Machine learning to detect behavioural anomalies in dairy cows under subacute ruminal acidosis

Nicolas Wagner[a,b], Violaine Antoine[a], Marie-Madeleine Mialon[b], Romain Lardy[b], Mathieu Silberberg[b], Jonas Koko[a], Isabelle Veissier[b,*]

[a] *Université Clermont Auvergne, CNRS, UMR 6158 LIMOS, Campus des Cézeaux – BP 10125, F-63173 Aubière cedex, France*
[b] *Université Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, F-63122 Saint-Genès-Champanelle, France*

ABSTRACT

Sickness behaviour is characterised by a lethargic state during which the animal reduces its activity, sleeps more and at times when normally awake, reduces its feed and water intake, and interacts less with its environment. Subtle modifications in behaviour can materialise just before clinical signs of a disease. Recent sensor developments enable continuous monitoring of animal behaviour, but the shift to abnormal animal activity remains difficult to detect. We explored the use of Machine Learning (**ML**) to detect abnormal behaviour from continuous monitoring. We submitted 14 cows (Bos taurus) to Sub-Acute Ruminal Acidosis (**SARA**), a disease known to induce changes in behaviour. Another 14 control cows were not submitted to SARA. We used a ruminal bolus to monitor pH and detect when a cow experienced SARA. We used a positioning system to infer an animal's activity based on its position in relation to specific elements in the barn (feeder, resting area, and alleys). We tested several ML algorithms: K Nearest Neighbours for Regression (**KNNR**); Decision Tree for Regression (**DTR**); MultiLayer Perceptron (**MLP**); Long Short-Term Memory (**LSTM**); and an algorithm where activity is assumed to be similar from one day to the next. First, we developed ML models to predict activity on a given day from the previous 24 h, considering all cows together. Then, we calculated the error between observed and predicted values for a given cow. Finally, we compared the error to a threshold chosen to optimise the distinction between normal and abnormal values. KNNR performed best, detecting 83% of SARA cases (true-positives), but it also produced 66% of false-positives, which limits its use in practice. In conclusion, ML can help detect anomalies in behaviour. Further improvements could probably be obtained by applying ML on very large datasets at animal rather than group level.

1. Introduction

When animals get infected, it can cause a range of behavioural modifications leading to a lethargic state during which the animal reduces its activity, sleeps more and at times when it is normally awake, reduces its feed and water intake, and interacts less with conspecifics or with humans (Hart, 1988; Dantzer & Kelley, 2007; Byrd & Lay, 2018). This sickness behaviour is observed in many species, including humans. It is induced by cytokines released at the onset of infection that act on the brain. Sickness behaviour is thought to reduce energy expenditure to sustain the high metabolic cost of the fever response and thus facilitate recovery from infection (Hart, 1988; Dantzer & Kelley, 2007). Alternatively, sickness behaviour could help reduce the spread of diseases among groups of animals (Shakhar & Shakhar, 2015), but this hypothesis has not yet been confirmed (Lopes et al., 2018).

Sickness behaviour can start before actual clinical symptoms of disease. For instance, cows are less reactive to their environment and change activity less often a few hours before fever due to *Escherichia coli*-induced mastitis (De Boyer Des Roches et al., 2017). In addition, the circadian rhythm of cattle activity can be less marked from as early as two days before clinical symptoms are detected by caretakers (Veissier et al., 1989b; Veissier et al., 2017). There is therefore scope to use such behaviours as early signs of a disease in order to detect and remedy health disorders very early on, and thus reduce medicine intakes and improve animal welfare.

Up to recently, it was difficult to routinely detect such behavioural changes because they require long periods of closely observing the animals. Today, there are readily-available sensors and Precision Livestock Farming (**PLF**) tools that offer the possibility to continuously monitor animal behaviour. Accelerometers can record gross activities

---

* Corresponding author.
  E-mail address: isabelle.veissier@inrae.fr (I. Veissier).

such as walking, eating, lying down, positioning systems can connect an animal positions to specific resources and thus infer its putative activity, and image analysis can provide further input (González et al., 2015; Andriamandroso et al., 2017; Barwick et al., 2018). As argued by Dawkins et al. (2012), the issue is not about collecting data but how to process it. More specifically, to detect early signs of diseases through behavioural modifications, we need to be able to predict normal behaviour and to estimate deviations from this norm. Modelling behaviour should be of great help for this purpose.

Machine Learning (**ML**) can greatly help the study of behaviour (Valletta et al., 2017). ML consists in learning a model that corresponds to the performance of the data with no *a priori* hypotheses. The model can then help to classify data into groups according to a given criterion. Examples of the use of ML in behavioural studies include tracking mice to analyse their social behaviour, assigning eggs to individual female pheasants in mixed clutches, identifying social networks in jackdaws, counting wildebeest from aerial photos, or detecting oestrus from vaginal temperature and conductivity (Hong et al., 2015; Valletta et al., 2017; Higaki et al., 2018). So far, to our knowledge, ML has not been used to predict future behaviour.

Here we explored ML to predict cow behaviours in order to detect anomalies potentially reflecting a health disorder. We used a Real-Time Locating System (**RTLS**) to continuously record the position of cows in a barn. As the various in-barn activities of a cow are associated with in-barn locations, we inferred the activity of each cow according to its position in relation to resources such as feeders and the resting area. We submitted cows to a Sub-Acute Ruminal Acidosis (**SARA**) as a model of a disease that induces behavioural changes (Commun et al., 2012; Silberberg et al., 2017). We analysed the activity as a time series and predicted behaviour from one day to the next. A deviation from the prediction reveals abnormal behaviour, and could thus signal a health disorder.

2. Materials and methods

2.1. Ethical note

The experiment was conducted at the INRAE 'Herbipôle' experimental facility (https://doi.org/10.15454/1.5572318050509348E12, UE 1414, Marcenat, France), which has French Ministry of Agriculture approval to carry out experiments on live animals (EEA accreditation #C15-114-01). All scientists and technicians involved in the experiment have received initial training for experiments on live animals and are regularly retrained to maintain and refresh their capacities, in line with French regulations governing experiments on animals. The protocol was submitted to and approved by the regional ethics committee (approval: APAFIS366). All these measures are aligned with the framework of the EU Directive 2010/63 for the protection of animals used for scientific purposes. In addition, cow health was checked closely every day by caretakers using direct observations and data from sensors (see below).

2.2. Animals and husbandry

A total of 28 Holstein-breed dairy cows aged 3–7 years were monitored for 9 weeks from 3 months after the start of lactation. The cows were housed in a 11.7 × 21.4 m pen equipped with as many feeding places and resting cubicles as cows (Fig. 1B). They were fed forage (hay and wrapped hay) and starch-containing concentrates. The cows were divided into two treatments (n = 14 in each treatment). In the first treatment (controls), the cows were fed a diet with 75% forage and 25% concentrates (corresponding to 10.5% starch in the diet) during the entire observation period. In the second treatment, after 4 weeks of control diet, the proportion of concentrates was progressively increased to reach 46% in 10 days (31.5% starch) and the cows were maintained on this diet for 2 weeks before resuming the control diet for 3 weeks.

Table 1 summarises the characteristics of the two diets. This treatment was designed to induce a SARA, which is characterised by a low pH in the rumen (Villot et al., 2018).

2.3. Observation

Cow activities were captured and collected using an RTLS system (CowView, GEA Farm Technologies, Bönen, Germany). The system consists of a locating sensor fixed on top of each cow collar that emits a signal in the ultra-wideband area that is captured by antennas fixed in the barn ceiling (Fig. 1). The position of the cow is determined by triangulation every second. If the cow is found in a cubicle, she is considered to be resting, if she is found near a trough, she is considered to be eating, and if she is found in the alley, she is considered to be walking (if she is moving) or standing (if not moving). For the purposes of this experiment, we merged walking with standing, as direct observations proved that the device did not precisely distinguish these two activities.

As in Veissier et al. (2017), we performed a factorial correspondence analysis with hours of the day as observations and proportion of cows in each activity as variables. We used the weights of activities on the first component to build a new summary variable called activity level. These weights are: −0.34 for resting, +0.29 for being in alleys, and +0.52 for eating. A high value obtained for a cow at a given time (i.e. a specific hour on a specific day) means that she is fairly active; a low value means that she is fairly inactive. We then exploited the time series composed of the activity level of each cow per hour. Fig. 2 provides an example of variations in activity level during a day.

We monitored the ruminal pH of each cow every 15 min during the entire experiment using a specific pH sensor in the cow rumen (Farm bolus; eCow, Exeter, UK). Following the method proposed by Villot et al. (2018), we normalised the ruminal pH values of each cow to take into account inter-individual variability, sensor drift and sensor noise. We considered that a cow was under SARA when the time she spent with a normalised ruminal pH (**NpH**) decreased by at least 0.3 for more than 50 min/d and the daily standard deviation in NpH was above 0.2 or the daily NpH range was above 0.8.

Every day, the caretakers observed the cows and noted any apparent disorders (lameness, low ingestion, diarrhoea, respiratory disorder, colic, injury, isolation, apathy, and oestrus) as well as any potential disturbances in the barn due to e.g. electric failure or a fire alarm.

2.4. Machine learning

The data we used are time series as the values correlate with time. Following a classification approach, a ML model can be built to classify the data from a time series into normal *vs.* abnormal. For instance, Rajpurkar et al. (2017) used a neural network to classify electrocardiograms into normal heart activity *vs.* several types of arrhythmia. Using this approach assumes that the patterns of all possible anomalies are known. This was not the case in our study: although cows' activity follows a circadian rhythm as in other animals (Buijs et al., 2016), each cow has its own rhythm of activity during the day and may react differently to a particular disorder. This implies that we cannot detect a specific pattern of anomaly among the time series. However, we can attempt to predict future values of the series and detect the anomalies by comparing the observed values against predicted values. Such an approach was proposed by Malhotra et al. (2015), then applied by Chauhan and Vig (2015) to detect anomalies in electrocardiograms. The framework proposed by Malhotra et al. (2015) is organized in three steps: (1) ML prediction, (2) Prediction error computing, and (3) Statistical analysis (Fig. 3). In Step 1, the ML algorithm predicts the future values of a given time-window. In Step 2, the error between predicted and observed values is calculated. If there is no anomaly in the window, the system is able to accurately predict the future values. In Step 3, the error between predicted and observed values is compared to a given
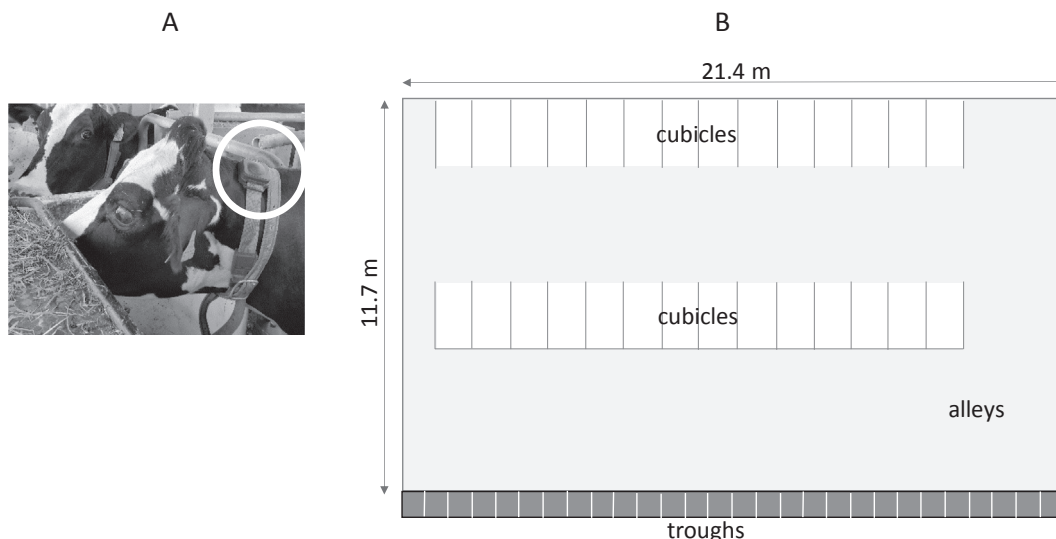
Fig. 1. RTLS-enabled detection of the cow position in the barn. Fig. 1A, RTLS sensor on top of the cow collar; Fig. 1B, layout of the barn.

Table 1

Diets provided to the cows.

Item	Control	SARA inducing diet
Forage / concentrate ratio	75 / 25	54 / 46
Ingredients (% dry matter)		
Hay	19	14
Wrapped hay	56	40
Concentrates for production	25	0
Cereal concentrates[1]	0	46
Chemical composition of diet (g/100 g dry matter)		
Organic matter	93.3	94.7
Starch	10.5	39.4
Fibers (cellulose, lignin, …)[2]	52.1	21.6
Crude protein	17.4	14.2

[1] Barley, corn, and wheat mix.

[2] Neutral Detergent Fibers.

threshold, chosen to optimise the distinction between normal and abnormal values. An anomaly is detected in the time window if the error is statistically higher than the threshold. We applied this framework to detect anomalies in cow activity.

Here, we used 24-h time-windows, because the circadian rhythm is a strong determinant of activity (Buijs et al., 2016). We thus considered 24 h of activity represented by the vector:

$$x = <x_1, \cdots, x_{24}>, x \in \mathbb{R}^{24}$$

and predicted the activity of the future hours as:

$$y^p = <x_{25}^p, x_{26}^p, \cdots, x_{24+l}^p>, y^p \in \mathbb{R}^l$$

where:

$x_1, \cdots, x_{24}$ is the activity level of a cow over the 24 h of a given day

$x_{25}^p, x_{26}^p, \cdots, x_{24+l}^p$ is the activity level predicted for the next day

l is the number of future hours for which cow activity is predicted (from 1 to 12 in our study).

We used the same 24-h sample to train the models whatever the number of hours to be predicted, taking into account only the maximum number of hours to be predicted (here 12, so the sample includes a time windows of 24 observed + 12 h predicted). The sample available from the dataset gets smaller with increasing number of hours to be predicted. Beyond 12 h to be predicted, the models were unstable and the quality of the results decreased. We thus decided to limit the prediction to 12 h from the previous 24 h. Using floating 24-h time-windows, we were nevertheless able to predict 24 h of activity.

In Step 1 of the framework borrowed from Malhotra et al. (2015), we tested several widely-used ML algorithms to predict the activity of the future hours (y^p):

- K Nearest Neighbours for Regression **(KNNR)**, which is a lazy algorithm that gives a prediction based on the average of the *k* nearest
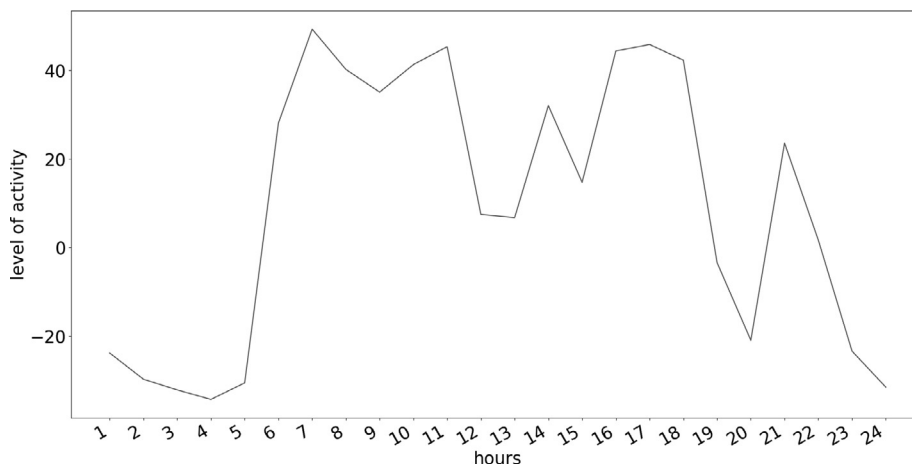

Fig. 2. Changes in activity level of an individual cow during one day (Cow 2008605 from the control group, on March 2, 2015).
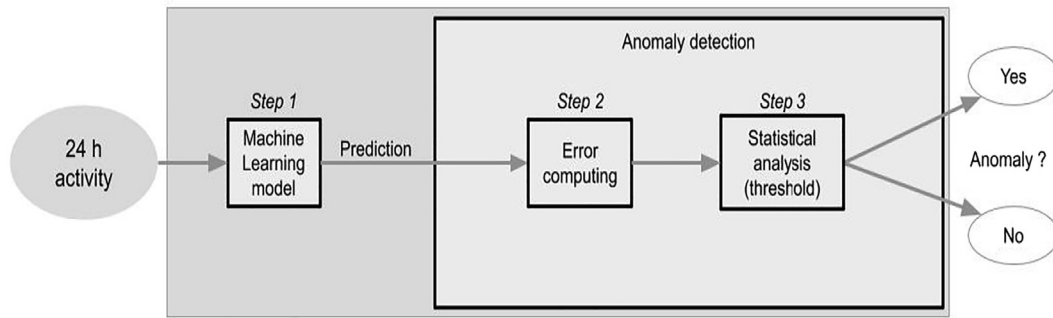
Fig. 3. Framework developed by Malhotra et al. (2015) to detect anomalies in time series.

observations from the dataset (Cover & Hart, 1967). To do so, we screened the dataset to find the k 24 consecutive hours with the minimum Euclidian distance from the observed vector x. We tested k = 1, 10 and 20;

- Decision Tree for Regression (**DTR**), which takes a decision (the prediction) from a particular structure called a decision tree (Quinlan, 1986). Each node of the tree corresponds to a rule, i.e. a condition test on the data. Each possible result of that test leads to a branch, then possibly to a new node with a new rule. The algorithm starts by the first node of the tree, called root, then follows a path through all branches and finishes by a leaf that gives the best prediction;
- MultiLayer Perceptron (**MLP**), which is a basic neural network architecture (Rumelhart et al., 1986) composed of artificial neurons organized in layers. Each neuron from a layer is connected to all neurons from the following layer. A connection consists of a signal that is transmitted from one neuron to another, like synapses in the brain. A neuron collects all signals from the previous layer and applies a non-linear function, the result of which is sent to the following neurons.
- Long Short-Term Memory (**LSTM**), which is a neural network architecture designed for time series (Hochreiter & Schmidhuber, 1997) that takes into account former values in the series. It is the original algorithm used by Malhotra et al. (2015);
- **H-24**, which is a homemade basic algorithm assuming a stable rhythm on consecutive days h, that is the activity on Hour 25 equals that on Hour 1, that activity on Hour 26 equals that on Hour 2, etc.

In Step 2, we calculated error vectors. For each 24-h time-window x, we have l ($l \in [1, 12]$) predicted values and l observed values (vectors y^p and y^o). For error computing, we calculated the prediction error vector:

$$e = y^p - y^o = < x_{25}^p - x_{25}, x_{26}^p - x_{26}, \cdots, x_{24+l}^p - x_{24+l}>$$

with $e \in \mathbb{R}^l$

For the statistical analysis in step 3, we assumed that all computed error vectors follow a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$, with $\mu = [0]^l$ and Σ is the covariance matrix computed by the maximum likelihood estimation method.

A threshold τ was calculated to define a limit of acceptance, i.e. to distinguish a normal *vs.* abnormal activity. It was computed in a way to maximise the difference between True Positive Rate (**TPR**, see below) and the False Positive Rate (**FPR**, see below). If the density of an error vector was below the threshold, we classified the corresponding 24-h time-window as normal. If the error vector density was above the threshold, we considered the corresponding 24-hour time-window as abnormal.

2.5. Dataset

We split our dataset into normal *vs.* abnormal cow*days. To extract

the normal cow*days, we removed the cow*days when a disorder was detected on a cow by the caretakers or via the ruminal pH sensor and the days when a disturbance was recorded affecting the whole barn (e.g. an electricity failure or fire alarm). We also removed two cow*days before a disorder was detected on a cow and seven days after in order to avoid pre-clinical states or the disorder recovery phase. To extract the abnormal cow*days due to SARA, we kept the cow*days on which cows were detected under SARA due to abnormal ruminal pH and the following cow*day, based on the fact that ruminal pH correlates with behavioural symptoms of SARA, i.e. an animal is sick when pH goes down and recovers quickly when pH resumes normal values (Villot et al., 2018). As this study is focused on SARA, we did not use the remaining cow*days, which were days when a disturbance occurred in the barn or days considered abnormal due to another cow disorder (disease or oestrus). We obtained 196 normal cow*days and 265 cow*days when cows were under SARA.

Then we used:

– 70% of the normal cow*days to build the ML models to learn the normal patterns of a cow activity. Algorithms based on a neural network architecture (MLP and LSTM) need to separate training and validation, so we used 80% of the sub-dataset for training and 20% for validation;

– 10% of the normal cow*days and 50% of the SARA cow*days to find the parameters of the multivariate Gaussian error distribution and optimise the threshold τ to distinguish normal *vs.* abnormal days. We used different percentages for normal and SARA cow*days to obtain a similar number of days in the two conditions;

– 20% of the normal cow*days and 50% of the SARA cow*days to compare the performances of the ML models.

All normal and SARA cow*days were chosen at random from the initial dataset.

2.6. Training and comparison of ML models

The training of a MLP and LSTM models is stochastic, i.e. each execution of the training process can bring a different result. We trained the models 10 times to cover for variations in models produced by each training. The results given in this paper are the average of the results from the 10 executions.

We calculated the TPR and FPR obtained with each ML model, with:

– TPR, the proportion of 24-h time-windows labelled as abnormal and detected as abnormal in all 24-h time-windows labelled as abnormal,

– FPR, the proportion of 24-h time-windows labelled as normal and detected as abnormal in all 24-h time-windows labelled as normal.

We also recorded the time needed to perform all three steps of the Malhotra method and the time needed to detect anomalies for each ML

Table 2

Performances of the algorithms tested. The algorithms were: K-Nearest Neighbours for Regression (KNNR) applied to 1, 10 and 20 neighbours; Decision Tree for Regression (DTR); MultiLayer Perceptron (MLP); Long Short-Term Memory (LSTM); and an algorithm called H-24 where activity is assumed similar to the day before. For each algorithm, the activity of the cows was predicted from 1 to 12 h. Performances are expressed as True Positive Rate (TPR) and False Positive Rate (FPR).

l^1		1	2	3	4	5	6	7	8	9	10	11	12
Algorithm													
KNNR1	TPR	0.59	0.51	0.52	0.57	0.57	0.57	0.65	0.68	0.66	0.52	0.66	0.83
	FPR	0.58	0.50	0.49	0.52	0.49	0.46	0.51	0.53	0.51	0.40	0.50	0.66
KNNR10	TPR	0.45	0.45	0.49	0.56	0.54	0.61	0.71	0.70	0.70	0.64	0.60	0.50
	FPR	0.46	0.46	0.46	0.50	0.45	0.48	0.55	0.53	0.53	0.47	0.47	0.45
KNNR20	TPR	0.50	0.62	0.42	0.52	0.56	0.58	0.65	0.63	0.61	0.57	0.41	0.25
	FPR	0.51	0.63	0.39	0.45	0.48	0.46	0.47	0.44	0.44	0.39	0.29	0.21
DTR	TPR	0.41	0.10	0.70	0.30	0.24	0.35	0.20	0.25	0.30	0.34	0.11	0.24
	FPR	0.42	0.11	0.71	0.31	0.23	0.32	0.17	0.23	0.27	0.30	0.10	0.26
MLP	TPR	0.51	0.68	0.59	0.55	0.65	0.63	0.63	0.50	0.55	0.41	0.34	0.16
	FPR	0.49	0.66	0.53	0.44	0.51	0.49	0.48	0.37	0.42	0.31	0.27	0.16
LSTM	TPR	0.21	0.24	0.24	0.32	0.35	0.33	0.28	0.35	0.33	0.26	0.14	0.78
	FPR	0.21	0.22	0.23	0.28	0.24	0.23	0.22	0.30	0.26	0.19	0.08	0.72
H-24	TPR	0.78	0.78	0.78	0.77	0.77	0.77	0.77	0.62	0.60	0.48	0.48	0.48
	FPR	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.61	0.58	0.46	0.46	0.46

[1] Number of hours for which we predict the cows' activity.

model.

3. Results

Table 2 gives the TPR and FPR to detect behavioural anomalies due to SARA for all algorithms tested and the number of hours for which we predict cow activity.

The performances of the algorithms tested for detecting SARA ranged from 0.11 to 0.83 in terms of TPR and from 0.75 to 0.08 in terms of FPR. We obtained the highest TPR with the KNNR1 algorithm and the prediction of 12 h of cow activity. In general, FPR increased with TPR, so that the difference between TPR and FPR remained low. KNNR algorithms yielded the largest difference between TPR and FPR: 0.18 for KNNR20 with 7–10 h predicted. KNNR1 with 12 h prediction produced the best compromise between a high TPR and a large difference between TPR and FPR (TPR, 0.83; TPR – FPR, 0.17).

We obtained the lowest performances with the H-24 and DTR algorithms. Both these algorithms had very similar TPR and FPR values. The algorithms based on a neural network architecture (MLP and LSTM) provided intermediate results: the TPR reached values slightly lower than with the KNNR algorithms but the difference between TPR and FPR was lower than with KNNR algorithms (largest difference was 0.15 for MLP with 7 h predicted and 0.11 for LSTM with 5 or 6 h predicted).

The KNNR models were fairly quick to build (between 51 and 84 s) and the detection time was 35 s for KNNR1 and KNNR10 and 39 s for KNNR20, whatever the number of hours predicted. The other algorithms took a bit less time to be built (36–51 s) and the detection took only 1 s (Table 3).

4. Discussion

The main findings of this study are that the activity rhythm of a cow varies from one day to another, but it is nevertheless possible to predict series of 24 h of activity and detect significant variations (anomalies) in them using ML. Depending on the algorithm used, these anomalies coincide more or less with the occurrence of a disease (here, SARA).

We obtained the best results with the KNN1 algorithm that predicts individual cow activity based on the closest (i.e. most similar) observations from the initial dataset, and with 12 h of prediction: 83% cases of SARA were detected. In this study, the gold standard for detecting SARA was ruminal pH. However, not all cows suffer from a low ruminal pH. When sheep are submitted to short but repeated episodes of SARA, they change their behaviour on the first occurrence of SARA (they are more agitated and more aggressive to each other) but these

reactions fade thereafter (Commun et al., 2012). The 17% of SARA cases that we did not detect through abnormal behaviour may correspond to cows undergoing SARA without showing signs of suffering from it.

At the same time, with the KNN1 algorithm and 12 h of prediction, 66% of the normal cow*days were also detected as abnormal, giving a TPR-to-FPR difference of only 0.17. The other algorithms tested did not provide better results in terms of FPR or TPR-to-FPR difference. We thus conclude that even though the method is fairly sensitive, it still lacks specificity, which limits its practicability for helping farmers detect animals that need attention, as farmers may give up responding to alerts provided by the system due to too many false-positives.

The lack of specificity may come from the fact that the rhythm of activity is highly variable from one day to another, making it difficult to identify an abnormal rhythm among variations of the rhythm. Indeed, we obtained the worst performances with an algorithm that assumes a stable rhythm from one day to another (named H-24). More recent methods of ML, such as Residual Networks used by Yadav and Bist to identify flies activity (2019) should be investigated to try to improve the specificity of our detection. Furthermore, we applied the various ML algorithms to describe rhythm of activity at group level, i.e. where the rhythm considered as normal is the same for all cows. By contrast, each cow may have its own rhythm of activity. Building models at individual level could probably increase the precision in the distinction between normal vs. abnormal activity, but would require larger datasets. Here we observed 28 cows for 9 weeks, whereas to obtain the same precision in results at cow level, we would need 28*9 weeks, i.e. nearly 5 years of observations, before being able to detect anomalies in the next years, which is clearly not compatible with the objective of routinely detecting diseases in a herd.

Cattle behaviour generally follows a well-marked circadian rhythm (Veissier et al., 1989a; Veissier et al., 2001; Nikkhah, 2014). Cattle are generally active during the day and rest at night. In addition, when kept indoors, dairy cattle display peaks of activity at the time food is delivered and around milking. Here we considered 24-h time-windows but we did not look at specific periodic components of the circadian rhythm. Extracting such components could improve the accuracy of the modelling of the 24 h time-windows and may allow models to be built at animal level rather than group level.

This study used SARA, a metabolic disease, as disease model. SARA may not affect animal behaviour as much as other diseases that induce inflammation, fever, or pain (e.g. in cattle, Borderas et al., 2009; De Boyer Des Roches et al., 2017). Further exploration of behavioural changes under such diseases is necessary in order to make further progress in using ML to detect sickness behaviour early. This would

Table 3

Calculation time for the algorithms tested. The algorithms were: K-Nearest Neighbours for Regression (KNNR) applied to 1, 10 and 20 neighbours; Decision Tree for Regression (DTR); MultiLayer Perceptron (MLP); Long Short-Term Memory (LSTM); and an algorithm called H-24 where assumed is assumed similar to the day before. For each algorithm, the activity of the cows was predicted from 1 to 12 h. We distinguish the time needed to build the model (including training and validating the model and calculation of the anomaly detection threshold) and the time needed to detect anomalies. All times are given in s.

l[1]		1	2	3	4	5	6	7	8	9	10	11	12
Algorithm													
KNNR1	Model building	74	84	82	79	77	73	70	66	65	61	59	54
	Detection	35	35	35	35	35	35	35	35	35	35	35	35
KNNR10	Model building	73	84	83	78	78	73	71	67	65	61	59	54
	Detection	35	35	35	35	35	35	35	35	35	35	35	35
KNNR20	Model building	57	66	64	62	60	57	55	53	51	48	46	43
	Detection	39	39	39	39	39	39	39	39	39	39	39	39
DTR	Model building	51	58	57	54	52	50	48	45	43	41	39	36
	Detection	1	1	1	1	1	1	1	1	1	1	1	1
MLP	Model building	53	62	60	58	56	55	53	51	50	48	46	44
	Detection	1	1	1	1	1	1	1	1	1	1	1	1
LSTM	Model building	59	68	69	68	69	70	71	71	69	70	70	71
	Detection	1	1	1	1	1	1	1	1	1	1	1	2
H-24	Model building	51	60	61	60	61	60	60	61	62	61	61	60
	Detection	1	1	1	1	1	1	1	1	1	1	1	1

[1] Number of hours for which we predict the cows' activity.

necessitate inducing either an infectious disease or at least inflammation via e.g. LPS injection, or monitoring all diseases occurring spontaneously in a herd over a large period.

Finally, the time needed to build the models was fairly quick, with the longest being 84 s for KNNR with 1 neighbour and 12 h of prediction. Detection was also quick, at 35–39 s for KNNR and just 1 s for the other algorithms. These times are compatible with a use of ML in real-farm practice, allowing fast detection of anomalies. However, the sample used in our study was relatively small (28 cows). The time to build the models may be far longer in large herds when a KNNR algorithm is used, because for each 24-h window analysed, the algorithm has to look for the nearest neighbour among all data. The practicality of KNNR algorithms thus needs checking on large herds.

In conclusion, Machine Learning helps analyse the activity rhythm of animals. It is possible to detect anomalies in the activity rhythm that are possibly related to health disorders (in our study, an experimentally-induced episode of SARA). However, the current method lacks specificity for detecting health disorders. Investigations should be pursued on larger datasets, possibly from routine observation of diseases and other artefacts in the animal's life, with modelling applied at animal level and taking into account the circadian nature of the activity rhythm.

CRediT authorship contribution statement

Nicolas Wagner: Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Violaine Antoine:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration. **Marie-Madeleine Mialon:** Formal analysis, Investigation, Data curation, Writing - review & editing, Project administration. **Romain Lardy:** Conceptualization, Writing - review & editing, Project administration. **Mathieu Silberberg:** Investigation, Data curation, Writing - original draft, Writing - review & editing. **Jonas Koko:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Isabelle Veissier:** Conceptualization, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, 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.

References

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.

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.

Borderas, T.F., Rushen, J., von Keyserlingk, M.A.G., de Passillé, A.M.B., 2009. Automated measurement of changes in feeding behavior of milk-fed calves associated with illness. J. Dairy Sci. 92, 4549–4554.

Buijs, F.N., León-Mercado, L., Guzmán-Ruiz, M., Guerrero-Vargas, N.N., Romo-Nava, F., Buijs, R.M., 2016. The circadian system: a regulatory feedback network of periphery and brain. Physiology 31, 170–181.

Byrd, C., Lay, D., 2018. Can baseline heart rate variability be used as a predictor of the swine behavioral and febrile response to a sickness challenge? J. Anim. Sci 96, 9.

Chauhan, S., Vig, L., 2015. Anomaly detection in ECG time signals via deep long-term memory networks. In: Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (Ieee Dsaa 2015), 834–840.

Commun, L., Silberberg, M., Mialon, M.M., Martin, C., Veissier, I., 2012. Behavioural adaptations of sheep to repeated acidosis challenges and effect of yeast supplementation. Animal 6, 2011–2022.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27.

Dantzer, R., Kelley, K.W., 2007. Twenty years of research on cytokine-induced sickness behavior. Brain. Behav. Immun. 21, 153–160.

Dawkins, M.S., Cain, R., Roberts, S.J., 2012. Optical flow, flock behaviour and chicken welfare. Anim. Behav. 84, 219–223.

De Boyer Des Roches, A., Faure, M., Lussert, A., Herry, V., Rainard, P., Durand, D., Foucras, G., 2017. Behavioral and patho-physiological response as possible signs of pain in dairy cows during Escherichia con mastitis: A pilot study. J. Dairy Sci. 100, 8385–8397.

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.

Hart, B.L., 1988. Biological basis of the behavior of sick animals. Neurosci. Biobehav. Rev. 12, 123–137.

Higaki, S., Moiura, R., Suda, T., Andersson, I.M., Okada, H., Zhang, Y., Itoh, T., Miwalkeichi, F., Yoshioka, K., 2018. Estrous detection by continuous measurements of vaginal temperature and conductivity with supervised machine learning in cattle.

Theriogenology 123, 90–99.

Hochreiter, S., Schmidhuber, J., 1997. long short-term memory. Neural Comput. 9, 1735–1780.

Hong, W., Kennedy, A., Burgos-Artizzu, X.P., Zelikowsky, M., Navonne, S.G., Perona, P., Anderson, D.J., 2015. Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. Proc. Natl. Acad. Sci. 112, E5351–E5360.

Lopes, P.C., Block, P., Pontiggia, A., Lindholm, A.K., König, B., 2018. No evidence for kin protection in the expression of sickness behaviors in house mice. Sci. Rep. 8, 16682.

Malhotra, P., Vig, L., Shroff, G., Agarwal, P., 2015. Long short-term memory networks for anomaly detection in time series, ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Available from http://www.i6doc.com/en/. Presses universitaires de Louvain, Bruges (Belgium), pp. 89–94.

Nikkhah, A., 2014. Time of feeding orchestrates circadian post-feeding intake patterns in dairy cows. Cattle Pract. 22, 92–94.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81–106.

Rajpurkar, P., Hannun, A.Y., Haghpanahi, M., Bourn, C., Ng, A.Y., 2017. Cardiologist-level arrhythmia detection with convolutional neural networks. ArXiv Preprint ArXiv 1707.01836.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.

Shakhar, K., Shakhar, G., 2015. Why do we feel sick when infected—can altruism play a role? PLoS Biol. 13, e1002276.

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

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

Veissier, I., Boissy, A., de Passillé, A.M., Rushen, J., van Reenen, C.G., Roussel, S., Andanson, S., Pradel, P., 2001. Calves' responses to repeated social regrouping and relocation. J. Anim. Sci. 79, 2580–2593.

Veissier, I., Le Neindre, P., Trillat, G., 1989a. Adaptability of calves during weaning. Biologie du Comportement (Biology of Behaviour) 14, 66–87.

Veissier, I., Le Neindre, P., Trillat, G., 1989b. The use of circadian behaviour to measure adaptation of calves to changes in their environment. Appl. Anim. Behav. Sci. 22, 1–12.

Veissier, I., Mialon, M.-M., Sloth, K.H., 2017. Short communication: Early modification of the circadian organization of cow activity in relation to disease or estrus. J. Dairy Sci. 100, 3969–3974.

Villot, C., Meunier, B., Bodin, J., Martin, C., Silberberg, M., 2018. Relative reticulo-rumen pH indicators for subacute ruminal acidosis detection in dairy cows. Animal 12, 481–490.

Yadav, S., Bist, A.S., 2019. Residual nets for understanding animal behavior. J. Anim. Behav. Biometeorol. 7, 97–103.