviour.

Behaviour	Duration (HH:MM:SS)	Number of cows
Grazing	27:38:57	62
Walking	01:10:19	60
Ruminating – Lying	08:05:26	44
Ruminating – Standing	03:34:27	35
Resting – Lying	08:00:00	57
Resting – Standing	03:40:10	66
Other	05:12:38	79
Total	57:21:57	86

Table 4
Number of sequences including each behaviour and number of time-windows associated with every behaviour for the selected configuration (10 s; 90%).

	Number of sequences including the behaviour	Number of windows associated with the behaviour
Grazing	814	63,136
Walking	110	2155
Ruminating – Lying	504	18,203
Ruminating – Standing	200	7024
Resting – Lying	94	20,084
Resting – Standing	131	7222
Total	1794	117,824

- *C*: cost of constraints violation. The larger, the smaller the margin of the hyperplane will be.
- *sigma*: inverse kernel width for the Radial Basis kernel function.

The SVM algorithm was applied using the R package kernlab (Karatzoglou et al., 2004) in R 3.6.1 software (R Core Team, 2019).

- Adaboost (ADA)

Adaboost is a boosting algorithm where the weak classifiers are decision trees with a single split. Adaboost works by affecting more weight on instances which are difficult to classify and less on those already correctly classified. In this way, boosting promotes new models capable of correctly classifying instances which were misclassified by earlier models. The impact of each model is weighted depending on its performance instead of offering the same weight to all models. Thus, the final model converges to a strong classifier (Subasi et al., 2018). Hyperparameters to tune are:

- *max_depth*: maximum depth of a tree. High values will tend to complicate the model with likely overfitting.
- *mfinal*: number of iterations for which boosting is run or the number of trees to use.

The Adaboost algorithm was applied using the R package adabag (Alfaro et al., 2013) in R 3.6.1 software (R Core Team, 2019).

Table 5
Abbreviation for each of the calculated features and list of the associated publications.

Abbreviation	Full name	References
$\overline{A}$	Average	Barwick et al. (2018) and Smith et al. (2016)
σ^2	Variance	Figo et al. (2010)
σ	Standard deviation	Bersch et al. (2014) and Smith et al. (2016)
Min	Minimum	Barwick et al. (2018) and Figo et al. (2010)
Max	Maximum	Barwick et al. (2018) and Figo et al. (2010)
Range	Range	Figo et al. (2010)
Q_2	Median	Fida et al. (2015)
Q_1	First quartile	Figo et al. (2010)
Q_3	Third quartile	Figo et al. (2010)
IQ	Interquartile range	Figo et al. (2010)
RMS	Root Mean Square X-axis	Bersch et al. (2014)
SMA	Signal Magnitude Area	Barwick et al. (2018)
AI	Average Intensity	Barwick et al. (2018)
MV	Movement Variation	Barwick et al. (2018) and Lush et al. (2018)
β_1	Skewness	Dutta et al. (2015) and Martiskainen et al. (2009)
β_2	Kurtosis	Dutta et al. (2015) and Martiskainen et al. (2009)
$Max\_corr_{x;y}$	The maximum of the correlation	Figo et al. (2010)
H_s	Spectral entropy	Zaccarelli et al. (2013)
$OBDA_1$	Overall Body Dynamic Acceleration	Benaissa et al. (2018)
$OBDA_2$	Overall Body Dynamic Acceleration	Wilson et al. (2008)
VeDBA	Vector of the Dynamic Body Acceleration	Walker et al. (2015)
pitch	Pitch angle	Walker et al. (2015)
roll	Roll angle	Walker et al. (2015)

• Random Forest (RF)

Random Forest is a bagging algorithm where many decision trees are randomly built. A subset of variables and samples are randomly chosen to build each decision tree. Each node of the decision trees aims at segmenting the subset of samples using the randomly selected variables. The target is to find the segmentation leading to the purest classification, *i.e.* leading to a maximum of instances of the same class in every node. For the classification tree, the Gini index is usually used to evaluate the decrease of impurity after each split. The final classification is carried out using every random decision trees (Breiman, 2001). Hyperparameters to tune are:

- *ntree*: number of iterations to carry out, *i.e.* number of decision trees to build.
- *mtry*: number of variables to choose at each iteration to build the decision tree.

The RandomForest algorithm was applied using the randomForest package (Liaw and Wiener, 2002) in R 3.6.1 software (R Core Team, 2019).

2.4.2. Models fitting and prediction of behaviour time-windows

The 4 Machine Learning models were fitted using the training data. For this purpose, several combinations of hyperparameters were tested from a generated grid of values specific to each algorithm (Section 2.4.1). For each combination of hyperparameters, the model was evaluated using a 10-fold cross validation procedure repeated three times. The accuracy and Cohen's Kappa associated with the prediction of each tested model were collected. This step was carried out with the R package caret (Kuhn et al., 2018) in R 3.6.1 (R Core Team, 2019). For each algorithm, the model leading to the best performance, called the fitted model, was applied to predict a behaviour in each 10 s-window using the testing data. At the end of this supervised classification step, the sequences of predicted behaviours for every 10 s-window were obtained for the 4 Machine Learning algorithms.

2.5. Consideration of the temporal structure within behaviours sequences

In order to account for the fact that the probability for a cow of expressing a specific behaviour at a given time depends on the behaviour she expressed just before, an HMM-based Viterbi algorithm using the behaviours predicted by the Machine Learning algorithms as input was used to predict its actual behaviours. The assumptions common to all HMMs are i) that the observations (behaviours predicted by the Machine Learning algorithms) are imperfect measures of a true state (actual behaviours) and that ii) the true states undergo Markovian transitions whereby the probability of being in a state at time t (within a given time window), only depends on the true state at time t − 1 (in the previous time window). Therefore, as proposed by Yang (2009), the goal of this smoothing step was to reassess the predictions of the behaviours obtained with the Machine Learning algorithms for each sequence of observation. The principle of the HMM-based Viterbi algorithm smoothing is illustrated in Fig. 3. All the matrices and vectors defining the HMM and mentioned in Fig. 3 are explained in the following sections.

The HMM-model noted Λ was used to estimate the unknown states of the process noted S, *i.e.* the behaviours actually carried out by the

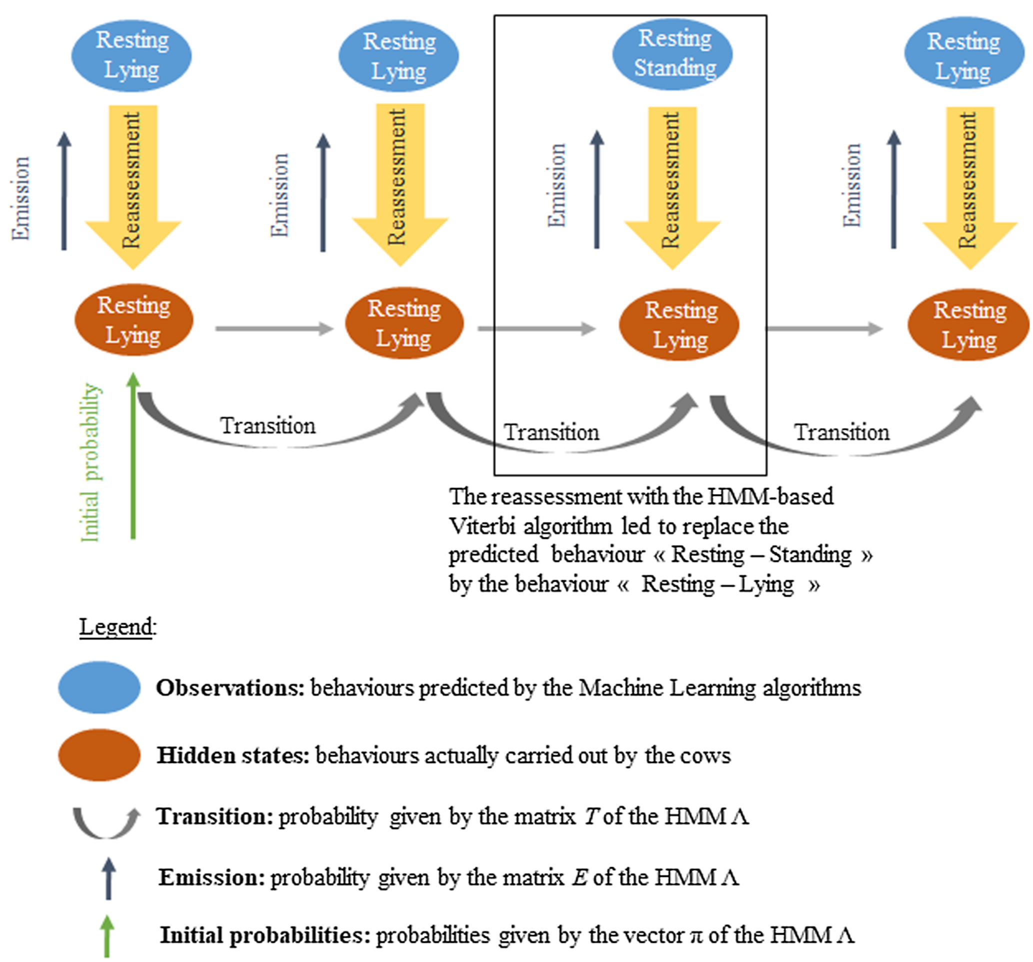

Fig. 3. Principle of the smoothing with the HMM-based Viterbi algorithm.

Table 6
Hyperparameters of fitted models for every algorithm.

Algorithm	Parameters of fitted models
XGB	nrounds = 250 max_depth = 6 eta = 0.4 gamma = 0 colsample_bytree = 0.8 min_child_weight = 1 subsample = 0.75
SVM	C = 128 sigma = 0.05
ADA	max_depth = 29 mfinal = 150
RF	mtry = 15 ntree = 2000

cows, knowing the measurable observations noted V, *i.e.* the behaviours predicted by the Machine Learning algorithms (Witten and Frank, 2011). The HMM Λ was also based on a transition matrix T providing the probability to move from a state, *i.e.*, a behaviour, to another at the next time-window. The transition matrix was obtained empirically, based on the data collected experimentally. The HMM was also defined by its emission matrix E, representing the probability that each observation was generated by each state. The emission matrix corresponds to the confusion matrix resulting from the prediction of the model. Finally, initial probabilities of each state, noted π, were finally used to define the HMM. The vector of initial probabilities was computed with the experimental data. All the matrices and vector used to define the HMM Λ in this study are detailed in Appendix B. A more detailed definition of the HMM is provided in the Appendix C. The function "initHMM" of the R package HMM (Himmelmann, 2010) in R 3.6.1 software (R Core Team, 2019) was used to define HMM-models.

The Viterbi algorithm was finally applied to find the most probable hidden behaviour from the preceding behaviour, based on the HMM $\Lambda = (T, E, \pi)$. We refer the reader to Forney (1973) for a more detailed description of the Viterbi algorithm. The Viterbi algorithm was applied with the function "viterbi" of the R package HMM (Himmelmann, 2010).

2.6. Evaluation of the prediction

The prediction was evaluated before and after smoothing with the HMM-based Viterbi algorithm. The overall performance of the models was evaluated with the accuracy, corresponding to the percentage of well-classified instances. The Cohen's Kappa (Cohen, 1960) was also used to assess performance. This metric provides the agreement between observation and prediction from a model. According to the criteria proposed by Cohen (1960), a Kappa higher than 0.81 was considered as an almost perfect agreement. The sensitivity and specificity of every behaviour were also calculated to evaluate the prediction of every behaviour. The sensitivity assessed the ability to return a positive outcome when the hypothesis is true. The specificity assessed the ability to return a negative outcome when the hypothesis is false. The equations for each parameter are provided in Appendix D. All the indicators were computed with the caret package (Kuhn et al., 2018) in R 3.6.1 (R Core Team, 2019).

3. Results

3.1. Results of model fitting

The fitted models obtained for each algorithm are presented in Table 6.

The accuracy and Cohen's Kappa got for every fitted model are provided in Fig. 4.

Both accuracy and Cohen's Kappa were higher than 0.99 for all the algorithms using a 10-fold cross validation with the training data. The best performance was obtained with the XGB algorithm followed by the RF algorithm (Fig. 4.).

3.2. Results of the models using independent observation sequences

3.2.1. Overall performance

The overall performance of every model before and after the HMM-based Viterbi smoothing are presented in Table 7. The confusion matrices obtained for every prediction before smoothing are shown in Appendix E and Appendix F, respectively.

The highest performance was obtained with XGB after smoothing (accuracy: 0.98; Cohen's Kappa: 0.96) followed by RF (accuracy: 0.97; Cohen's Kappa: 0.95). The lowest performance was obtained with the ADA algorithm before smoothing (accuracy: 0.95; Cohen's Kappa: 0.91). It should be noted that the smoothing with the HMM-based Viterbi algorithm did not appear to improve the overall performance of the models.

3.2.2. Discrimination of every behaviour

The prediction with the fitted models of XGB, RF, SVM and ADA was also evaluated for every behaviour, before and after smoothing. The sensitivities and the specificities of the predictions for every behaviour for the four algorithms are detailed in Fig. 5.

The best prediction results were obtained with XGB for every behaviour, except for the behaviour "Resting – Standing" where the sensitivity was higher with SVM (XGB: 0.78 and SVM: 0.82). The lowest performance was obtained with ADA, except for the sensitivity of "Walking" (0.78) and the specificity of "Grazing" (0.99) where the performance was lower with SVM (0.70 and 0.99, respectively).

The best prediction was obtained for the behaviour "Grazing" (sensitivity: 1.00 ± 0.00; specificity: 0.99 ± 0.00). The behaviour "Ruminating – Lying" also reached a sensitivity and a specificity higher than 0.97 and 0.99 after the HMM-based Viterbi smoothing, respectively. Similarly, the sensitivity and the specificity of the behaviour "Ruminating – Standing" were higher than 0.92 and 0.99 after smoothing, respectively. However, the sensitivity of the behaviour "Resting – Standing" reached a maximum of 0.82 with XGB after smoothing. The models also failed in the prediction of the behaviour "Walking" with a sensitivity lower than 0.80, except for the algorithm XGB where a sensitivity of 0.84 was obtained after smoothing.

It should be noted that the HMM-based Viterbi smoothing improved substantially the sensitivity of the behaviour "Resting - Standing" ($+0.04 \pm 0.01$) for which performance with the Machine Learning algorithm were the lowest.

4. Discussion

4.1. Improved performance in behaviour prediction in comparison with similar studies

Within the methodological framework proposed in the present study, the best prediction of dairy cow behaviours was obtained by combining (i) a preliminary study to find the most appropriate preprocessing of the raw accelerometer signal (window size of 10 s; overlap of 90%), (ii) a classification-prediction of the behaviours with the XGB algorithm, and (iii) a smoothing of the predictions using the HMM-based Viterbi algorithm. In the test dataset, the observed behaviours were predicted with an accuracy of 0.98 and a Cohen's Kappa of 0.96. The sensitivities and specificities were above 0.90 for all the behaviours, except for the behaviour "Resting–standing" behaviour (sensitivity: 0.82) and for the behaviour "Walking" (sensitivity: 0.84). Considering (1) the range of predicted behaviours, (2) the number of accelerometers fixed on the cows and (3) the method used to validate the models, these results appear better than those reported in the

Fig. 4. Results of model fitting for every algorithm.

Table 7
Prediction results (accuracy and Cohen's Kappa) before and after smoothing for every algorithm.

Algorithm	Accuracy		Cohen's Kappa	
	Before smoothing	After smoothing	Before smoothing	After smoothing
XGB	0.97	0.98	0.96	0.96
RF	0.97	0.97	0.95	0.95
SVM	0.96	0.97	0.94	0.95
ADA	0.95	0.95	0.91	0.92

literature.

4.1.1. *Prediction with a large range of behaviours*

A satisfactory prediction has already been obtained in similar studies when few behaviours were predicted. For instance, Andriamandroso (2017) have reached an accuracy of 0.92 to predict the 3 behaviours "Ruminating", "Grazing" and "Other" behaviours. Benaissa et al. (2018) have also predicted the same 3 behaviours with an accuracy of 0.93 using SVM. However, it should be noted that the behaviour "Other" was predicted in these two studies, which is also challenging as this class combined several different behaviours. The 3 behaviours "Grazing", "Ruminating" and "Resting" were also predicted

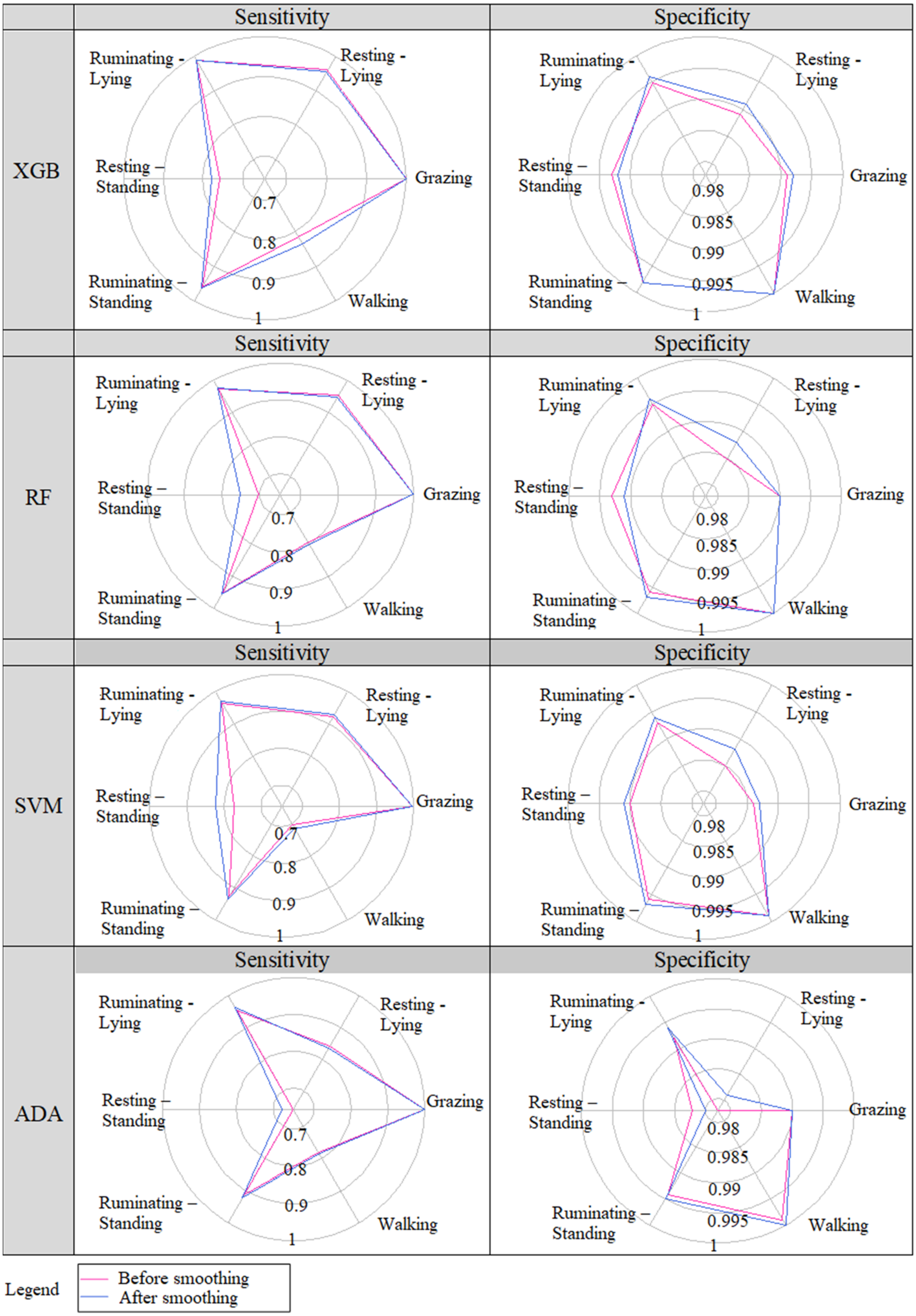

Fig. 5. Sensitivity and specificity of each behaviour for every algorithm, before and after smoothing with the HMM-based Viterbi algorithm.

with the satisfactory accuracy of 0.93 using a discriminant analysis (Giovanetti et al., 2017). However, in most studies, the performance drastically decreased when more behaviours were considered. For instance, Martiskainen et al. (2009) predicted 8 behaviours involving feeding behaviours as well as posture of cows using SVM. Their indicator called precision reached a value of 0.78 and a Cohen's Kappa of 0.69. In the same way, Alvarenga et al. (2016) predicted the behaviours "Grazing", "Lying", "Running", "Standing" and "Walking" of sheep and obtained an accuracy of 0.85 and a Cohen's Kappa of 0.79 using a window size of 5 s with a decision tree. Thus, our methodological framework seems to provide better performance than those obtained in similar studies in which both posture and feeding behaviours have been considered.

4.1.2. Prediction with a single sensor fixed on the neck

High performance has already been obtained when 2 sensors were fixed at 2 different positions on the animal. Indeed, the chosen behaviours are more or less well-predicted depending on the position of the sensor on the animal (Benaissa et al., 2017). For this reason, in order to predict both feeding behaviours and posture ("Lying", "Standing" and "Feeding") of cows, Benaissa et al. (2017) fixed an accelerometer on a neck-collar and another one on the leg to reach an accuracy of 0.98. However, fixing several sensors is tedious for both animals and handlers, as well as costly. As the procedure developed in this study aims to be used for practical implementations in farming, we chose to use a single sensor. It is indeed safer and more convenient for both farmers and animals than fixing several sensors at different places.

4.1.3. Reliable validation of the models

In many similar studies, each labelled window is affected either to the training dataset or to the testing dataset, independently of the sequence of observations to which the window belongs to (Lush et al., 2018; Martiskainen et al., 2009). Consequently, time-windows from the same cow at the same period of observation can be found in both training data and testing data. Due to the temporal structure of subsequent behaviours, the time-windows from the same sequence of observation are highly correlated. As explained by Rahman et al. (2018), the feature space distribution between the training data and the testing data is therefore very similar. The prediction is systematically better with this approach but this is not suitable to validate the algorithm. For instance, Lush et al. (2018) have predicted the behaviours "Foraging", "Walking", "Running", "Standing", "Lying" and "Urination event" in sheep with an accuracy of 0.96 and a Cohen's Kappa of 0.95 using Random Forest. However, 75% of the labelled windows were used to train the model and the remaining 25% were used to evaluate it, independently of the sequence of observation to which each window belongs. In our study, performance remained strong while windows from the same sequence of observation were either in the training data or in the testing data.

4.2. Methodology used to get a high performance

4.2.1. Field observation data

As noted by other authors (Rahman et al., 2018), the high between-cow variability in the recorded signal can decrease the performance of the prediction. For this reason, during data collection, particular attention was paid to cover as much variability as possible. Observations were carried out on 86 Holstein cows from 4 farms, of different parities and stages of lactation. The observations were also carried out in different paddocks (temporary *versus* permanent grasslands) and at different periods (summer of 2017 and spring and summer of 2018). In similar studies, the number of observed animals usually ranges between 5 and 40 animals from the same farm (Alvarenga et al., 2016; Lush et al., 2018). To the authors' knowledge, the dataset collected is larger and covers a wider range of conditions than those usually met in literature. As a consequence, the model was trained on a wider range of situations than what is usually found in the literature which could have decreased the performance of the prediction but which makes our conclusions more robust.

4.2.2. Appropriate pre-processing

The effect of the accelerometer signal pre-processing was widely approved in detection of human activities (Bersch et al., 2014). This aspect has received less attention in livestock monitoring behaviour, except for the impact of the window size (Robert et al., 2009; Smith et al., 2016). In a preliminary study, we identified window size and the percentage of overlap between windows to be critical when pre-processing the data (Riaboff et al., 2019). In the present work, two different window sizes and two different percentages of overlap were evaluated. This allowed predicting the behaviours with a pre-processing configuration favourable to strong performance of prediction.

4.2.3. Comparison of Machine Learning algorithms including eXtreme Gradient boosting

High performance has already been obtained in similar studies with SVM (Smith et al., 2016) and RF (Lush et al., 2018). The Adaboost algorithm has only been tested by Dutta et al. (2015) for monitoring livestock behaviour, whereas its efficiency has already been approved in other prediction problems (Subasi et al., 2018). To the best of our knowledge, the XGB algorithm has never been applied to the prediction of livestock behaviour, despite its high performance in other areas (Shi et al., 2019). The best prediction was reached with XGB in this study. We therefore advise considering the XGB algorithm to predict livestock behaviours with accelerometer data.

4.2.4. Consideration of the temporal structure within the behaviour sequences

The Viterbi algorithm based on HMM uses the temporal structure within observation sequences to find the most probable sequence of behaviours, given a sequence of predicted behaviours obtained with the Machine Learning algorithm. This approach may be particularly relevant to predict behaviours for which there is *a priori* no relevant feature from accelerometer data to achieve the discrimination. For instance, this is the case for the behaviours standing and lying as both the level of activity and the head position are similar (Martiskainen et al., 2009).

Yang (2009) showed that the accuracy of the prediction of daily activities in humans using a decision tree increased by 0.06 after the Viterbi smoothing. Although the improvement was less substantial in our study, the sensitivity of the behaviour "Resting – Standing" for which the lowest performance was obtained with the Machine Learning algorithms was increased considerably ($+ 0.04 \pm 0.01$).

4.3. Practical implications of the study

There is currently no reliable decision support tool to help pasture-based systems, neither for the pasture management, nor for the monitoring of animal health and welfare. The methodology developed in this study could be integrated in such a tool.

4.3.1. Pasture management support

In rotational grazing systems, a variation of 10% of milk yield can be explained by the fresh grass available to animals (Fulkerson et al., 2005). The allocation of fresh grass regulated by the rotation of the cows in the pastures is therefore critically important to keep high milk yield. Thus, a decision support tool indicating to farmers the appropriate time to allocate new fresh grass to the dairy cows would be relevant in pasture-based systems. Both changes in the lying behaviour (O'Driscoll et al., 2019) and feeding behaviour (Werner et al., 2019) are expected when the quantity and the quality of grass decline. In this respect, our methodology could be used to predict such behaviours. Other algorithms are also required to detect when there is not enough grass, based on the predicted behaviours. An automatic detection system to assist pasture management in rotational grazing systems could thus be developed, even if a work is still required on the technological aspects to make the system functional (battery life, computation time, automatic data transfer, etc.). In particular, a trade-off has to be found for the sampling rate in order to obtain both a satisfactory accuracy and a battery life compatible with the desired application (Benaissa et al., 2018).

4.3.2. Animal monitoring at pasture

- Detection of health events: the particular case of lameness detection

Some studies have shown that lameness is associated to changes to the behaviours of dairy cows (Chapinal et al., 2009), suggesting that behaviours could be used to detect lame cows (Barker et al., 2018). So far, the association between behaviour and lameness has been assessed considering either the lying behaviour (Yunta et al., 2012) or the feeding behaviour (Norring et al., 2014). As mentioned by Willshire and Bell (2009), the combination of lying behaviour with feeding behaviour should help to improve both sensitivity and specificity of lameness detection, which are currently low (Rutten et al., 2013). Consequently, the methodological framework developed in our study may relevant to the development of sensors predicting cow behaviour for the detection of lameness. Other algorithms are needed to detect lameness based on the predicted behaviours. Some technological aspects of the sensor should also be improved (battery life, automatic data transfer, computation time, etc.). This would alert automatically the farmer when a lame cow is detected to provide earlier treatments and mitigate the pain and costs associated with the disease (Willshire and Bell, 2009).

- Detection of a lack of comfort at pasture

Even if grazing is favourable to animal welfare (Burow et al., 2013), some stressful situations can also occur at pasture. For instance, Schütz et al. (2010) have shown that heat stress can occur when the temperature increases and there is no shadow in the paddock. In such situations, the distance of animals from the barn compromises the detection of the event by the farmer. As explained by Rushen et al. (2012), cow behaviour is a relevant indicator to assess welfare in farming but the biggest problem is the limited range of behaviours that have been measured automatically. In that regard, our methodological framework gives the opportunity to get information about the whole range of the main dairy cows' behaviours at pasture. An abnormal expression of one the 6 considered behaviours could allow the detection of welfare issues in particular if the behaviours are combined with the position of the cows collected with a GPS sensor. As an example, Riaboff et al. (2018) identified a heat stress situation through an overexpression of the behaviour "Resting" close to the drinking trough. Thus, our methodology could be used as a basis for the development of algorithms for detecting stressful events, even if some improvements on the technological aspects of the sensor are still needed, as previously explained. Such a tool would be a way to monitor 4 out of the 5 fundamental freedoms (Farm Animal Welfare Committee (FAWC), 2011), namely "Do not suffer from hunger and thirst", "Do not suffer from discomfort", "Do not suffer from pain, injuries and illnesses" and "Do not be afraid or stressed".

4.4. Limitations of the studies and potential solutions

4.4.1. Overestimation of the model accuracy in an experimental context

- Validation of the models with the same cows

In this study, each algorithm was evaluated at the observation sequence level, as discussed in Section 4.1.3. Consequently, the signal windows from the same observation sequence were either in the training data or in the testing data. However, windows from the same cow can be found in both datasets as the cows were observed several times during the experiment. Considering the cow-level, *i.e.* with windows from the same cow either in the training data or in the testing data, is the most accurate way to assess the model performance. This approach is sometimes considered in similar studies (Rahman et al., 2018) and leads to moderate performance, with an accuracy usually below 0.90 when a single sensor has been used. This method led also to a decrease of performance in our study, with an accuracy of 0.85 and a Cohen's Kappa of 0.76 using the XGB algorithm (*data not shown*). This decrease in performance is mainly due to the differences in the physical movement between dairy cows, leading to different motion patterns from a cow to another (Rahman et al., 2018). A way to improve the genericity of the algorithm may consist in considering a higher number of cows from various farms. Adding other sensors in the same electronic box fixed on the neck-collar could be also a way to better predict posture behaviours and then improve the genericity of the model.

- Removal of the remaining behaviours from the dataset

As explained in Section 2.1, behaviours other than the 6 under investigation, *i.e.* the "Other" class, were removed from the dataset. However, dairy cows can express these behaviours in real field situations. In these cases, the model will fail in the prediction because it has not been trained to predict these specific behaviours. Consequently, the prediction accuracy in real field situations will certainly be lower than in our development context. As the "Other" class is a combination of several behaviours, it is not possible to extract representative features for this class. Consequently, training a Machine Learning algorithm to predict the "Other" class does not seem appropriate. A solution would be to collect observations about each specific additional behaviour and then train the model to predict each event, as proposed by Lush et al. (2018) for the urination event.

- Using signal windows associated to a unique behaviour

As explained in Section 2.2, the signal sequences were split into 10 s-windows associated to a unique behaviour. These windows were then used to train the Machine Learning models. In a real field situation, dairy cows can express different behaviours within the same time window even if we chose a short window duration to prevent this issue. As the model has not been trained with heterogeneous time windows, the prediction may be wrong in such cases. This is another reason why the accuracy of the model might be reduced in real field situations. It would be interesting to train and test the models with heterogeneous windows in a second step to compare the accuracies, as proposed by Benaissa et al. (2018).

4.4.2. Confusion between the prediction of standing and lying

The lowest performance was obtained for the behaviour "Resting – Standing" (sensitivity: 0.82, specificity: 0.99 with XGB after smoothing) because of the confusion between the behaviour "Resting – Lying" (Appendix E). This confusion has already been mentioned in the literature when a single accelerometer is fixed on the animal neck (Benaissa et al., 2018; Martiskainen et al., 2009). As explained by Hamalainen et al. (2011), both the level of activity and the position of the head are similar between these two postures. Thus, it is difficult to find relevant features from accelerometer data alone to discriminate these two postures. Although this confusion was reduced with the Viterbi-based-HMM smoothing (Appendix F), the discrimination between these two postures remains an important challenge when using a single accelerometer on the neck. A solution might be to add other sensors in the same electronic box, like a magnetometer, a gyrometer or an altimeter. By merging data from different sensors, new relevant features could be found to discriminate lying and standing positions.

5. Conclusion

This study aimed to develop a methodological framework to predict feeding behaviour and posture of dairy cows at pasture from accelerometer data. For this purpose, a suitable pre-processing of the raw accelerometer signal was applied (10 s time-window, 90% overlap) and several methods of Machine Learning were compared (eXtreme Gradient Boosting, Random Forest, Support Vector Machine and Adaboost). The subsequent application of a Viterbi algorithm was used to account for the temporal structure within behaviour sequences. The best prediction capacity was obtained using the XGB algorithm followed by the Viterbi smoothing, leading to an accuracy of 0.98 and a Cohen's Kappa of 0.96 on a test dataset. These prediction results were higher than those obtained in the similar studies. Our methodology therefore has great potential for the development of decision-support tools in grazing systems, both for assisting pasture management and for monitoring animal health and welfare.

The methodology led to an excellent accuracy of the prediction of the behaviours but some improvements are still needed. It is still difficult to discriminate the standing from the lying posture of the cows using a single accelerometer fixed on the neck. Another important aspect is the genericity of the developed procedure. These two issues remain a main challenge for the automatic monitoring of livestock behaviour. One potential solution would consist in combining accelerometer data with data from other sensors, like a gyrometer, a magnetometer and an altimeter, embedded in the same electronic box fixed on the neck-collar. Merging data from different signals could help to better discriminate the postures of the cows and ensure that the methodology is applicable to any farms.

CRediT authorship contribution statement

L. Riaboff: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **S. Poggi:** Conceptualization, Methodology, Formal analysis, Writing - review & editing. **A. Madouasse:** Conceptualization, Writing - review & editing. **S. Couvreur:** Conceptualization, Writing - review & editing. **S. Aubin:** Conceptualization, Writing - review & editing. **N. Bédère:** Conceptualization, Writing - review & editing. **E. Goumand:** Conceptualization, Project administration. **A. Chauvin:** Conceptualization. **G. Plantier:** Conceptualization, Funding acquisition.

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.

Appendix A. Equation of the raw signal magnitude and the calculated features presented in Table 5

Equation of the raw signal magnitude

$a_{mag} = \sqrt{{a_x}^2 + {a_y}^2 + {a_z}^2}$

Equation of features

$\overline{Ax} = \frac{1}{M}\sum_{i=1}^{M} a_{xi}$

${\sigma_x}^2 = \frac{1}{M}\sum_{i=1}^{M} (a_{xi} - \overline{A_x})^2$

$\sigma_x = \sqrt{{\sigma_x}^2}$

$Range_x = Max_x - Min_x$

$IQ_x = \mathrm{Q}_{3;x} - \mathrm{Q}_{1;x}$

$RMS = \sqrt{\frac{1}{M}\sum_{i=1}^{M} {a_{xi}}^2}$

$SMA = \frac{1}{M}\left(\sum_{i=1}^{M} |a_{xi}| + \sum_{i=1}^{M} |a_{yi}| + \sum_{i=1}^{M} |a_{zi}|\right)$

$AI = \frac{1}{M}\sum_{i=1}^{M} a_{mag}$

$MV = \frac{1}{M}\left(\sum_{i=1}^{M-1} |a_{x,i+1} - a_{x,i}| + \sum_{i=1}^{M-1} |a_{y,i+1} - a_{y,i}| + \sum_{i=1}^{M-1} |a_{z,i+1} - a_{z,i}|\right)$

$\beta_{1;x} = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{a_{xi} - \overline{Ax}}{\sigma}\right)^3$

$\beta_{2;x} = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{a_{xi} - \overline{Ax}}{\sigma}\right)^4$

$H_s = -\sum_{k=1}^{N} P_k \, ln(P_k)$ with $P_k = \frac{|\lambda_k|^2}{\sum_i |\lambda_i|^2}$ and for the frequency λ_k

$OBDA_1 = |a_{dynamic_{amag_i}}|$

$OBDA_2 = |a_{dynamic_{axi}}| + |a_{dynamic_{ayi}}| + |a_{dynamic_{azi}}|$

$VeDBA = \sqrt{{a_{dynamic\_ax\_i}}^2 + {a_{dynamic\_ay\_i}}^2 + {a_{dynamic\_az\_i}}^2}$

$Pitch = Arcsin\left(\frac{-a_{static\_y}}{\sqrt{{a_{static\_x}}^2 + {a_{static\_y}}^2 + {a_{static\_z}}^2}}\right) * \frac{180}{\pi}$

$Roll = 2Arctan\left(\frac{a_{static\_x}}{a_{static\_y} + \sqrt{{a_{static\_y}}^2 + {a_{static\_x}}^2}}\right) * \frac{180}{\pi}$

Normalization of features

$c'_i = \frac{c_i - c_{min}}{c_{max} - c_{min}}$ where c'_i the normalized value, c_i is the original value, c_{min} is the minimum value of feature c for all observations and c_{max} is the maximum value of feature c for all observations

Appendix B. Transition matrix *T*, emission matrices *E* and vector of probabilities π used to reassess the prediction of each Machine Learning algorithm

HMM-model is noted $\Lambda = (T, E, \pi)$. The transition matrix *T*, the emission matrix *E* and the vector of probabilities π used to reassess the predicted behaviours in this study are the following:

- The transition matrix T was obtained empirically using the data collected during the experiment. The following matrix was used:

$T =$

		Instant t + 1					
	Hidden states	Grazing	Resting – Lying	Ruminating – Lying	Resting –Standing	Ruminating – Standing	Walking
Instant t	Grazing	0.9964	0.0001	0.0001	0.0004	0.0001	0.0029
	Resting – Lying	0.0001	0.9983	0.0013	0.0001	0.0001	0.0001
	Ruminating – Lying	0.0001	0.0041	0.9955	0.0001	0.0001	0.0001
	Resting – Standing	0.01325	0.0001	0.0001	0.97775	0.00735	0.00145
	Ruminating – Standing	0.00305	0.0001	0.0001	0.00155	0.99055	0.00465
	Walking	0.10460	0.0001	0.0001	0.02080	0.0001	0.8743

It should be noted that when the observed probability associated to one transition was zero, it was replaced by 0.0001.

- The emission matrix E corresponds to the confusion matrix from the prediction with each Machine Learning algorithm. The following matrices were used:

XGB	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	0.993	0.000	0.000	0.000	0.000	0.007
	Resting – Lying	0.000	0.910	0.003	0.086	0.001	0.000
	Ruminating – Lying	0.0.000	0.004	0.981	0.011	0.002	0.002
	Resting – Standing	0.000	0.077	0.016	0.850	0.057	0.000
	Ruminating – Standing	0.001	0.000	0.022	0.003	0.974	0.000
	Walking	0.003	0.000	0.000	0.000	0.000	0.997

SVM	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	0.989	0.000	0.001	0.000	0.001	0.010
	Resting – Lying	0.008	0.874	0.022	0.086	0.004	0.006
	Ruminating – Lying	0.000	0.007	0.972	0.007	0.012	0.002
	Resting – Standing	0.000	0.139	0.017	0.797	0.044	0.003
	Ruminating – Standing	0.000	0.000	0.041	0.003	0.951	0.004
	Walking	0.023	0.000	0.004	0.000	0.010	0.963

ADA	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	0.992	0.000	0.000	0.000	0.000	0.008
	Resting – Lying	0.000	0.811	0.032	0.145	0.012	0.001
	Ruminating – Lying	0.000	0.009	0.967	0.005	0.015	0.003
	Resting – Standing	0.000	0.303	0.015	0.636	0.045	0.001
	Ruminating – Standing	0.001	0.000	0.070	0.008	0.920	0.001
	Walking	0.014	0.000	0.023	0.000	0.007	0.956

RF	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	0.992	0.000	0.000	0.000	0.000	0.008
	Resting – Lying	0.000	0.878	0.010	0.111	0.001	0.000
	Ruminating – Lying	0.000	0.002	0.981	0.009	0.006	0.002
	Resting – Standing	0.000	0.103	0.017	0.824	0.055	0.001
	Ruminating – Standing	0.000	0.000	0.044	0.002	0.954	0.000
	Walking	0.007	0.000	0.005	0.000	0.000	0.988

In these matrices, the true behaviours are in rows and the predicted behaviours in columns. Each occurrence in the confusion matrix was expressed as a percentage of the total number of occurrences of the corresponding row. In this way, the sum of each row was equal to 1.

- The vector π was obtained empirically using the experimental data. The following vector was obtained:

$\pi =$

Hidden states	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Initial probabilities	0.5289	0.1607	0.1627	0.0622	0.0635	0.0219

Appendix C. Definition of the HMM algorithm

A HMM-model is noted $\Lambda = (T, E, \pi)$ and based on:

• Its hidden states which constitute the ensemble $S = \{s_1, s_2, \cdots, s_n\}$ where n is the number of hidden states. The state where is the HMM at the instant t is noted q_t ($q_t \in S$).

• m observable symbols in each state. All the possible observations constitute the ensemble $V = \{v_1, v_2, \cdots, v_m\}$. $O_t \in V$ is the observed symbol at the instant t.

• A matrix of transition probabilities between states, noted T and defined as follows:

$T(i, j) = P(q_{t+1} = s_j | q_t = s_i)\ \forall\ i, j \in \{1, \ldots, n\}$ and $\forall\ t \in \{1, \ldots, T\}$ with:

$T(i, j) > 0\ \forall\ i, j$ and.$\sum_{j=1}^{n} T(i, j) = 1$

• A matrix of probabilities of observation of each symbol in each state, called emission probability and noted E. This matrix is defined as follows:

$E(j, k) = P(O_t = v_k | q_t = s_j)\ \forall\ j \in \{1, \ldots, n\}$ and $\forall\ k \in \{1, \ldots, m\}$ with:

$E(j, k) > 0$ and.$\sum_{k=1}^{m} E(j, k) = 1$

• A vector π of initial probabilities: $\pi = \{\pi_1, \pi_2, \cdots, \pi_n\}$ where for each state i, π_i is the probability that the initial state of the HMM was i:

$\pi_i = P(q_1 = s_i)\ \forall\ i, \in \{1, \ldots, n\}$ with $\pi_i > 0\ \forall$ i and.$\sum_{i=1}^{n} \pi_i = 1$

Once the three parameters (T, E, π) of the HMM were determined, the optimal sequence of hidden states, *i.e.*, the most probable hidden states given a sequence of observations, is obtained with the Viterbi algorithm.

Appendix D. Performance metrics used to assess the models

$$accuracy = \frac{number\ \ of\ \ well\ \ classified\ \ instances}{number\ \ of\ \ instances}$$

$$Cohen's\ \ Kappa = \frac{p_o - p_e}{1 - p_e}$$

$$sensitivity_i = \frac{TP_i}{TP_i + FN_i}$$

$$specificity_i = \frac{TN_i}{FP_i + TN_i}$$

where p_o is the relative observed agreement among raters and p_e is the hypothetical probability of chance agreement. The index i is the considered behaviour; TP_i (True positives) is the number of instances where behaviour i was observed and correctly predicted; FN_i (False negatives) is the number of instances where behaviour i was observed but another behaviour was predicted, FP_i (False positives) is the number of instances where behaviour i was predicted but another behaviour was observed; TN_i (True negatives) is the number of instances where behaviour i was not observed and not predicted.

Appendix E. Confusion matrices obtained with every model using the test dataset before the HMM-based Viterbi smoothing

XGB	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	16,513	0	0	0	0	118
	Resting – Lying	0	2857	10	269	4	1
	Ruminating – Lying	0	25	5663	64	12	11
	Resting – Standing	0	96	20	1057	71	0
	Ruminating – Standing	3	0	46	6	2052	0
	Walking	2	0	0	0	0	571

SVM	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	16,479	0	11	0	10	163
	Resting – Lying	26	2751	68	271	12	20
	Ruminating – Lying	0	39	5549	38	69	13
	Resting – Standing	0	188	23	1080	60	4
	Ruminating – Standing	1	0	86	7	1983	9
	Walking	12	0	2	0	5	492

ADA	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	16,508	0	0	0	0	131
	Resting – Lying	0	2501	99	446	36	3
	Ruminating – Lying	0	48	5458	31	87	18
	Resting – Standing	0	429	21	901	64	1
	Ruminating – Standing	2	0	148	18	1948	2
	Walking	8	0	13	0	4	546

RF	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	16,514	0	0	0	0	131
	Resting – Lying	0	2844	33	359	3	1
	Ruminating – Lying	0	11	5589	51	35	11
	Resting – Standing	0	123	20	981	66	1
	Ruminating – Standing	0	0	94	5	2035	0
	Walking	4	0	3	0	0	557

Appendix F. Confusion matrices obtained with every model using the test dataset after the HMM-based Viterbi smoothing

XGB	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	16,514	0	0	0	0	103
	Resting – Lying	0	2840	5	237	0	0
	Ruminating – Lying	0	16	5666	62	5	11
	Resting – Standing	0	122	25	1088	74	1
	Ruminating – Standing	2	0	43	9	2060	0
	Walking	2	0	0	0	0	586

SVM	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	16,480	0	10	0	10	154
	Resting – Lying	24	2768	59	204	9	19
	Ruminating – Lying	0	32	5573	36	53	15
	Resting – Standing	0	178	22	1149	61	3
	Ruminating – Standing	1	0	75	7	2001	9
	Walking	13	0	0	0	5	501

ADA	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	16,510	0	0	0	0	127
	Resting – Lying	0	2487	68	406	20	2
	Ruminating – Lying	0	39	5521	24	69	22
	Resting – Standing	6	452	16	942	79	1
	Ruminating – Standing	2	0	122	24	1971	0
	Walking	0	0	12	0	0	549

RF	Prediction	Grazing	Resting – Lying	Ruminating – Lying	Resting – Standing	Ruminating – Standing	Walking
Reference	Grazing	16,514	0	0	0	0	128
	Resting – Lying	0	2820	23	295	0	0
	Ruminating – Lying	0	9	5604	46	29	11
	Resting – Standing	0	149	25	1048	64	2
	Ruminating – Standing	0	0	84	7	2046	0
	Walking	4	0	3	0	0	560

References

Alfaro, E., Gámez, M., García, N., 2013. adabag: An R package for classification with boosting and bagging. J. Stat. Softw. 54, 1–35. https://doi.org/10.18637/jss.v054.i02.

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. Anim. Behav. Sci. 181, 91–99. https://doi.org/10.1016/j.applanim.2016.05.026.

Andriamandroso, A.L.H., 2017. Cattle grazing dynamics under contrasted pasture characteristics at temporal and spatial scales (Sciences agronomiques et ingénierie biologique). Université de Liège, Gembloux Agro-Bio Tech, Liège, Belgique.

Barker, Z.E., Vázquez Diosdado, J.A., Codling, E.A., Bell, N.J., Hodges, H.R., Croft, D.P., Amory, J.R., 2018. Use of novel sensors combining local positioning and acceleration to measure feeding behavior differences associated with lameness in dairy cattle. J. Dairy Sci. 101, 6310–6321. https://doi.org/10.3168/jds.2016-12172.

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. Anim. Behav. Sci. 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. https://doi.org/10.1016/j.rvsc.2017.10.005.
Bersch, S., Azzi, D., Khusainov, R., Achumba, I., Ries, J., 2014. Sensor Data Acquisition and Processing Parameters for Human Activity Classification. Sensors 14, 4239–4270. https://doi.org/10.3390/s140304239.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. 2, 121–167.
Burow, E., Rousing, T., Thomsen, P.T., Otten, N.D., Sørensen, J.T., 2013. Effect of grazing on the cow welfare of dairy herds evaluated by a multidimensional welfare index. Animal 7, 834–842. https://doi.org/10.1017/S1751731112002297.
Carvalho, P.C.F., 2013. Harry Stobbs memorial lecture: can grazing behavior support innovations in grassland management? Trop. Grassl. 1, 137–155.
Chapinal, N., de Passillé, A.M., Weary, D.M., von Keyserlingk, A.G., Rushen, J., 2009. Using gait score, walking speed, and lying behavior to detect hoof lesions in dairy cows. J. Dairy Sci. https://doi.org/10.3168/jds.2009-2115.
Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., 2018. xgboost: Extreme Gradient Boosting.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46. https://doi.org/10.1177/001316446002000104.
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. 18–28. https://doi.org/10.1016/j.compag.2014.12.002.
Farm Animal Welfare Committee (FAWC) [WWW Document], 2011. URL https://www.gov.uk/government/groups/farm-animal-welfare-committee-fawc.
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.
Figo, D., Diniz, P.C., Ferreira, D.R., Cardoso, J.M.P., 2010. Preprocessing techniques for context recognition from accelerometer data. Pers. Ubiquitous Comput. 14, 645–662. https://doi.org/10.1007/s00779-010-0293-9.
Forney, G.D., 1973. The Viterbi algorithm. Proc. IEEE 61, 268–278.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.
Fulkerson, W.J., McKean, K., Nandra, K.S., Barchia, I.M., 2005. Benefits of accurately allocating feed on a daily basis to dairy cows grazing pasture. Aust. J. Exp. Agric. 45, 331. https://doi.org/10.1071/EA03109.
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. Livest. Sci. 196, 42–48. https://doi.org/10.1016/j.livsci.2016.12.011.
Gonzalez, 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. 91–102. https://doi.org/10.1016/j.compag.2014.10.018.
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), pp. 831–836. https://doi.org/10.1109/ISDA.2011.6121760.
Himmelmann, L., 2010. HMM: HMM – Hidden Markov Models.
Kamphuis, C., Delarue, B., Burke, C.R., Jago, J., 2012. Field evaluation of 2 collar-mounted activity meters for detecting cows in estrus on a large pasture-grazed dairy farm. J. Dairy Sci. 95, 3045–3056. https://doi.org/10.3168/jds.2011-4934.
Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. kernlab – An S4 Package for Kernel Methods in R. J. Stat. Softw. 11, 1–20. https://doi.org/10.18637/jss.v011.i09.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2018. caret: Classification and Regression Training.
Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2, 18–22. https://CRAN.R-project.org/doc/Rnews/.
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.
Martiskainen, P., Järvinen, M., Skön, J.P., Tiirikainen, J., Kolehmainen, M., Mononen, J., 2009. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. Appl. Anim. Behav. Sci. 119, 32–38.
Mathie, M.J., Coster, A.C.F., Lovell, N.H., Celler, B.G., 2004. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiol. Meas. 25, R1–R20. https://doi.org/10.1088/0967-3334/25/2/R01.
McSweeney, D., Foley, C., Halton, P., O'Brien, B., 2015. Calibration of an automated grass measurement tool to enhance the precision of grass measurement in pasture based farming systems. In: Teagasc Ag Conference.
Norring, M., Häggman, J., Simojoki, H., Tamminen, P., Winckler, C., Pastell, M., 2014. Short communication: Lameness impairs feeding behavior of dairy cows. J. Dairy Sci. 97, 4317–4321. https://doi.org/10.3168/jds.2013-7512.
O'Driscoll, K., Lewis, E., Kennedy, E., 2019. Effect of feed allowance at pasture on the lying behaviour of dairy cows. Appl. Anim. Behav. Sci. 213, 40–46. https://doi.org/10.1016/j.applanim.2019.02.002.
Penning, P.D., 1983. A technique to record automatically some aspects of grazing and ruminating behaviour in sheep. Grass Forage Sci. 38, 89–96. https://doi.org/10.1111/j.1365-2494.1983.tb01626.x.
R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
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. Inf. Process. Agric. 5, 124–133. https://doi.org/10.1016/j.inpa.2017.10.001.
Riaboff, L., Aubin, S., Bédère, N., Couvreur, S., Madouasse, A., Goumand, E., Chauvin, A., Plantier, G., 2019. Evaluation of pre-processing methods for the prediction of cattle behaviour from accelerometer data. Comput. Electron. Agric. 165, 104961. https://doi.org/10.1016/j.compag.2019.104961.
Riaboff, L., Bedere, N., Couvreur, S., Aubin, S., Goumand, E., Magnier, J., Madouasse, A., Chauvin, A., Plantier, G., 2018. Influence of pasture characteristics and time of day on dairy cow behaviour predicted from GPS-data. In: 7 Gen Meet. Eur. Grassl. Fed. EGF 23. pp. 803–809.
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, 80–84.
Rushen, J., Chapinal, N., de Passillé, A., 2012. Automated monitoring of behavioural-based animal welfare indicators. Anim. Welf. 21, 339–350. https://doi.org/10.7120/09627286.21.3.339.
Rutten, C.J., Velthuis, A.G.J., Steeneveld, W., Hogeven, H., 2013. Sensors to support health management on dairy farms. J. Dairy Sci. 1928–1952.
Schlecht, E., Dickhoefer, U., Gumpertsberger, E., Buerkert, A., 2009. Grazing itineraries and forage selection of goats in the Al Jabal al Akhdar mountain range of northern Oman. J. Arid Environ. 73, 355–363. https://doi.org/10.1016/j.jaridenv.2008.10.013.
Schütz, K.E., Rogers, A.R., Poulouin, Y.A., Cox, N.R., Tucker, C.B., 2010. The amount of shade influences the behavior and physiology of dairy cattle. J. Dairy Sci. 93, 9.
Shahriar, Md.S., Smith, D., Rahman, A., Freeman, M., Hills, J., Rawnsley, R., Henry, D., Bishop-Hurley, G., 2016. Detecting heat events in dairy cows using accelerometers and unsupervised learning. Comput. Electron. Agric. 128, 20–26. https://doi.org/10.1016/j.compag.2016.08.009.
Shi, H., Wang, H., Huang, Y., Zhao, L., Qin, C., Liu, C., 2019. A hierarchical method based on weighted extreme gradient boosting in ECG heartbeat classification. Comput. Methods Programs Biomed. 171, 1–10. https://doi.org/10.1016/j.cmpb.2019.02.005.
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.
Subasi, A., Dammas, D.H., Alghamdi, R.D., Makawi, R.A., Albiety, E.A., Brahimi, T., Sarirete, A., 2018. Sensor based human activity recognition using adaboost ensemble classifier. Procedia Comput. Sci. 140, 104–111. https://doi.org/10.1016/j.procs.2018.10.298.
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. Biotelemet. 3. https://doi.org/10.1186/s40317-015-0045-8.
Walker, J.S., Jones, M.W., Laramee, R.S., Holton, M.D., Shepard, E.L., Williams, H.J., Scantlebury, D.M., Marks, N., 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. https://doi.org/10.1186/s40462-015-0056-3.
Werner, J., Umstatter, C., Kennedy, E., Grant, J., Leso, L., Geoghegan, A., Shalloo, L., Schick, M., O'Brien, B., 2019. Identification of possible cow grazing behaviour indicators for restricted grass availability in a pasture-based spring calving dairy system. Livest. Sci. 220, 74–82. https://doi.org/10.1016/j.livsci.2018.12.004.
Willshire, J.A., Bell, N.J., 2009. An economic review of cattle lameness. Cattle Pract. 17, 136–141.
Wilson, R., Shepard, E., Liebsch, N., 2008. Prying into the intimate details of animal lives: use of a daily diary on animals. Endanger. Species Res. 4, 123–137. https://doi.org/10.3354/esr00064.
Witten, I.H., Frank, E., 2011. Data Mining: Practical Machine Learning Tools and Techniques. Elsevier https://doi.org/10.1016/C2009-0-19715-5.
Yang, J., 2009. Toward physical activity diary: motion recognition using simple acceleration features with mobile phones. In: Proceedings of the 1st International Workshop on Interactive Multimedia for Consumer Electronics, pp. 1–10.
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, 6546–6549. https://doi.org/10.3168/jds.2012-5670.
Zaccarelli, N., Li, B.-L., Petrosillo, I., Zurlini, G., 2013. Order and disorder in ecological time-series: Introducing normalized spectral entropy. Ecol. Indic. 28, 22–30. https://doi.org/10.1016/j.ecolind.2011.07.008.