Research article

The new design of cows' behavior classifier based on acceleration data and proposed feature set

Phung Cong Phi Khanh[1], Duc-Tan Tran[2], Van Tu Duong[3], Nguyen Hong Thinh[1] and Duc-Nghia Tran[4,*]

[1] VNU University of Engineering and Technology, 144 Xuan Thuy, Hanoi City, Vietnam
[2] Faculty of Electrical and Electronic Engineering, Phenikaa University, Hanoi City, Vietnam
[3] NTT Hi-Tech Institute–Nguyen Tat Thanh University, 300A Nguyen Tat Thanh Street, Ward 13, District 4, Ho Chi Minh City, Viet Nam
[4] Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi City, Vietnam

* **Correspondence:** Email: nghiatd@ioit.ac.vn; Tel: +84936866335.

Abstract: Monitor and classify behavioral activities in cows is a helpful support solution for livestock based on the analysis of data from sensors attached to the animal. Accelerometers are particularly suited for monitoring cow behaviors due to small size, lightweight and high accuracy. Nevertheless, the interpretation of the data collected by such sensors when characterizing the type of behaviors still brings major challenges to developers, related to activity complexity (i.e., certain behaviors contain similar gestures). This paper presents a new design of cows' behavior classifier based on acceleration data and proposed feature set. Analysis of cow acceleration data is used to extract features for classification using machine learning algorithms. We found that with 5 features (mean, standard deviation, root mean square, median, range) and 16-second window of data (1 sample/second), classification of seven cow behaviors (including feeding, lying, standing, lying down, standing up, normal walking, active walking) achieved the overall highest performance. We validated the results with acceleration data from a public source. Performance of our proposed classifier was evaluated and compared to existing ones in terms of the sensitivity, the accuracy, the positive predictive value, and the negative predictive value.

Keywords: cow; monitoring; acceleration; sensor; classification

1. Introduction

Commercial dairy farms face major challenges in monitoring and maintaining cow well-being and comfort, which related directly quantities of dairy produce. It is difficult for medium-to-large size farms to monitor their herds through observation, resulting in a financial loss for the farm.

Dairy cows are high value farm animals requiring careful management since they are particularly susceptible to health problems. When they have health problems or physiological conditions, different behavior will be exhibited [1]. Cows alter their behavior to enable them to deal with stressors such as infection, satiety, or social and environmental changes [2]. So, behavior is an indicator of health and well-being dairy cows, detection of changes in cow behavior on a daily basis can support to provide alerts to execute specific management tasks that improve dairy farm management [3–8].

Over the past decade, there has been a huge increase in the use of remote monitoring devices such as global positioning (GPS) trackers, location sensors and accelerometers for automated recording of animal behavior [9]. In [10], the classification of bull behaviors is designed and implemented using video data from assembling cameras. Behavior events of interest in this study included lying, standing, walking, and mounting. In [11] the authors describe the potential benefits and challenges of remotely monitoring cattle behavior with various kinds of methodologies, including clinical illness scores, visual monitoring, accelerometers, pedometers, feed intake and behavioral monitoring, global position systems, and real-time location systems.

Recent advances in sensor technology of electronic devices that allow high sensitivity and provide new scenarios for recording cow activities [12]. Some existing systems based on sensor technology have been developed for automatic dairy cow behavior analysis [13,14]. With the advantages of being small size, light weight and low power consumption, accelerometers provide a noninvasive and objective method of measuring cow behavior under farm conditions [9,15–22].

The approach based on accelerometer sensor combine with has a greater emphasis on individual well-being and performance rather than a more traditional herd based approach. Nevertheless, the interpretation of the data collected by such sensors when characterizing the type of behaviors still brings serious challenges to developers, related to activity complexity (i.e., certain behaviors contain similar gestures), the extraction of relevant features that allow to differentiate the behavior, to the data loss that characterizes any wireless transmitter, and the complex data processing required to deal with the noise inherent in the collected measurements [23]. This has led to the need for more efficient and accurate methods of analyzing the vast amounts of movement and behavioral data that are being collected [24].

The complex problem of recognizing animal behavior has motivated different groups of researchers to benchmark different real-world scenarios with wearable sensing solutions [13,25,26]. Machine learning provides an excellent approach to improve model accuracy, based on data structures that might dynamically change, while dealing with complex and large datasets acquired from a particular environment [27]. For example, Martiskainen et al. [28] developed a method that uses acceleration data and multi-class support vector machines (SVM) to automatically classify several behaviors in dairy cows. In a similar study, Diosdado et al. [9] implemented a decision-tree algorithm to classify different cattle behaviors. Arcidiacono et al. [19] computed an acceleration threshold to classify the feeding and standing activities of dairy cows in a free-stall barn. Their algorithm also estimated the number of steps of each cow from the acceleration data by making use

of statistically defined thresholds. In their next study [29], these authors showed that their new approach is possible to be implemented in the real-time manner due to the low sampling frequency (i.e. 4 Hz) and low complexity. They also explained two main reasons for the misclassification between feeding and standing is that: 1) when the cow rotated its head during the feeding activity, the sensor's position was deflected; 2) in some cases, even the cow was standing, its head was still down. Finally, some suggestions are given in order to improve both the classification performance and the real-time implementation.

In [20], recently, Jun Wang et al. proposed a Multi-BPAda Boost classification algorithm that allows classifying seven cow behaviors (feeding, lying, standing, lying down, standing up, normal walking, active walking) from data of three-dimensional accelerometers.

Although the works of [9,19,20,28] demonstrated the potential of using machine learning in classification of cow behavior, they only classify two or three behaviors [9,19], the positive predictive value of classification is not high [28] and lack of analysis about features of data [20,28]. As a matter of fact, the performance of this approach is rigorous dependence on features and window (related to the number of samples of a record) of data.

The main objective of this study is the development and evaluation of a new method based on our proposed feature set and window of data for characterizing leg-mounted acceleration data to improve the performance of cow behavior classification. This paper focused on classifying cow behavior into seven categories (feeding, lying, standing, lying down, standing up, normal walking, active walking). We proposed 5 features (mean, standard deviation, root mean square, median, range), 16-second window of data (1 sample/second), and a Gradient Boosted Decision Tree (GBDT) algorithm for classification. We evaluated our method and compared to Jun Wang et al. with the same dataset [20]. Furthermore, we also made comparisons to the work of P. Martiskainen and M. Jarvinen [28] which based on collar-mounted acceleration data.

2.   Materials and methods

2.1. The breeding environment and animals

The five Holstein dairy cows (Figure 1c) from a free-stall barn farm (Nanyang, Henan Province, China) were chosen for the trial on the basis of similar body size, all of them were in the early lactation stage. The cows were in the separation area in the free-stall barn which had a rectangular layout of 180 × 31 m. According to Jun Wang et al. [20], the barn included a feeding passage, two rows of self-locking headlocks, and two rows of head to head stalls arranged with sand beds. The roof was covered with light-weighted color steel plates with the symmetric structure with a 1:3 slope. The height of the barn and the eaves was 10 and 4.65 m, respectively. The cows were loose-housed in the studied area which was located in the middle of the barn and separated by fences. The internal facilities included a watering trough, a row of self-locking headlocks, and seven groups of head to headstalls. See Figure 1a,b for plan and section of the studied area in the barn.

Cows were milked twice a day with a fish-bone milking machine, floors were cleaned every day with a scraper blade. Cows were fed the total mixed ration (TMR) diet. They were healthy without any signs of serious lameness or other diseases that would affect their behavior.
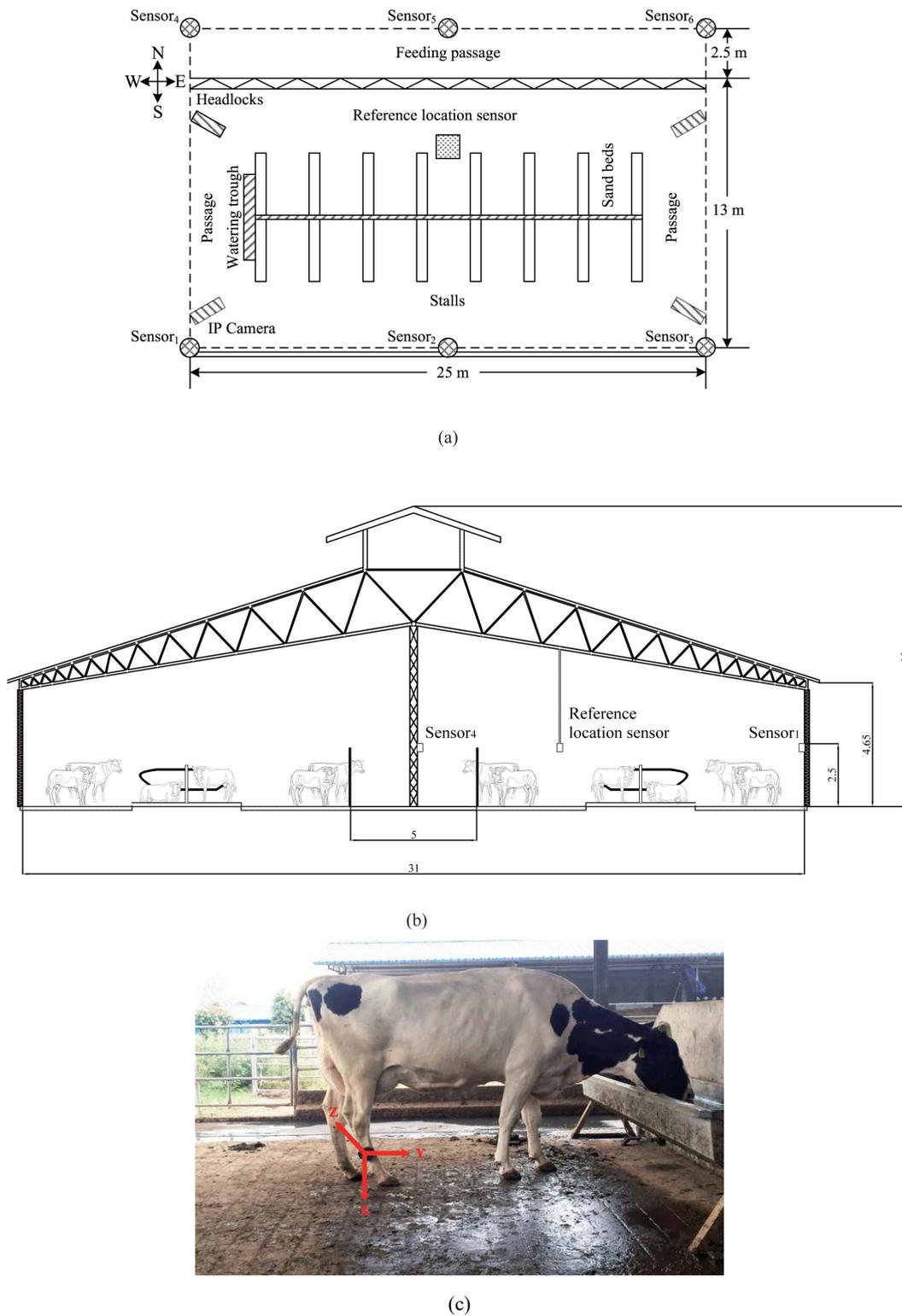
(a)

(b)

(c)

Figure 1. Plan and section of the studied area in the barn [20]. (a) Plan. (b) Section. Sensor is attached to the leg of a cow. (c) Copyright: © 2018 Wang et al. [20].

2.2. The acceleration sensor

The acceleration sensor can be attached to the leg or the neck of a cow (see an example in

Figure 1c). The aim of the sensor is to determine acceleration of the leg and the neck. A 3-DOF (degrees of freedom) acceleration sensor (with a selectable measurement range of $\pm$ 2 g, $\pm$ 4 g, $\pm$ 8 g, or $\pm$ 16 g with g = 9.81 (m/s^2) that measures acceleration value in 3 directions (X, Y, Z) can be used to measure both dynamic and static accelerations. Accordingly, acceleration can be calculated using equation as follows:

$$A_k = (\text{Sam}_i / 1024 \times R - O_i) / S_i, \tag{1}$$

where A_k is the acceleration value in direction k $(k = X, Y, Z)$:

Sam_i is the value after sampling of axis k;

R is the voltage reference; O_i is the offset;

S_i is the sensitivity of the accelerometer on axis k.

Bandwidths of the accelerometer can be configured within a range of 0.5 to 1600 Hz for X and Y axes, and a range of 0.5 to 550 Hz for the Z axis [30].

In this study, we only focused on the acceleration data taken from the leg.

2.3. Machine learning algorithms for cow's behaviors classification

Our method for estimating cow behavior utilizes 5 features of acceleration data, such as mean, standard deviation (SD), root mean square (RMS), median, range, as features for classification. We proposed a window selection and feature extraction scheme for the recognition process. The recognition process started with the features extraction of acceleration data. We used a fixed-width sliding window of n-second (record i has n samples, including n-6 last samples of record i-1) on data of each behavior. From each window, a vector of 5 features was obtained. The explanation of these choices in section 2.5. Finally, these patterns were used as input of the trained classifier for the recognition of behaviors. The entire behavior recognition process is shown in Figure 2.
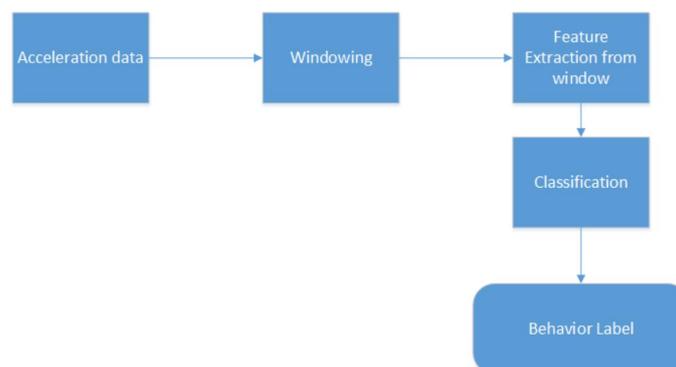

Figure 2. Behavior recognition process pipeline.

Figure 3 outlines the constructive process of the classifier in this study. 60% of the data were randomly selected as the training data set, and the remaining 40% were used as the test data set. The classifier was trained with the features calculated from the acceleration measurements (input) and the matching behavior (output). Resulting classifier model was tested on and the model performance

indicators were calculated from the independent test data set. A program written in Python 3.5 was developed that applies various supervised learning algorithms for testing the effectiveness of the proposed method.
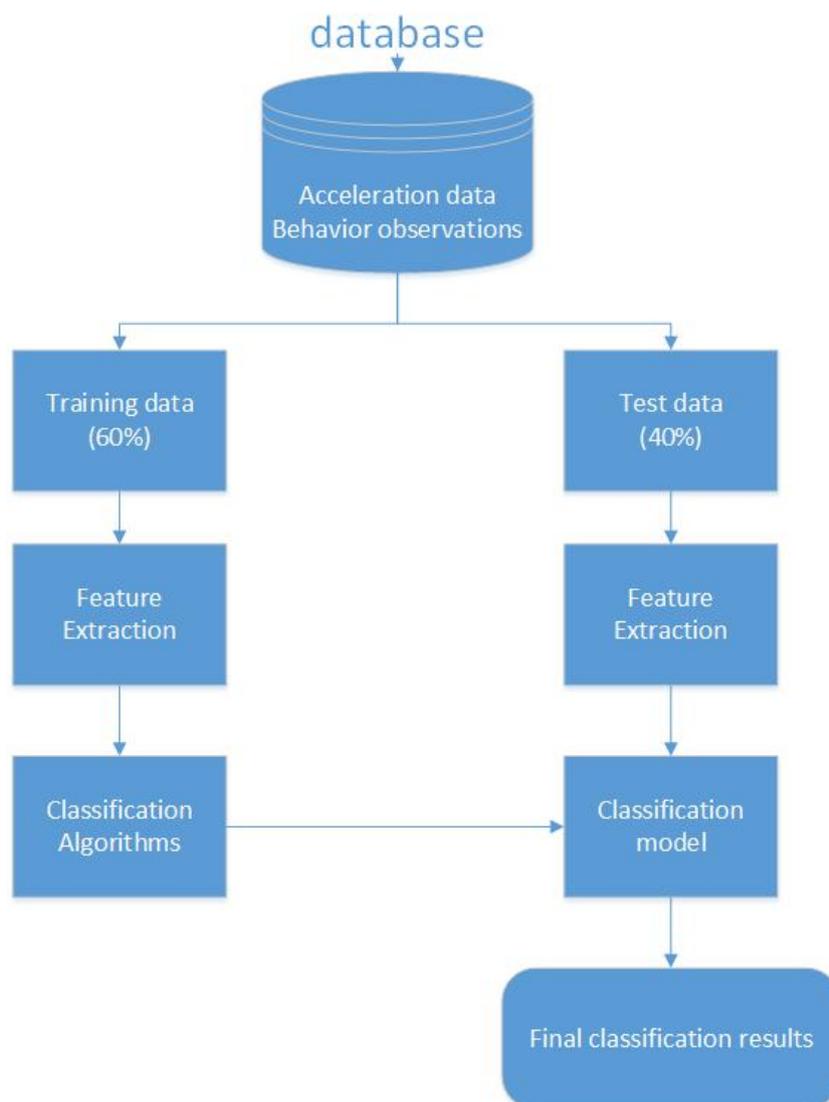

Figure 3. Flow chart of constructive process of the classifier.

We would like to assess the classification performance of 4 individual machine learning algorithms (Gradient Boosted Decision Tree, Support Vector Machines, Random Forest and K-Nearest Neighbor) with different windows (6-second, 12-second, 16-second, and 20-second) (section 2.5) and 5 features (mean, median, standard deviation, root mean square and range). After that, we analyzed the detailed results of the best combination. The performance of the model was evaluated based on four indicators (section 2.6). Finally, we made comparisons in term of overall performance to the works of Jun Wang et al. [20] and Martiskainen and M. Jarvinen [28].

2.4. Data used in this research

We addressed the problem of classifying cow behavior by defining a learning framework based on the study of acceleration data characteristics. Our framework was experimentally validated on data extracted from the online public dataset in Jun Wang et al.'s work [20]. In particular, we analyzed data acquired from leg tag sensors attached to 5 healthy dairy cows, as they recorded in this dataset. This dataset contains 3685 records recorded by Jun Wang et al. [20]. It is publicly available at https://doi.org/10.1371/journal.pone.0203546.t006. The reason we chose this dataset is that the author provided the raw data of acceleration. Performance of our proposed classifier was latter evaluated and compared to Jun Wang et al.'s work in term of the sensitivity, the accuracy, and the positive predictive value.
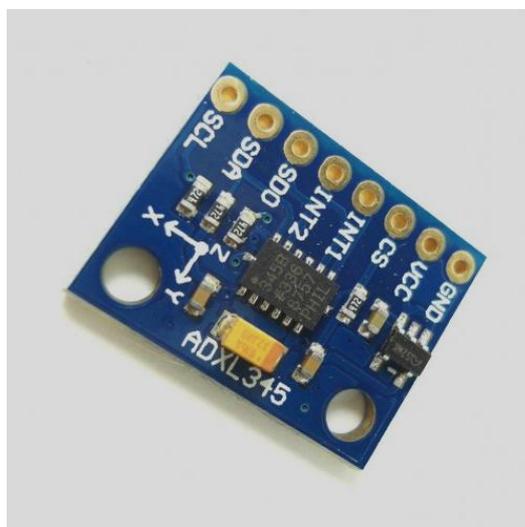

Figure 4. Acceleration sensor ADXL345.

According to Jun Wang et al.'s project's technical description [20], the leg tag sensor used was a three-dimensional accelerometer (ADXL345, Analog Devices Inc., USA) (Figure 4). The accelerometer used to obtain acceleration data in X, Y, Z dimensions, each has a range of ± 8 g and a sampling frequency of 1 Hz. It integrates a 12 bit A/D converter to change the analogue voltage into digital data [20]. Records were labelled according to seven primitive classes, including: feeding, lying, standing, lying down, standing up, normal walking, active walking. An example of acceleration data is shown in Table 1. Table 2 summaries the behavior's definition in this dataset.

Jun Wang et al. verified no restriction or influence of the leg tag on the cows' activities. There was no obvious difference in the behavior before and after installing the leg tag.

Jun Wang et al. used a video recording system in the barn to verify and classify the data. The behaviors shown in the video images were compared with the acceleration data acquired from the sensor system via vision processing software. The video images with the leg tag data were synchronized in order to match the video analysis with the acceleration data for each cow. So, when the activities for all cows from video logs were determined, the data were classified.

Table 1. A record observation of feeding (6 samples/record).

Feeding		
Acc in x (g)	Acc in y (g)	Acc in z (g)
−0.8	0.2	0.2
−0.9	0.1	0.2
−0.9	0	0.3
−0.8	0.1	0.4
−1.1	−0.2	0.5
−0.8	0.1	0.3

Table 2. Description of the behaviors (adapted from [20]).

Behavior category	Definition
Feeding	The cow is at feeding zone and searches for or masticate the feed.
Lying	The cow is in a cubicle in a lying down position.
Standing	The cow stands entirely on its four legs.
Lying down	The cow bents one foreleg, lowers its forequarters, then hindquarters, settles down in a state of lying.
Standing up	The cow rises from a lying state to stand on all four feet.
Normal walking	The activity characterized by at least 3 consecutive limb movements (a progressive step within the 1 s video period).
Active walking	The cow walks forward quickly with long strides (two progressive steps within the 1 s video period).

The dataset in [20] contains only those observations lasting in 6 s. If the length of the time window was too short, the differences between the measurements for each behavior were not obvious or significant [31]. By increasing the time window to 6 s, the measurements effectively contained the whole process of all behavior activities to ensure the integrity of behavior data [20]. In this paper, we customized this dataset following our proposed dataset (see section 2.3), a labeled observation in the dataset contained 16 samples of acceleration data, each recorded in three perpendicular axes with the sampling rate of 1 Hz. Table 3 shows an example of a record of our proposed dataset, and Table 4 shows composition of behavior observations.

To sum up, we attempted to recognized cow's behaviors using a single tri-axial accelerometer worn on the leg of the cow. Behavior recognition was formulated as a classification problem. Data generated by the accelerometer were used to train a set of classifiers, which included K-Nearest Neighbors, Support Vector Machine, Random Forest and Gradient Boosted Decision Tree found in the Scikit-learn library[1].

---

[1]Scikit-learn is an open source Python library that has powerful tools for data analysis and data mining.

Table 3. A record observation of feeding (16 samples/record).

Feeding		
Acc in X (g)	Acc in Y (g)	Acc in Z (g)
−0.8	0.2	0.2
−0.9	0.1	0.2
−0.9	0	0.3
−0.8	0.1	0.4
−1.1	−0.2	0.5
−0.8	0.1	0.3
−0.8	0.1	0.3
−0.8	0.1	0.3
−0.8	0.1	0.3
−0.8	0.1	0.3
−0.8	0.1	0.3
−0.8	0.1	0.3
−0.8	0.1	0.3
−0.8	0.1	0.3
−0.8	0.1	0.3
−0.8	0.1	0.3

Table 4. Composition of behavior observations (16 samples/record).

Behavior pattern	Number of observations
Feeding	613
Lying	731
Standing	451
Lying down	326
Standing up	303
Normal walking	738
Active walking	516
Total	3678

2.5. Feature extraction for acceleration data from a leg-mounted

The features and window of data are critical for the performance of classification algorithms. The features of the data are based on temporal window. To find the best window length, we made some comparisons with the human activities recognition (HAR) problem where the acceleration signal is sampled at a much higher sampling rate (at e.g., 50 Hz) than in our cow dataset (at 1 Hz). In the context of the HAR problem, mid-sized windows (from 5 to 7 s long) perform best from a range of windows from 1 to 15 s for wrist-placed accelerometers [32]. Because dairy cows are less active

than human, we could use a longer window length for our problem.

2.5.1. Feature selection

In this machine learning study, a feature is an individual measurable property of acceleration data. Design informative, discriminating and independent features are a crucial step for effective algorithms in classification. The new contribution of the paper is that we proposed an efficient set of features for classifying seven cow behaviors. With the good features, the machine learning algorithms will find what we are looking for. For example, the mean (one of our features) of acceleration frame is clearly different between Lying and Feeding (Tables 5–7).

Tables 5–7 represent the statistical measurements on seven portions of each class. In particular, the calculated features (mean, SD, RMS, median, range) of each data of static states (feeding, lying and standing) and dynamic states (lying down, standing up, normal walking, active walking) in X, Y and Z dimension were presented.

Table 5. Acceleration data in X dimension.

	Feeding	Lying	Standing	Lying down	Standing up	Normal walking	Active walking
mean	−0.81	−0.04	−0.83	−0.02	0.01	−0.01	−0.02
SD	0.06	0.05	0.07	1.01	0.97	0.50	1.49
RMS	0.82	0.07	0.83	1.01	0.97	0.50	1.49
median	−0.8	0	−0.8	0	0	0	0
range	2	1.8	1.7	7.3	6.5	3.5	10.7

Table 6. Acceleration data in Y dimension.

	Feeding	Lying	Standing	Lying down	Standing up	Normal walking	Active walking
mean	0.01	0.24	0.04	0.02	0.03	0.01	−0.01
SD	0.12	0.10	0.12	1.00	1.01	0.50	1.53
RMS	0.12	0.26	0.12	1.00	1.01	0.50	1.52
median	0	0.3	0	0	0	0	0
range	3.3	1.1	2.1	6.8	6.8	3.3	12.3

Table 7. Acceleration data in Z dimension.

	Feeding	Lying	Standing	Lying down	Standing up	Normal walking	Active walking
mean	0.31	−0.62	0.29	−0.02	0.00	−0.01	−0.04
SD	0.08	0.04	0.08	1.00	1.02	0.50	1.51
RMS	0.32	0.62	0.30	1.00	1.02	0.50	1.51
median	0.3	−0.6	0.3	0	0	0	0
range	1.2	0.4	1.2	6.7	7	4	10.4

It can be shown from Tables 5–7, in each dimension, the dynamic states have wider ranges (up

to 12.3 in the case of active walking) of value compared to the static states (less than 3.3). Thus, *range* can be chosen as a feature to capture the difference between the highest and lowest value between those states.

$$range(X_j) = \left[\min_{i=1}^{N} \{x_i\}, \max_{i=1}^{N} \{x_i\} \right] \tag{2}$$

where X is the data of X-axis;

X_j is the record j;

N is the number of samples of a record (a window size);

x_i is the sample i of record X_j;

min is the minimum value of X_j;

max is the maximum value of X_j;

range(X_j) is variation between the minimum value and the maximum value of X_j.

Static states data tend to have a different typical value. The *mean* can refer to measurements of central tendency and *median* which gives a measurement of typical value as features for distinguishing static states from dynamic states and also for the classification of static states from each other. So, the mean is a good feature.

$$m(X_j) = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{3}$$

$$median(X_j) = \frac{x_{[\#N/2]} + x_{[\#N/2+1]}}{2} \tag{4}$$

where $m(X_j)$ is mean of X_j
for (4), x_i values in X_j is sorted.

Between the four dynamic states, normal walking data (SD about 0.5) tend to be more concentrated than the others (SD more than 1). Active walking data (SD about 1.5) have the largest range of value and also are the most sparse data. So, the standard deviation captured those differences between these states.

$$\sigma(X_j) = \sqrt{\frac{1}{N} \sum_{i=1}^{k} (x_i - m)^2} \tag{5}$$

where $\sigma(X_j)$ is standard deviation of X_j.

Lying down data have more non zero values than standing up data. This leads us to the usage of RMS to measure this characteristic.

$$RMS_{X_j} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2} \tag{6}$$

where RMS_{X_j} is root mean square of X_j.

Note that all the formulae (2–6) are for X-axis; the formulae for Y-axis, Z-axis are similar.

2.5.2. Verification of features

We plotted the data in the training set (section 2.4) to give an insight into data distribution. t-SNE[2] tool allows displaying multidimensional data on 2-dimension space while ensuring distance characteristic between data points. Figure 5 shows the result when using the t-SNE tool for mapping each data original point (X, Y, Z) in 3-Dimension space into a 2-Dimension space.
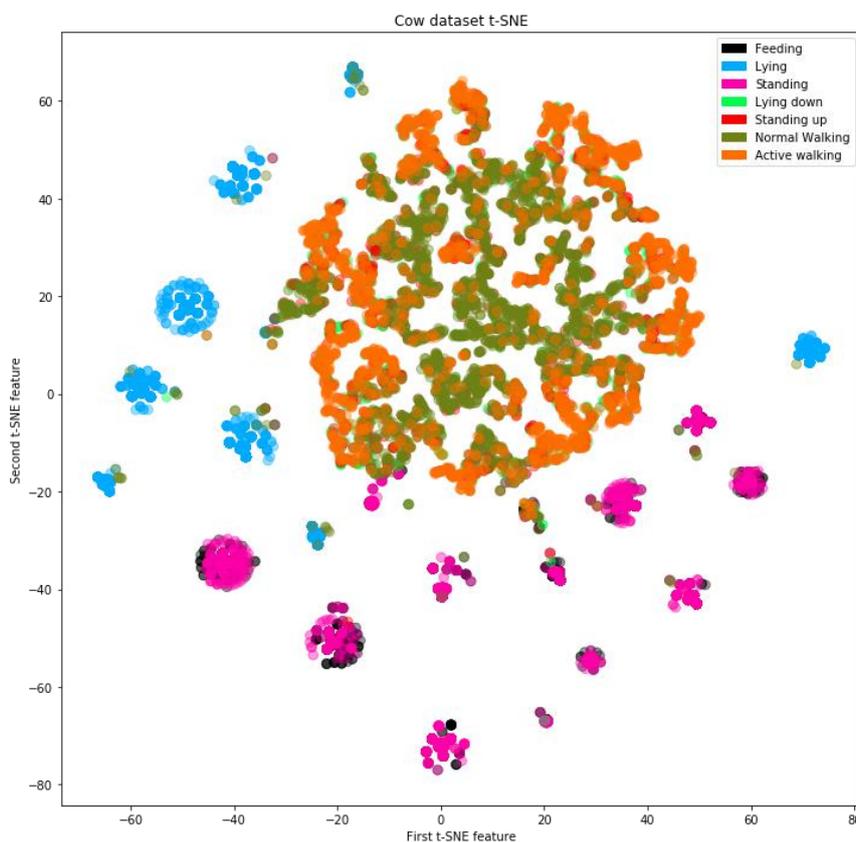

Figure 5. Representation of original data. Each color point represents for a 3-Dimension data point (X, Y, Z).

As we can see in Figure 5, the four states (standing up, lying down, normal walking and active walking) are overlapped each other. It is also noted that feeding state overlapped significantly with standing while lying state is dispersed. From this point, applying machine learning to improve the classification performance is reasonable if we could extract features of the acceleration data.

We conducted some experiments that converted each 16-sample-length signal into a vector which characterized of the 5 proposed features (mean, median, SD, RMS and range were calculated by (2−6) for all records of X, Y and Z). Figure 6 presents the result when the combination of 5 features was experimented.

---

[2]t-Distributed Stochastic Neighbor Embedding (t-SNE) is a technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional datasets
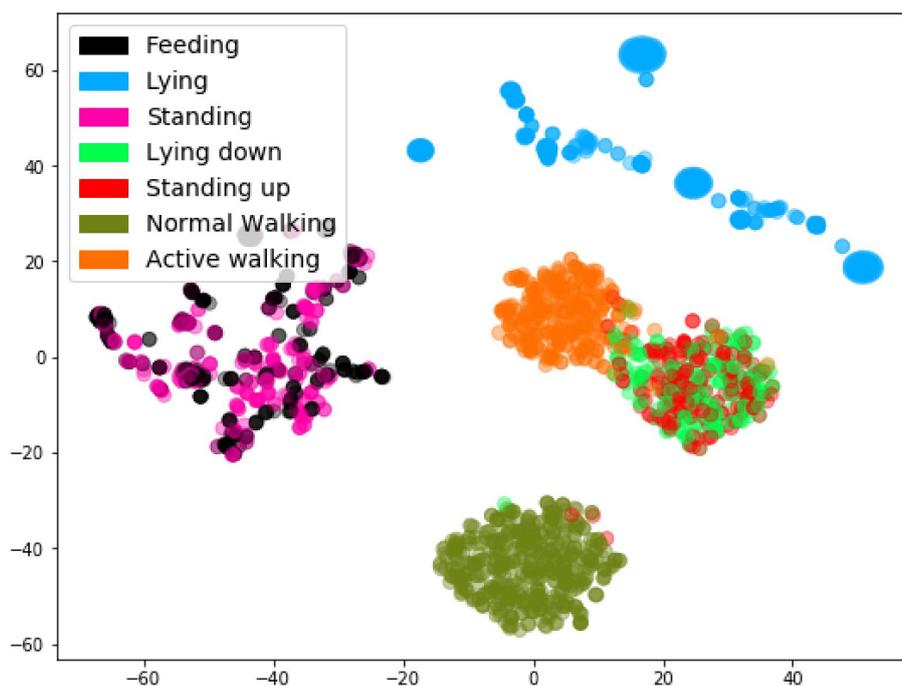
Figure 6. t-SNE representation of data when using all 5 features. Each colored point is represented for a vector which obtained by calculating features (mean, median, SD, RMS and range).

2.6. Evaluation method

We provide definitions and some results for classifications that detect the presence of behavior (a classification result is either positive or negative, which may be true or false):

- A true positive classification result is one that detects the behavior when the behavior is present.
- A true negative classification result is one that does not detect the behavior when the behavior is absent.
- A false positive classification result is one that detects the behavior when the behavior is absent.
- A false negative classification result is one that does not detect the behavior when the behavior is present.

We evaluated the performance of the algorithm based on accuracy, sensitivity, positive predictive value and negative predictive value.

- Accuracy is the fraction of classifications was correct, equals to "number of correct predictions"/"total number of predictions".
- Sensitivity measures the ability of a classification to detect the behavior when the behavior is present.
- Positive predictive value (PPV) presents the proportion of positive classifications.
- Negative predictive value (NPV) presents the proportion of negative classifications.

From the definitions, we have formulae:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{7}$$

$$Sen = \frac{TP}{TP + FN} \tag{8}$$

$$PPV = \frac{TP}{TP + FP} \tag{9}$$

$$NPV = \frac{TN}{TN + FN} \tag{10}$$

where *TP* is True Positives (all true positive), *TN* is True Negatives, *FP* is False Positives and *FN* is False Negatives; *Acc* is denoted for the accuracy, *Sen* is the sensitivity, *PPV* is the positive predictive value, and *NPV* is the negative predictive value.

3. Results and discussions

3.1. Experiment results

The experiment was devised to assess the classification performance of 4 individual machine learning methods tested with different windows (6-second, 12-second, 16-second and 20-second) and 5 features (mean, median, standard deviation, RMS and range) (section 2.5). We made comparisons between the performance of Gradient Boosted Decision Tree (GBDT), Support Vector Machines (SVM), Random Forest (RF) and K-Nearest Neighbor (KNN) in term of accuracy and sensitivity (calculated by (7–8) for 7 behaviors). Figure 7 shows the performance of these classification algorithms.

In any case, there was a great dependence on the window size, the overall accuracy and sensitivity increased with window size for all the algorithms considered. From the results, it can be stated that the general trend is that window size does have an influence on classification performance.

A suitable window will reduce the computational burden of the classification algorithms, the effects of noise, and the temporal dependence of subsequent examples. However, there is a trade-off: the longer the window length, the more these positive benefits are realized; if the window length becomes too large, the probability that a given window contains more than one activity is increased, the delay before a classification output can be generated is increased, and the number of training examples for the classification will also be reduced [33].

Considering all the options above, we found that the GBDT algorithm (16-s window) probably achieved the overall highest performance, with overall accuracy 86.3% and overall sensitivity 80.6%.
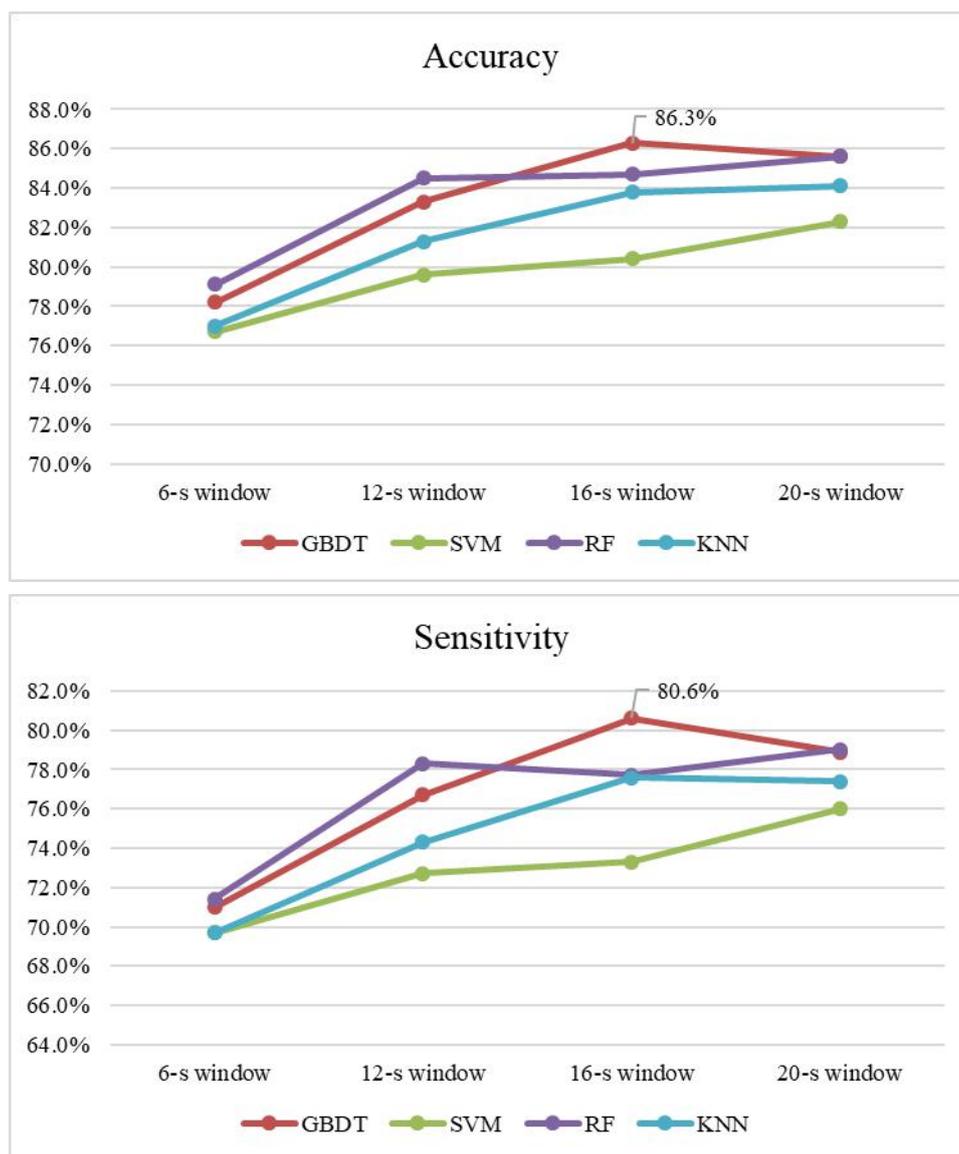
Figure 7. Performance comparison of 4 different machine learning algorithms using the same input data set (section 2.4). Performance measures (accuracy and sensitivity) were obtained using 6-s, 12-s, 16-s and 20-s windows. Overall accuracy was calculated as the arithmetic mean accuracy for 7 behaviors. Overall sensitivity was calculated in a similar manner. GBDT algorithm (16-s window) achieved the highest performance (accuracy 86.3% and sensitivity 80.6%). Legend: Gradient Boosted Decision Tree (GBDT), Support Vector Machines (SVM), Random Forest (RF) and K-Nearest Neighbor (KNN).

3.2. Performance evaluation of Gradient Boosted Decision Tree model

The detailed results of using GBDT classifier (16-s window) which give the best overall performance (section 3.1) is presented in Table 8. Table 8 shows the number of cases that have been correctly identified as positive (the modeled behavior), as well as correctly identified as negative (other behaviors). Cases where a negative sample has been misclassified as positive and vice versa, are called false positives and false negatives, respectively (section 2.6).

Table 8. The confusion matrix achieved from the classification of dairy cow behaviors test data (1456 records, explained in sections 2.3 and 2.4) by the Gradient Boosted Decision Tree algorithm (16-s window).

True Label	Predicted behavior							Total
	Feeding	Lying	Standing	Lying down	Standing Up	Normal walking	Active Walking	
Feeding	**220**	0	23	0	0	0	0	243
Lying	0	**290**	0	0	0	0	0	290
Standing	57	0	**120**	0	0	2	0	179
Lying down	0	0	0	**71**	53	0	5	129
Standing up	0	0	0	47	**65**	3	5	120
Normal walking	0	0	0	1	0	**290**	0	291
Active walking	0	0	0	2	2	0	**200**	204
Total	277	290	143	121	120	295	210	1456

Adapting from Table 8, lying (290/290), normal walking (290/291) and active walking (200/204) were well classified, only few misclassifications of the behaviors occurred (1/291 normal walking and 4/204 active walking). Feeding and standing were misclassified with each other (57/179 feeding were misclassified as standing and 23/243 standing were misclassified as feeding). Lying down were misclassified with standing up (53/129 lying down were misclassified as standing up and 47/120 standing up were misclassified as lying down).

Performance of the model was evaluated based on four indicators (section 2.6), namely accuracy (7), sensitivity (8), PPV (9) and NPV (10) as shown in Table 9.

Table 9. Summary of the performance indicators of the GBDT algorithm for all behavior categories.

Behavior pattern	Algorithm performance indicators			
	Accuracy	Sensitivity	PPV	NPV
Feeding	94.0%	90.5%	79.4%	97.8%
Lying	100%	100%	100%	100%
Standing	93.9%	67.0%	83.9%	95.1%
Lying down	92.1%	55.0%	58.7%	95.3%
Standing up	91.9%	54.2%	54.2%	95.6%
Normal walking	99.5%	99.7%	98.3%	99.9%
Active walking	98.9%	98.0%	95.2%	99.6%

The overall performance of GBDT model was good. Sensitivity was high (>90%) for all class except standing (67%), lying down (55%) and standing up (54.2%). Accuracy was excellent for all classes (>91.9%) of behaviors, as well as for the overall classification performance. PPV was good for lying (100%), standing (83.9%), feeding (79.4%), normal walking (98.3%), active walking (95.2%). Lower PPV values for standing up (54.2%) and lying down (58.7%) indicate that the classifier had problem predicting positive cases correctly in these classes, which further suggests that these behavior patterns became most easily confused with other behaviors. NPV was

also excellent for all classes (>95%) of behaviors.

3.3. Discussions

This study pointed out that using machine learning with 5 features (mean, SD, RMS, median, range) and 16-second window of data (1 sample/second) perform the best results for distinguishing behaviors in the proposed system. In particular, the GBDT algorithm provided excellent overall classification performance (in term of accuracy, sensitivity, PPV and NPV) for 5/7 classes of behavior. We are able to discriminate more types of behavior than the monitoring system reported by Diosdado et al. [9] and Arcidiacono et al. [19]. The statistical performance of GBDT algorithm addressed in Table 9, was obviously higher than the previous studies [20,28] (more detailed discussions below).

P. Martiskainen and M. Jarvinen [28] used support vector machines to classify various behaviors in dairy cows based on collar-mounted acceleration data. We would like to compare our results with theirs for the same behaviors. In detail, we obtained better accuracy for standing (93.9% versus 87%), lying (100% versus 84%), and normal walking (99.5% versus 99%). Our method offers the better sensitivity for lying (100% versus 80%), feeding (90.5% versus 75%), and normal walking (99.7% versus 79%). For PPV, we also had better performances in standing (83.9% versus 65%), lying (100% versus 83%), normal walking (98.3% versus 79%). The only behavior P. Martiskainen and M. Jarvinen [28] obtained better results is feeding (accuracy 96%, PPV 81%) (ours: accuracy 94.0%, PPV 79.4%). The reason for that is they used collar-mounted acceleration data instead of leg-mounted acceleration data. It made distinguishing feeding and standing better (as we can see in Table 8, our method misclasssified between feeding and standing).

We made detailed comparisons in term of overall performance to the work of Jun Wang et al. [20] which used the same dataset. Table 10 shows the comparison result when we used macro-average evaluation method which computes the metric independently for each class and then take the average (hence treating all classes equally). Table 11 shows the comparison result when we used the micro-average evaluation method which aggregates the contribution of all classes to computed average values.

Table 10. Overall performance indicators used macro-average evaluation method.

Indicators	Jun Wang et al. [20]	Our work
Accuracy	92.3%	95.8%
Sensitivity	79.1%	80.6%
PPV	82.1%	81.4%
NPV	Not provided	97.6%

Table 11. Overall performance indicators used micro-average evaluation method.

Indicators	Jun Wang et al. [20]	Our work
Accuracy	86.6%	96.6%
Sensitivity	85.2%	86.3%
PPV	79.8%	86.0%
NPV	Not provided	98.3%

Jun Wang et al. [20] did not provide NPV value, so we could not make a comparison of NPV value. As we can see from Tables 10 and 11, in both case NPV values were excellent (97.6% with macro-average evaluation method, 98.3% with micro-average evaluation method).

In the case of macro-average method, our result outperformed the Jun Wang et al. [20]'s result in terms of accuracy (95.8% versus 92.3%) and sensitivity (80.6% versus 79.1%) but theirs was slightly better in term of PPV (82.1% versus 81.3%). When taking micro-average as an evaluation method, our performance was better than Jun Wang et al. [20] in term of accuracy (96.6% versus 86.6%), sensitivity (86.3% versus 85.2%) and PPV (86.0% versus 79.8%). This means our solution gave better result in predicting the classes which have majority observations.

There was an imbalance in the proportion of behavior data (class). In our dataset, we have 3678 records/7 classes (average is about 525 records/class), class lying down is only 326 records and class standing up is only 303 records (Table 4). So, micro-average evaluation method is preferable for the evaluation classification of this dataset. In this case, our performance was proved is better than Jun Wang et al. [20] for all considered indicators.

On the other hand, sensitivity values were generally high, meaning that not many negatives cases were falsely classified as positive. The performance of GBDT algorithm as shown in Table 9 was high with exceptions of lying down and standing up. Standing up and lying down have significant similarity in data characteristics, which directly lead to the confusion in behavior identification [20].

The misclassification between feeding and standing (shown in Table 8) was clearly explained in [29]. In this paper, the authors pointed out that when the cow rotated its head during the feeding activity, the sensor's position was deflected. Furthermore, in some cases, even the cow was standing, its head was still down. The authors also suggested that the assembling of the sensor to the cow should be improved in order to reduce the misclassification between feeding and standing.

Although our approaches provided better performance than the previous works in literature, it still remains some limitations such as: low classification performance for standing up and lying down. Fortunately, in real world application, when standing up and lying down states do not occur as frequently as other states, our excellent results in predicting the remaining 5 states seemed to be reasonable for applying in real-time system [34,35]. We also noticed the poor result of distinguishing lying down and standing up states indicates that different parameters (e.g. time windows) should be applied to the modeling process, depending on the characteristics of the observed behavior.

In this study, we re-used data already available of Jun Wang et al. [20]. Researchers like Jun Wang et al. may be motivated to share data if it results in greater visibility of their work and increasing reputation. Academic researchers have many novel research opportunities with open research data (e.g. analyzing large volumes of data, testing novel hypotheses, research replication, etc.). The combination of data from multiple sources enables the generation of new datasets, information, and knowledge. In particular, the amount of data is critical for machine learning algorithms' performance.

4. Conclusion

In this paper, we designed an efficient set of features then applied for cows' behavior classifiers. The selection of the window length (16 s) also has an important role in classification. We demonstrated that the Gradient Boosted Decision Tree method allows accurate determination seven

cow motion states of from acceleration data (feeding, lying, standing, lying down, standing up, normal walking, active walking). Compared to the method of Jun Wang et al. [20], we provided better performance (in terms of the sensitivity, the accuracy, the positive predictive value, and the negative predictive value) of the estimation, which is critical for classification of cow behavior. The provided approach corresponds to the acceleration data of the leg tag sensor systems. However, since this study re-used the database of Jun Wang et al. [20], more validation should be done with another dataset (e.g. in the barn by using other systems which validated by direct observation or computer vision). In particular, our results should be extended to the case by using the time tag recorded to improve the classification of lying down and standing up.

Acknowledgments

Conflict of interest

The author declares that he has no conflict of interest.

References

1. G. Mattachini, E. Riva, C. Bisaglia, J. C. A. M. Pompe, G. Provolo, Methodology for quantifying the behavioral activity of dairy cows in free-stall barns, *J. Anim. Sci.*, **10** (2013), 4899–4907.

2. A. Rahmana, D.V. Smitha, Cattle behaviour classification from collar, halter, and ear tag sensors, *Inf. Process. Agric.*, **5** (2018), 124–133.

3. S. M. C. Porto, C. Arcidiacono, Localization and identification performances of a real-time system based on ultra wide band technology for monitoring and tracking dairy cow behavior in semi-open free-stall barn, *Comput. Electro. Agric.*, **108** (2014), 221–229.

4. M. R. Borchers, Y. M. Chang, A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors, *J. Dairy Sci.*, **999** (2016), 7458–7466.

5. G. M. Pereira, J. H. Bradley, I. E. Marcia, Validation of an eartag accelerometer sensor to determine rumination, eating, and activity behaviors of grazing dairy cattle, *J. Dairy Sci.*, **101** (2018), 2492–2495.

6. H. C. Weigele, L. Gygax, A. Steiner, B. Wechsler, J. B. Burla, Moderate lameness leads to marked behavioral changes in dairy cows, *J. Dairy Sci.*, **3101** (2018), 2370–2382.

7. F. Mahmoud, B. Christopher, A. Maher, H. Jürg, S. Alexander, S. Adrian, H. Gaby, Prediction of calving time in dairy cattle, *Anim. Reprod. Sci.,* **187** (2017), 37–46.

8. N. Bareille, F. Beaudeau, S. Billon, A. Robert, P. Faverdin, Effects of health disorders on feed intake and milk production in dairy cows, *Livest. Prod. Sci.,* **83** (2003), 53–62.

9. J. A. V. Diosdado, Z. E. Barker, Classification of behavior in housed dairy cows using an accelerometer-based activity minitoring system, *Anim. Biotelemetry*, **3** (2015), 1–14.

10. K. M. Abell, M. E. Theurer, R. L. Larson, B. J. White, D. K. Hardin, R. F. Randle, Predicting bull behavior events in a multiple-sire pasture with video analysis, accelerometers, and classification algorithms, *Comput. Electro. Agric.*, **136** (2017), 221–227.

11. M. E. Theurer, D. E. Amrine, B. J. White, Remote noninvasive assessment of pain and health status in cattle, *Vet. Clin. Food Anim.*, **29** (2013), 59–74.

12. C. W. Maina, *IoT at the Grassroots–Exploring the Use of Sensors for Livestock Monitoring*. Ist-Africa Week Conference, 2017, 1–8.

13. N. Zehner, C. Umstatter, System specification and validation of a noseband pressure sensor for measurement of ruminating and eating behavior in stable-fed cows, *Comput. Electro. Agric.*, **136** (2017), 31–41.

14. J. Wang, Z. He, J. Ji, K. Zhao, H. Zhang, IoT-based measurement system for classifying cow behavior from tri-axial accelerometer, *Cienc. Rural*, **49** (2019), 1–13.

15. E. S. Nadimi, H. T. Søgaard, Observer Kalman flter identifcation and multiple-model adaptive estimation technique for classifying animal behaviour using wireless sensor networks, *Comput. Electro. Agric.*, **68** (2009), 9–17.

16. K. O'Driscoll, L. Boyle, A brief note on the validation of a system for recording lying behavior in dairy cows, *Appl. Anim. Behav. Sci.*, **111** (2008), 195–200.

17. M. S. Shahriar, D. Smith, Detecting heat events in dairy cows using accelerometers and unsupervised learning, *Comput. Electro. Agric.*, **128** (2016), 20–26.

18. J. M. Talavera, L. E. Tobón, J. A. Gómez, M. A. Culman, J. M. Aranda, D. T. Parra, et al., Review of IoT applications in agro-industrial and environmental fields, *Comput. Electro. Agric.*, **142** (2017), 283–297.

19. C. Arcidiacono, S. M. Porto, M. Mancino, G. Cascone, A threshold-based algorithm for the development of inertial sensor-based systems to perform real-time cow step counting in free-stall barns, *Biosyst. Eng.*, **153** (2017), 99–109.

20. J. Wang, Z. He, Development and validation of an ensemble classifier for real-time recognition of cow behavior patterns from accelerometer data and location data, *PLoS One*, **13** (2018).

21. B. Robert, B. J. White, D. G. Renter, R. L. Larson, Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle, *Comput. Electro. Agric.*, **67** (2009), 80–84.

22. B. D. Robért, B. J. White, D. G. Renter, R. L. Larson, Determination of lying behavior patterns in healthy beef cattle by use of wireless accelerometers, *Am. J. Vet. Res.*, **72** (2011), 467–473.

23. L. Atallah, B. Lo, R. King, G. Yang, Sensor positioning for activity recognition using wearable accelerometers, *IEEE Trans. Biomed. Circ. Syst.*, **5** (2011), 320–329.

24. J. Krause, S. Krause, Reality mining of animal social system, *Trends Ecol. Evol.*, **28** (2013), 541–551.

25. J. C. Davila, A. M. Cretu, M. Zaremba, Wearable sensor data classification for human activity recognition based on an iterative learning framework, *Sensors*, **17** (2017), 1287.

26. R. Muller, L. Schrader, A new method to measure behavioral activity levels in dairy cows, *Appl. Anim. Behav. Sci.*, **83** (2003), 247–258.

27. M. Sugiyama, M. Kawanabe, *Machine learning in Non-Stationary Environments*, MIT Press, 2012.

28. P. Martiskainen, M. Jarvinen, Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines, *Appl. Anim. Behav. Sci.*, **119** (2009), 32–38.

29. C. Arcidiacono, S. M. C. Porto, M. Mancino, G. Cascone, Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data, *Comput. Electro. Agric.*, **134** (2017), 124–134.

30. S. Bhattacharya, A. M. Krishna, D. Lombardi, A. Crewe, N. Alexander, Economic MEMS based 3-axis water proof accelerometer for dynamic geo-engineering applications, *Soil Dyn. Earthq. Eng.*, **36** (2012), 111–118.

31. M. Alsaaod, J. J. Niederhauser, G. Beer, N. Zehner, G. S. Regula, A. Steiner, Development and validation of a novel pedometer algorithm to quantify extended characteristics of the locomotor behavior of dairy cows, *J. Dairy Sci.*, **98** (2015), 6236–6242.

32. M. Janidarmian, A. R. Fekr, K. Radecka, Z. Zilic, A comprehensive analysis on wearable acceleration sensors in human activity recognition, *Sensors*, **17** (2017), 1–26.

33. N. Twomey, T. Diethe, A comprehensive study of activity recognition using accelerometers, *Informatics*, **5** (2018), 1–37.

34. C. P. K. Phung, T. K. Nguyen, D. C. Nguyen, D. N. Tran, D. T. Tran, Classification of cow's behaviors based on 3-DoF accelerations from cow's movements, *Int. J. Electr. Comput. Eng.*, **9** (2019), 1656–1662.

35. Q. T. Hoang, C. P. K. Phung, T. N. Bui, T. P. D. Chu, D. T. Tran, Cow behavior monitoring using a multidimensional acceleration sensor and multiclass SVM, *Int. J. Mach. Learn. Networked Collab. Eng.*, **2** (2018), 110–118.